Linear Aggregation in Linear Regression

Harry Lütjohann

Stockholm 1974

Linear Aggregation in Linear Regression

Harry Lütjohann

Stockholm 1974

ISBN 91–7222–049–X
Sundt Offset Stockholm 74

Linear Aggregation in Linear Regression

Akademisk avhandling

av

Harry Lütjohann

som med tillstånd
av Samhällsvetenskapliga fakulteten vid Stockholms universitet
för vinnande av filosofisk doktorsgrad
framställes till offentlig granskning
på lärosal F, Norrtullsgatan 2,
tisdagen den 28 maj 1974 kl. 10 f.m.

Stockholm 1974

PREFACE

An aggregation involves two sets of variables and three relations. The two sets of variables are called micro and macro variables. The first relation is a micro relation in the micro variables. The second relation is a macro relation in the macro variables. The third relation is an aggregating relation, which expresses the macro variables as functions of the micro variables.

Aggregation is intrinsically problematic in the following sense. If the micro, macro and aggregating relations are to be logically consistent, they must be carefully chosen so as to fit together. Starting from this fundamental point, one can formulate a great variety of more or less specific aggregation problems.

Aggregation has often been studied by econometricians, who formulated their aggregation problems in economic terms. But aggregation problems are formal in the sense that empirical knowledge does not contribute to their solution. Aggregation problems, like identification problems, are about the logic of mathematical or statistical models.

Two different approaches to aggregation are predominant in the econometric literature. The first approach concentrates on pure economic theory. The second approach is concerned with the statistical estimation of economic models.

The first approach usually assumes deterministic models. A frequent purpose is to find conditions such that the three relations of an aggregation are logically consistent. This may be called the consistency approach to aggregation. A survey of the consistency approach is given in J. Green, "Aggregation in Economic Analysis", Princeton 1964. The general result is that consistency in aggregation is a scarce commodity.

The second, statistical, approach originated with H. Theil, "Linear Aggregation of Economic Relations", Amsterdam 1954. This approach assumes micro and macro relations that consist of linear statistical models of the kind used in regression analysis. The parameters of these models are called micro and macro parameters. Micro data are assumed to be generated in accordance with the micro relation. The micro data are aggregated into macro data. The macro relation is estimated from the macro data by means of regression analysis. The general result is that the macro parameters thus estimated can always be expressed as linear functions of the micro parameters.

The consistency approach indicates that aggregations are as a rule inconsistent, also linear aggregations of linear relations. The statistical approach, on the other hand, seems to derive the macro relation from the micro and aggregating relations. The two approaches should be confronted. Do they contradict each other? This question is the point of departure of the present study.

* * * * *

This book is about linear aggregation in a context of linear regression analysis. It can be divided into three parts. The first part, chapters 1 to 3, introduces the concepts and theoretical tools required. The second part, chapters 4 to 8, outlines, performs and discusses a formal analysis of certain linear aggregations in linear regression. The third part, chapter 9, discusses some of the tacit assumptions of the formal analysis.

The nine chapters are divided into sections, and these into subsections. Sections and subsections are referred to by their underlined numbers, for example 1.2 and 2.5.1. Important results are formulated as propositions. A proposition carries the number of the subsection where it occurs, with a distinguishing letter added if necessary. For example, the second proposition of 2.4.1 is called P.2.4.1.B. Figures and tables are numbered analogously. References are identified by the author's name and the year of publication, for example Malinvaud [1956].

Without loss of essential continuity, section 2.5, chapter 6, and section 8.3 may be skipped. In any case, chapters 1, 4, 8 and 9 should not be skipped.

* * * * *

Chapter 1 introduces some basic concepts and broad problems of aggregation. The concept of consistency is slightly generalized, so as to become applicable to micro and macro relations that consist of linear statistical models. Starting from the fact the restrictions on the set of admissible data may be favourable to consistency, two basic types of consistency problems are distinguished. The central result of Theil's analysis of linear aggregation is formulated as an interpretation in micro terms of the macro regression coefficients.

Chapter 2 reviews the parts of least-squares theory needed later. Generalized least-squares and a regressor matrix of arbitrary rank are always admitted. Throughout, a distinction is maintained between properties that are based on the linear model, and model-free properties that are not. The concept of a unique coefficient function is introduced as a model-free counterpart of the concept of an estimable parameter function. The relation of estimability to identification is indicated. A distinction is introduced between designed and observed regressor data. Certain expressions in deviation form are explicitly derived, using matrix notation.

Chapter 3 indicates five different types, or structures, of linear aggregation. The macro data are in every case data from which one regression can be computed, the macro regression. Two of the types of aggregation are in the dimension of the regression variables, and two are in the dimension of the observations, which are called units of analysis. The fifth type of aggregation, which is well known to economists, is analyzed and decomposed into two of the simpler types.

Chapter 4 outlines the formal analysis to be applied in the next three chapters. Any micro or macro relation considered consists of linear statistical models. Any aggregating relation considered is linear. The scheme of analysis comprises two aggregation problems. The first problem is about the consistency or otherwise of an aggregation. The formulation of this problem takes into account the distinction between designed and observed regressor data. The second problem is concerned with the interpretation in micro terms of a unique coefficient function in the macro regression. The interpretation occurs in two variants, one of which is model-free, while the other one is based on the linear models of the micro relation. The latter interpretation is a translation into different terminology of Theil's analysis of linear aggregation. The theoretical core of the formal analysis is the three propositions P.4.2.6, P.4.3.4, and P.4.4.3.

Chapters 5, 6 and 7 apply the scheme of formal analysis to the five types of aggregation, three of which occur in one general and one special variant. The analysis of the fifth type of aggregation is based on that of the first two types.

Chapter 8 sums up and discusses the formal analysis. Some simple examples are used as a basis for an analysis of the macro parameters and macro relation established by Theil's analysis of linear aggregation. The conclusion is that

Theil's macro relation cannot, except in special cases, be a linear statistical model. Such models are usually assumed to be invariant with respect to the regressor data, but Theil's macro relation does not share this invariance property. Instead, Theil's macro relation is interpreted as a certain model-free linear regression. Further, the usefulness of the concept of aggregation bias is questioned. Finally, there is a brief comment on the role of the coefficient of determination in aggregation theory.

* * * * *

Even linear aggregations of linear relations are as a rule inconsistent. Therefore, if the macro relation derived by Theil is assumed to conform to the standard specification of a linear statistical model, then there is a contradiction. But the outcome of the formal analysis is that Theil's macro relation is not a linear statistical model. Consequently, Theil's statistical approach to aggregation cannot be used to bypass the fundamental difficulty that most aggregations are inconsistent.

Chapter 9 reconsiders the preceding analysis from a wider perspective. Simply following practice, the formal analysis has treated all regressor data as nonstochastic constants. This assumption is now questioned, and an alternative is suggested. The units of analysis are assumed to be drawn at random from a multivariate normal distribution. This distribution is assumed to be common to all units of analysis, so that there is a systematic similarity disturbed by random variation. The consequences are striking. The two aggregation problems of the formal analysis lose most of their appeal. The theme underlying the formal analysis was the conflict between the consistency approach and the statistical approach to aggregation. This conflict now loses its sharpness.

Aggregation problems are about the logic of the models assumed.

* * * * *

I wish to express my thanks to the following people for the support and assistance they have given me. Professor Sten Malmquist supervised my research on aggregation during the many years preceding the completion of this book. Professor Ove Frank and Docent Anders Klevmarken read and discussed with me various parts of earlier versions. The final typing and editing was done by Mrs. Marion Ekström and Mrs. Git Sundt.

* * * * *

Finally, a quotation. In one of his books, the German poet Christian Morgenstern described the exploits of the two friends Palmström and von Korf.

> Korf erfindet eine Tagnachtlampe,
> die, sobald sie angedreht,
> selbst den hellsten Tag
> in Nacht verwandelt.

LIST OF CHAPTERS

CONTENTS

1 SOME FUNDAMENTAL CONCEPTS AND PROBLEMS OF AGGREGATION

1.1 Some fundamental concepts

1.1.1 Aggregation and disaggregation

Statistical methods are not applied to the real world directly, but to data representing the real world. Preliminary to statistical analysis, the relevant information on the real world must be formulated as data. A set of statistical data can be looked upon as a numerical picture of a selected segment of the real world.

Numerical pictures, like maps, can be made in more than one way. In particular, two numerical pictures of the same aspects of the real world may differ as to degree of detail. More precisely, they may differ in such a manner that the more summary picture can be constructed from the more detailed picture.

For example, consider an imaginary econometric study. The aspects of the real world to be studied are the consumer behaviour, during eight weeks, of each of one thousand households of equal size. For each week separately, each household registers its income and its expenditure on each of nine groups of commodities and services. The resulting detailed data consists of 80,000 pieces of data.

Now suppose that the nine commodity groups are collapsed into three. Further, incomes and expenditures are registered for the whole eight-week period only. Finally, the households are grouped into fifteen income classes, and the only data reported are the averages for each group. The resulting summary data consist of 60 pieces of data.

The two sets of data in the example are numerical pictures of the same segment of the real world. The two pictures differ as to degree of detail. The summary picture can be derived from the detailed picture.

Some terminology will now be introduced. Consider two different procedures for describing a segment of the real world by means of statistical data. If the real world were sufficiently different, the data would be different. Thus, each single datum is to be regarded as a variable. Assume that the two procedures are such that, whatever the reality described, the more summary data can be computed from the more detailed data by a fixed rule of computation. The more

detailed data are then called the micro data, the more summary data the macro data. The process of exchanging the micro data for the macro data is called an aggregation, the reverse process a disaggregation.

Let the micro data be arranged into a variable vector d, the macro data into a variable vector D. The fixed rule by which the macro data are computed from the micro data is a vector function F,

$$D = F(d) \ .$$

This vector function is called the aggregating function.

As a rule, aggregations are such that the macro data vector D has fewer elements than the micro data vector d, and such that an inverse (disaggregating) function F^{-1} does not exist. In other words, as a rule some information is lost in the aggregation, and cannot be regained by a purely formal operstion.

1.1.2 Micro and macro relations

The purpose of a statistical analysis of data is often to investigate some theory. A theory restricts the behaviour of the real world. If the numerical picture used is relevant, the theory implies restrictions on the data. Often, but not always, these restrictions are to the effect that certain elements of the data are functions of the other elements.

In the example of 1.1.1, the theory could be that all demand functions are constant-elastic and common to all households. If the micro data are a relevant numerical picture, this theory implies that the logarithm of the expenditure on a given commodity group is a linear function of the logarithm of income. Further, the same function applies to every household and week.

Alternatively, one might consider the macro data relevant, and would then expect these to satisfy log-linear relations, each relation being valid for all fifteen income groups.

In the example, there is one single theory, although it is somewhat uncertain whether it restricts the micro data, the macro data, or perhaps both. There could also have been two different theories, one for the micro data and another one for the macro data.

Further, in the example different parts of the data can be compared in order

to provide a partial check on the theory. For example, the data can be plotted on double-logarithmic paper. In other cases there are fewer data, in particular macro data, and no such internal checking is possible.

Some more terminology and notation will now be introduced. A theory of the kind exemplified divides the relevant set of data into two subsets. Thus, the micro data vector is partitioned into the independent micro data vector x and the dependent micro data vector y. The restrictions on the micro data are expressed by the vector function

$$y = \varphi(x) ,$$

which is called the micro relation.

Analogously, the macro data are partitioned into the independent macro data vector z and the dependent macro data vector u. The restrictions on the macro data are expressed by the vector function

$$u = \psi(z) ,$$

which is called the macro relation.

1.1.3 Semi-aggregation and semi-disaggregation

The distinguishing feature of aggregation is that the macro data are a function $D = F(d)$ of the micro data. This need not mean that each single macro datum is a function of every micro datum. For instance, in the example of 1.1.1, macro income data are functions of the micro income data only, and similarly for the expenditure data. More generally, the independent macro data are functions of the independent micro data only, and similarly for the dependent macro and micro data. Aggregations which have this property will be called segregated.

In segregated aggregation, the aggregating function is split into two separate vector functions. One is the aggregating function for independent data

$$z = g(x) ,$$

and the other one is the aggregating function for dependent data

$$u = h(y) .$$

This terminology will be reconsidered in 3.1.1.

A segregated aggregation involves four variable data vectors and four fixed vector functions. Their interrelations are shown in Figure 1.1.3, which is, apart from notation, due to Malinvaud [1956]. As a rule, the four vector functions are not assumed to have inverse functions.

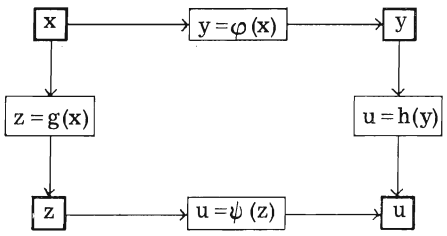


Fig. 1.1.3 The formal structure of segregated aggregation

If the micro relation φ is substituted into the aggregating function for dependent data h, the outcome is a composite vector function

$$u = h(\varphi(x)) \ ,$$

which will be called the semi-aggregated micro relation. In the figure, it is represented by the arrow path from x by way of y to u.

If the aggregating function for independent data g is substituted into the macro relation ψ, there results a composite vector function

$$u = \psi(g(x)) \ ,$$

which will be called the semi-disaggregated macro relation. Its representation in the figure is the arrow path from x by way of z to u.

This study is concerned with segregated aggregation only.

1.1.4 The concept of point-consistency

Four fundamental and two derived relations have been introduced above.

G:	$z = g(x)$	aggregating function for independent data,
H:	$u = h(y)$	aggregating function for dependent data,
Φ:	$y = \varphi(x)$	micro relation,
Ψ:	$u = \psi(z)$	macro relation,
$H\Phi$:	$u = h(\varphi(x))$	semi-aggregated micro relation,
ΨG:	$u = \psi(g(x))$	semi-disaggregated macro relation.

Throughout this study, the single and paired letters G, H, Φ, Ψ, $H\Phi$, and ΨG will be used as a brief symbolic notation.

The term "aggregation" can be given a narrower or a wider meaning. An aggregation in the narrower sense is defined by the two aggregating functions alone; in symbols $\{G, H\}$. Aggregation in this sense is unproblematic. An aggregation in the wider sense includes into the definition the micro and macro relations, too; in symbols $\{G, H, \Phi, \Psi\}$. Below, the term is used in the wider sense.

Any four vector functions of appropriate orders define an aggregation $\{G, H, \Phi, \Psi\}$. Consider such an aggregation and a given independent micro data vector x. What dependent macro data vector u corresponds to x ? The semi-aggregated micro relation $H\Phi$ gives an answer. The semi-disaggregated macro relation ΨG gives an answer, too. The two answers may well turn out to be different. The situation is illustrated by Figure 1.1.4.

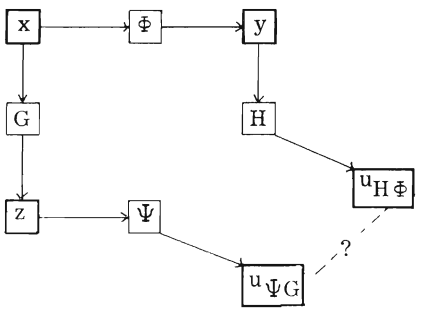


Fig. 1.1.4 The question of consistency

If an aggregation is such that, for a given admissible independent micro data vector x ,

$$h(\varphi(x)) = \psi(g(x)) \ ,$$

the aggregation is said to be point-consistent for that x. Otherwise it is point-inconsistent for the x in question. Loosely speaking, point-inconsistency is the rule and point-consistency the exception. This is the root of the complex of problems sometimes named, collectively, "the" aggregation problem.

A different but related concept of consistency will be introduced in 1.2.3.

1.1.5 A generalization: Stochastic relations

Up to this point, micro and macro relations have been assumed to be deterministic vector functions, where an independent vector determines a dependent vector exactly. This was a simplification. In econometrics, economic theories are not normally taken to imply such exact restrictions on the data. Instead, the typical micro or macro relation is what can be called a stochastic vector function. The dependent vector is assumed to be stochastic, and the independent vector determines its probability distribution.

Assume that the micro and macro relations are stochastic vector functions. The dependent "quantities" of the micro and macro relations are no longer the dependent micro and macro data vectors y and u, but probability distributions for them,

$$\Pr(y \leq y_o) = \Theta(y_o) ,$$
$$\Pr(u \leq u_o) = \Omega(u_o) ,$$

By the micro and macro relations, the independent micro and macro data vectors x and z determine the respective distributions. In symbols,

$$\Phi: \quad x \rightsquigarrow \Theta(\cdot) ,$$
$$\Psi: \quad z \rightsquigarrow \Omega(\cdot) .$$

The aggregating functions are as before but are now written

$$G: \quad x \rightsquigarrow z ,$$
$$H: \quad y \rightsquigarrow u .$$

in conformity with the notation for Φ and Ψ.

The aggregation now involves four data vectors, two probability distributions, two deterministic and two stochastic vector functions. Their interrelations are indicated by Figure 1.1.5.A, which is, apart from notation, due to Malinvaud [1956].

By the aggregating function H, the vector u is a function of the vector y. Given certain mathematical reservations on the nature of H (see e.g. Wilks [1962] section 2.8) this implies that the probability distribution of y determines that of u. In symbols,

$$H^*: \quad \Theta(\cdot) \rightsquigarrow \Omega(\cdot) .$$

If the micro relation Φ is combined with H^*, the outcome is the semi-aggregated micro relation

$$H\Phi: \quad x \rightarrow \Theta(\cdot) \rightarrow \Omega(\cdot) \ .$$

If the aggregating function G is combined with the macro relation Ψ, the semi-disaggregated macro relation

$$\Psi G: \quad x \rightarrow z \rightarrow \Omega(\cdot)$$

is immediately obtained.

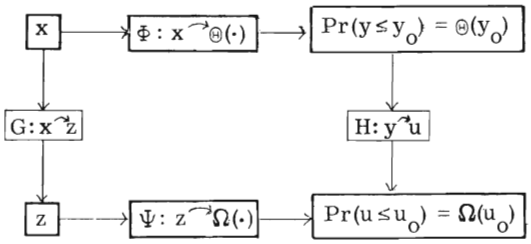


Fig. 1.1.5.A The formal structure of aggregation with stochastic relations

Any set of two deterministic and two stochastic vector functions of appropriate orders define an aggregation $\{G, H, \Phi, \Psi\}$. What probability distribution Ω of u corresponds to a given x ? The answers given by the semi-aggregated micro and semi-disaggregated macro relations may well differ. The situation is illustrated by Figure 1.1.5.B.

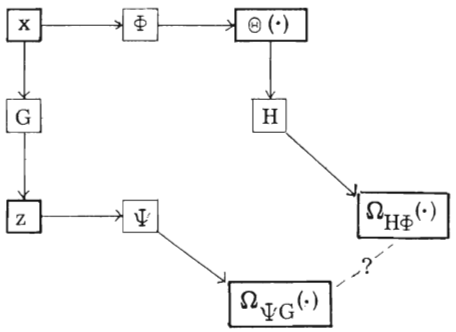


Fig. 1.1.5.B The question of consistency for stochastic relations

A possible generalization of the concept of point-consistency is as follows. An aggregation is considered to be point-consistent for a given admissible independent micro data vector x, if and only if for that x

$$\Omega_{H\Phi}(u_o) = \Omega_{\Psi G}(u_o) \quad \text{for every} \quad u_o \ .$$

This concept may be called distributional point-consistency, because it means that at the point x the two probability distributions of u are identical.

The above formal apparatus is very much inspired by Malinvaud [1956], section I.

1.1.6 Expectational point-consistency

Distributional point-consistency means that two probability distributions are identical. A less stringent requirement is often preferable. A weaker concept of consistency will now be defined, for use in 4.2 and later.

Let $E_{H\Phi}(u)$ denote the expected dependent macro data vector according to the distribution $\Omega_{H\Phi}(u)$, and similarly for $E_{\Psi G}(u)$ and $\Omega_{\Psi G}(u)$. Both expected vectors are functions of the independent micro data vector x. Often, $H\Phi$ and ΨG are such that no other properties of Ω change with x.

If an aggregation is such that, for a given admissible independent micro data vector x,

$$E_{H\Phi}(u) = E_{\Psi G}(u) ,$$

the aggregation is said to be expectationally point-consistent for that x. Otherwise it is expectationally point-inconsistent for the x in question. If the expectations do not both exist, the concept of expectational consistency is not applicable.

On the terminology in 1.1, see 1.4.2.

1.2 Problems related to consistency

1.2.1 Four examples

Four examples will be studied. Each is a different specialization of the following class of aggregations. The four data vectors are

$$x = \{x_0, x_1, \ldots, x_h, \ldots, x_q\} , \quad q \geq 2 ,$$
$$y = \{y_1, \ldots, y_h, \ldots, y_q\} ,$$
$$z = \{z_0, z_1\} ,$$
$$u = \{u\} .$$

The four vector functions are

$$G: \quad z_0 = x_0 \; ; \quad z_1 = \sum_{h=1}^{q} g_h x_h \; ,$$

$$H: \quad u = \sum_{h=1}^{q} y_h \; ,$$

$$\Phi: \quad y_h = \alpha_h x_0^{\varkappa_h} x_h^{\beta_h} \; , \quad h = 1, \ldots, q \; ,$$

$$\Psi: \quad u = \gamma z_0^{\lambda} z_1^{\delta} \; .$$

The semi-aggregated micro and semi-disaggregated macro relations are

$$H\Phi: \quad u = \sum_{h=1}^{q} \alpha_h x_0^{\varkappa_h} x_h^{\beta_h} \; ,$$

$$\Psi G: \quad u = \gamma x_0^{\lambda} \left(\sum_{h=1}^{q} g_h x_h \right)^{\delta} \; .$$

Only independent micro data $x_h > 0$, $h = 0, 1, \ldots, q$ are admissible. The parameters are restricted $\alpha_h > 0$, $\varkappa_h < 0$, $\beta_h > 0$, $h = 1, \ldots, q$; $\gamma > 0$, $\lambda < 0$, $\delta > 0$.

A possible interpretation is that Φ is the constant-elastic demand functions of q households, and Ψ the analogous "aggregate" demand function. Then x_0 is the price of the commodity in question, while x_h, $h \geq 1$, are the incomes, and y_h the quantities demanded.

Example 1.2.1.1. Here, $q = 2$, $g_1 = g_2 = 1$, $\beta_1 = \beta_2 = 1$, and $\delta = 1$. Further, $\alpha_1 = 1$ and $\varkappa_1 = -1$, $\alpha_2 = 6$ and $\varkappa_2 = -3$, $\gamma = 3$ and $\lambda = -2$. Thus

$$H\Phi: \quad u = x_0^{-1} x_1 + 6 x_0^{-3} x_2 \; ,$$

$$\Psi G: \quad u = 3 x_0^{-2} (x_1 + x_2) \; .$$

For what independent micro data vectors $x = \{x_0, x_1, x_2\}$ is the aggregation point-consistent?

When the aggregation is consistent, the two expressions for u are equal. Their difference vanishes,

$$x_0^{-2} (x_0 - 3) x_1 + 3 x_0^{-3} (2 - x_0) x_2 = 0 \; .$$

For what admissible vectors x does this equation hold? For any given x_0, it is the equation of a straight line through the origin in the (x_1, x_2) plane, say

$$w_1 x_1 + w_2 x_2 = 0 ,$$

where w_1 and w_2 are functions of x_0. The signs of w_1 and w_2 are as follows.

$0 < x_0 < 2$	$w_1 < 0$	$w_2 > 0$,
$x_0 = 2$	$w_1 < 0$	$w_2 = 0$,
$2 < x_0 < 3$	$w_1 < 0$	$w_2 < 0$,
$x_0 = 3$	$w_1 = 0$	$w_2 < 0$,
$3 < x_0$	$w_1 > 0$	$w_2 < 0$.

Thus if $2 \leq x_0 \leq 3$, the line does not pass inside the first quadrant.

A necessary and sufficient condition for point-consistency is that $x_0 < 2$ or $x_0 > 3$ while (x_1, x_2) is a point on a certain straight line, which is different for different values of x_0.

Example 1.2.1.2. Again, $q = 2$, $g_1 = g_2 = 1$, $\beta_1 = \beta_2 = 1$ and $\delta = 1$. Now, $\alpha_1 = 1$ and $\varkappa_1 = -1$, $\alpha_2 = 1$ and $\varkappa_2 = -3$, $\gamma = 1$ and $\lambda = -2$. This,

$$H\Phi: \quad u = x_0^{-1} x_1 + x_0^{-3} x_2 ,$$

$$\Psi G: \quad u = x_0^{-2} (x_1 + x_2) .$$

For what independent micro data vectors $x = \{x_0, x_1, x_2\}$ is the aggregation point-consistent?

As in the preceding example, consistency implies an equation,

$$x_0^{-2} (x_0 - 1) x_1 + x_0^{-3} (1 - x_0) x_2 = 0 .$$

If the equation is denoted $w_1 x_1 + w_2 x_2 = 0$, the signs of w_1 and w_2 are as follows.

$0 < x_0 < 1$	$w_1 < 0$	$w_2 > 0$,
$x_0 = 1$	$w_1 = 0$	$w_2 = 0$,
$1 < x_0$	$w_1 > 0$	$w_2 < 0$.

When $x_0 = 1$, the equation does not define a line in the (x_1, x_2) plane, but is satisfied by any (x_1, x_2).

A necessary and sufficient condition for point-consistency is that either $x_0 = 1$ while (x_1, x_2) is arbitrary, or else $x_1 \neq 1$ while (x_1, x_2) is a point on a certain straight line, which is different for different values of x_0.

Example 1.2.1.3. Tentatively, the whole class of aggregations is considered. What members of the class are point-consistent for all admissible independent micro data vectors x ?

Assume consistency. Then the two expressions for u are equal, and the equation

$$\sum_{h=1}^{q} \alpha_h x_0^{\varkappa_h - \lambda} x_h^{\beta_h} = \gamma \left(\sum_{h=1}^{q} g_h x_h \right)^{\delta}$$

holds identically in x. The RHS does not vary with x_0. If there is an h such that $\varkappa_h > \lambda$, the LHS increases infinitely as $x_0 \to \infty$. If there is an h such that $\varkappa_h < \lambda$, the LHS increases infinitely as $x_0 \to 0$. Thus $\varkappa_h = \lambda$ for every h.

The remaining equation is differentiated w.r.t. x_i where i is one h. This trick and the following reasoning are inspired from Green [1964]. The resulting equation

$$\alpha_i \beta_i x_i^{\beta_i - 1} = \gamma \delta g_i \left(\sum_{h=1}^{q} g_h x_h \right)^{\delta - 1}$$

holds identically in x. Since the LHS does not vary with any x_h whose $h \neq i$, neither does the RHS. Thus $\delta = 1$. But then the RHS does not vary with x_i, and so neither does the LHS. Thus $\beta_i = 1$, and consequently $\alpha_i = \gamma g_i$. The argument is repeated for $i = 1, \ldots, q$.

The aggregation is point-consistent for all admissible x if and only if

$$\varkappa_1 = \varkappa_2 = \ldots\ldots = \varkappa_q = \lambda ,$$
$$\beta_1 = \beta_2 = \ldots\ldots = \beta_q = \delta = 1 ,$$
$$\alpha_h = \gamma g_h , \quad h = 1, \ldots, q .$$

Interpreted in economic terms, all micro and macro price elasticities are equal mutually, all micro and macro income elasticities are equal mutually and to unity, and the micro incomes are weighted in the aggregation proportionately to certain micro parameters.

Example 1.2.1.4. Again, the whole class of aggregations is tentatively

considered. Now, the independent micro data vector x is subject to the restrictions

$$x_h = \xi_h t , \quad h = 1, \ldots, q ,$$

where ξ_h are given positive constants. What aggregations are point-consistent for all positive values of x_0 and t ?

Assume consistency. The two expressions for u are equal, and the equation

$$\sum_{h=1}^{q} \alpha_h \xi_h^{\beta_h} x_0^{\varkappa_h - \lambda} t^{\beta_h - \delta} = \gamma \left(\sum_{h=1}^{q} g_h \xi_h \right)^{\delta}$$

holds identically in (x_0, t). The RHS does not vary with x_0 or t. As in the preceding example, this implies that $\varkappa_h = \lambda$ for every h. By an analogous argument, it also implies that $\beta_h = \delta$ for every h . Finally, the equation

$$\sum_{h=1}^{q} \alpha_h \xi_h^{\delta} = \gamma \left(\sum_{h=1}^{q} g_h \xi_h \right)^{\delta}$$

is also implied.

The aggregation is point-consistent for all admissible vectors x, where the elements $x_1, \ldots, x_q$ maintain fixed proportions, if and only if

$$\varkappa_1 = \varkappa_2 = \ldots\ldots = \varkappa_q = \lambda ,$$
$$\beta_1 = \beta_2 = \ldots\ldots = \beta_q = \delta ,$$
$$\Sigma_h \alpha_h \xi_h^{\delta} = \gamma \left(\Sigma_h g_h \xi_h \right)^{\delta} .$$

In economic terms, all micro and macro price elasticities are equal, and similarly all micro and macro income elasticities. The last condition is better not verbalized.

1.2.2 Two basic types of consistency problems

Most aggregation problems concern a class of aggregations. Often, each of the four vector functions G, H, Φ, Ψ is specified only as a class of functions, whose members are distinguished by different numerical values for certain parameters. All the parameters of G, H, Φ, Ψ together form the total parameter vector p of the class of aggregations.

Problems concerning consistency also involve the independent micro data vector x. The elements of x are regarded as variables, and may be subject to restrictions.

The beginning of 1.2.1 provides an example, if q is fixed. The total parameter vector is

$$p = \{g_1, \ldots, g_q ; \alpha_1, \varkappa_1, \beta_1, \ldots, \alpha_q, \varkappa_q, \beta_q ; \gamma, \lambda, \delta\} .$$

The aggregating function H contributes no parameters in this case. The independent micro data vector x and the restrictions on p and x were specified in 1.2.1.

Consider a class of aggregations, i.e. a given set Π of total parameter vectors p, and a given set Ξ of independent micro data vectors x. The question of point-consistency can be raised for every pair of one aggregation $p \in \Pi$ and one data vector $x \in \Xi$. For every pair (p, x) the verdict is either "point-consistent" or "point-inconsistent".

Consider a subset Π_i of the set of aggregations Π. Those data vectors $x \in \Xi$ for which every aggregation $p \in \Pi_i$ is point-consistent, form a subset of Ξ, perhaps the null set, which will be denoted $\Xi(\Pi_i)$.

Similarly, consider a subset Ξ_i of the set of data vectors Ξ. Those aggregations $p \in \Pi$ which are point-consistent for every data vector $x \in \Xi_i$, form a subset of Π, perhaps the null set, which will be denoted $\Pi(\Xi_i)$.

The above formal apparatus is inspired by similar ideas in Alstadheim [1968].

There are two basic types of aggregation problems concerned with consistency. The first type of problems selects a set Π_i of aggregations and asks what is the corresponding set $\Xi(\Pi_i)$ of data vectors. Problems of this kind will be called direct consistency problems. The second type of problems selects a set Ξ_i of data vectors and asks what is the corresponding set $\Pi(\Xi_i)$ of aggregations. Problems of this kind will be called indirect consistency problems.

Example 1.2.1.1 is a direct consistency problem. The set Π_1 has a single member. The set $\Xi(\Pi_1)$ can be visualized as part of the locus of those points in the (x_0, x_1, x_2) space that satisfy the equation $h(\varphi(x)) = \psi(g(x))$. The locus is a curved two-dimensional surface. The part inside the positive orthant is retained. The solution has not much intuitive appeal. Loosely speaking, this is typical of direct consistency problems.

Example 1.2.1.2 is quite analogous to the preceding example. Nevertheless its solution $\Xi(\Pi_2)$ is qualitatively different. It consists of a surface like that just described, plus the plane $x_0 = 1$, both restricted to the positive orthant. The "additional" part of $\Xi(\Pi_2)$ is due to what can perhaps be called a singularity in the equation $h(\varphi(x)) = \psi(g(x))$.

Example 1.2.1.3 is an indirect consistency problem. The set Ξ_3 consists of the entire positive orthant. The solution described the set $\Pi(\Xi_3)$ by means of $3q+1$ independent restrictions. The corresponding micro and macro relations are

$$\Phi: \quad y_h = \gamma g_h x_0^{\lambda} x_h , \quad h = 1, \ldots, q ,$$

$$\Psi: \quad u = \gamma z_0^{\lambda} z_1$$

Interpreted in economic terms, this is a set of demand functions which is too restricted to be of much interest. Loosely speaking, this is typical of indirect consistency problems.

Example 1.2.1.4 is another indirect consistency problem. The set Ξ_4 is a subset of Ξ_3. By definition $\Xi_4 \subset \Xi_3$ implies $\Pi(\Xi_4) \supseteq \Pi(\Xi_3)$. The set $\Pi(\Xi_4)$ is less restricted than $\Pi(\Xi_3)$. The solution describes it by means of $2q+1$ independent restrictions. The corresponding micro and macro relations are

$$\Phi: \quad y_h = \alpha_h x_0^{\lambda} x_h^{\delta} , \quad h = 1, \ldots, q ,$$

$$\Psi: \quad u = \gamma z_0^{\lambda} z_1^{\delta} , \text{ where } \gamma = \Sigma_h \alpha_h \xi_h^{\delta} / (\Sigma_h g_h \xi_h)^{\delta} .$$

Compared with the preceding example, the restriction $\delta = 1$ has gone, and the proportionality requirements for α_h and g_h have been replaced by a single restriction.

Direct and indirect consistency problems do not form an exhaustive classification of all aggregation problems directly concerned with consistency.

1.2.3 Another concept of consistency

In the econometric literature, indirect consistency problems are sometimes formulated a little differently in two respects. An example is Green [1964], ch. 5.

First, the set Ξ_i of independent micro data vectors is not specified explicitly. Instead, it is tacitly understood that x is not restricted except as required by the economic interpretation. For example, any non-negative number is admissible as a price.

Second, the term "consistency" is not used in the point-wise sense of 1.1.4 or 1.1.6. Instead, an aggregation p is called consistent if and only if $p \in \Pi(\Xi_i)$. In other word, p is consistent if and only if it is point-consistent for every $x \in \Xi_i$.

In the terminology indicated, the question asked in an indirect consistency problem is simply: What aggregations p are consistent?

The terminology described implies a modified concept of consistency. In 1.1, consistency was defined as a possible property of the pair (p, x) of one aggregation and one independent micro data vector. According to the new terminology, consistency is a possible property of the pair (p, Ξ_i) of one aggregation and one (understood) set of independent micro data vectors.

The concept of consistency is now redefined as follows.

A set of independent micro data vectors Ξ is assumed to be given. An aggregation $\{G, H, \Phi, \Psi\}$, where Φ and Ψ are deterministic vector functions, is said to be consistent if and only if

$$h(\varphi(x)) = \psi(g(x)) \quad \text{for every } x \in \Xi .$$

If Φ, Ψ are stochastic vector functions, the aggregation is said to be (expectationally) consistent if and only if

$$E_{H\Phi}(u) = E_{\Psi G}(u) \quad \text{for every } x \in \Xi .$$

An analogous concept of distributional consistency relative to a set could also be defined.

The above concept of expectational consistency was formulated independently by Hannan [1972] and by Lütjohann [1970a].

1.2.4 Articulated consistency problems

Indirect consistency problems as described in 1.2.2 ask a single question about the total parameter vector p as a whole. By the definition of consistency

in 1.2.3, the question is: For what vectors p is there consistency?

Given an indirect consistency problem, the analysis can be brought further. The total parameter vector can be partitioned in some given way $p = (p_1, p_2)$, and different questions can be asked about the two subvectors. First, for what subvectors p_1 is consistency attainable by suitable selection of p_2? Second, if p_1 is such as to admit consistency, how should the subvector p_2 be chosen so as to achieve it? Problems of this more refined kind will be called articulated consistency problems. The elements of the subvector p_1 will be called the critical parameters. The elements of the subvector p_2 will be called the discretionary parameters.

For example, consider again the indirect consistency problem of example 1.2.1.3. A possible background is as follows. An investigator assumes that the micro relation Φ is valid, although he may not know its parameter values. He aggregates by the aggregating functions H and G, where he is willing to accept any weights g_h that turn out to be helpful. He would like the macro data to satisfy a macro relation of the form Ψ. The parameters γ, λ, δ of Ψ may take whatever values are required by the circumstances.

Given this background, it is natural for the investigator to consider the following articulated consistency problem. The critical parameters p_1 are the parameters of Φ. The first question is: What restrictions, if any, must be imposed on the parameters of Φ for consistency to be at all attainable? The discretionary parameters are the parameters of Ψ and G. The second question is: If the parameters of Φ admit consistency, how should those of Ψ and G be chosen to achieve it?

The solution of the indirect consistency problem in example 1.2.1.3 provides answers to both questions. First, consistency requires of Φ that $\alpha_h = \overline{\alpha}$ and $\beta_h = 1$, $h = 1, \ldots, q$, where $\overline{\alpha}$ is some negative number. Second, if these requirements are fulfilled, in order to fit Φ the parameters of Ψ and G must be chosen so that $\lambda = \overline{\alpha}$, $\delta = 1$, and $\gamma g_h = \alpha_h$, $h = 1, \ldots, q$.

Another possible background for example 1.2.1.3 is as follows. An investigator assumes that the macro relation Ψ is valid for specified parameter values, perhaps unknown. He also assumes that the aggregating functions H and G are valid, where $g_h = 1$ for all h. He would like to know whether a micro relation of the form Φ is consistent with these assumptions.

Given this alternative background, the appropriate articulated consistency problem is as follows. The critical parameters are those of Ψ. The discretionary parameters p_2 are those of Φ. The parameters of G are fixed, and are therefore omitted from the total parameter vector p.

Again, example 1.2.1.3 provides the answers. First, in order for consistency to be attainable, the parameters of the macro relation must satisfy the single restriction $\delta = 1$. Second, if $\delta = 1$, in order to achieve consistency, the parameters of the micro relation must be chosen to be $\alpha_h = \gamma$, $\varkappa_h = \lambda$, $\beta_h = 1$, $h = 1, \ldots, q$.

A particular class of articulated consistency problems will be introduced in 4.2. On the terminology in 1.2, see 1.4.2.

1.2.5 The classical Econometrica debate

One source of aggregation problems in economics is the wish to establish an explicit connection between micro-economic and macro-economic theories. In a classical debate in Econometrica, it was discussed what kind of consistency problems ought to be considered in this context. The following brief review does not cover the specific economic argumentation.

One idea was proposed by Klein [1946a]. Micro-economic theories should be accepted as given. Macro-economic theories should be postulated, but the definition of the aggregates involved should be left open. "Then construct aggregates which are consistent with the two theories."

Formally, Klein takes as given the micro and macro relations Φ and Ψ, and tries to find consistent aggregating functions G and H. He seems to intend the independent micro data vector set Ξ to be essentially unrestricted.

Another idea was proposed by May [1946]. A particular micro-economic theory should not be studied isolated from the total micro-economic context, say a general equilibrium system. The other relations of the complete model may limit the freedom of the variables involved in the particular theory. Macro-economic aggregates should be defined. Then, if the number of degrees of freedom left is appropriate, "the functions of the simplified model are derived as functionals of the functions of the general model."

Formally, May takes as given the micro relation Φ, the aggregating functions

G and H, and the restricted independent micro data vector set Ξ implied by, let us say, the micro-economy at large. If these givens are appropriate, the macro relation Ψ can be derived mathematically from them.

Yet another idea was proposed by Shou Shan Pu [1946]. The distribution among the micro-economic entities, say firms, of a quantity to be aggregated, say a factor of production, should not be assumed to be free to vary arbitrarily. It may be economically reasonable to assume that there exists a pattern to which the distribution always adheres, for example because the firms operate under perfect competition. Macro-economic aggregates should be defined. Then, "as long as there are any definite relations that determine the pattern of distribution, a unique aggregate production function can be formulated."

Formally, the approach of Shou Shan Pu agrees with that of May.

Shou Shan Pu objected to Klein's approach that it could well lead to unnatural macro-economic aggregates, for example geometric averages.

The ideas of May and Shou Shan Pu were criticized by Klein [1946b] for making Ψ depend on Ξ. The macro relation Ψ is thus determined not only by the micro relation Φ but also, indirectly, by other relations and conditions of the total micro-economic system. But "the aggregate production function should not depend upon profit maximization, but purely on technological factors."

In defense of his approach, May [1947] argued i.a. that Klein's requirement, that a production function ought to be purely technological, is not reasonable even in micro-economics.

Somewhat less sketchy reviews of the classical debate are given in Nataf [1962] and in Alstadheim [1968].

1.3 Problems of interpretation

1.3.1. Three examples

The linear models assumed in regression analysis are a special kind of stochastic vector functions. In the following three examples of aggregation, the micro and macro relations consist of one or two such models. Regression coefficients are computed from the given set of micro and macro data for $n \geq 2$

observations. (Formally, this anticipates chapter 2, but the examples are very simple.)

There are six data vectors

$$
\begin{aligned}
x' &= \{X_1, \ldots, X_n\}, \\
x'' &= \{X_{11}, \ldots, X_{1n}; X_{21}, \ldots, X_{2n}\}, \\
y' &= \{Y_1, \ldots, Y_n\}, \\
y'' &= \{Y_{11}, \ldots, Y_{1n}; Y_{21}, \ldots, Y_{2n}\}, \\
z &= \{Z_1, \ldots, Z_n\}, \\
u &= \{U_1, \ldots, U_n\}.
\end{aligned}
$$

Each example involves either x' or x'', either y' or y''.

Averages are defined as follows,

$$\overline{X}_2 = \frac{1}{n} \sum_{j=1}^{n} X_{2j} ,$$

and similarly for $\overline{X}$, $\overline{X}_1$, $\overline{Y}$, $\overline{Y}_1$, $\overline{Y}_2$, $\overline{Z}$ and $\overline{U}$.

Deviations are defined as follows,

$$x_{2j} = X_{2j} - \overline{X}_2 , \quad j = 1, \ldots, n,$$

and similarly for x_j, x_{1j}, y_j, y_{1j}, y_{2j}, z_j and x_j.

In the examples, any summation is over j from 1 to n. Further, the denominator of any computed regression coefficient is assumed not to vanish.

Example 1.3.1.1 involves x', y'', z and u. The four vector functions are as follows, where in each case $j = 1, \ldots, n$.

$$
\begin{aligned}
G: &\quad Z_j = X_j ; \; z_j = x_j , \\
H: &\quad U_j = Y_{1j} + Y_{2j} ; \; u_j = y_{1j} + y_{2j} , \\
\Phi: &\quad \begin{cases} Y_{1j} = \alpha_1 + \beta_1 X_j + \epsilon_{1j}, & E(\epsilon_{1j}) = 0, \\ Y_{2j} = \alpha_2 + \beta_2 X_j + \epsilon_{2j}, & E(\epsilon_{2j}) = 0, \end{cases} \\
\Psi: &\quad U_j = \gamma + \delta Z_j + \eta_j , \quad E(\eta_j) = 0.
\end{aligned}
$$

The probability distributions of y given x and of u given z are only incompletely specified.

From the macro data, the regression of U upon Z is computed. The regression coefficient d corresponding to the parameter δ is

$$d = \frac{\Sigma z_j u_j}{\Sigma z_j^2}$$

From the micro data, the regressions of Y_1 and Y_2 upon X are computed. The regression coefficients b_1 and b_2 corresponding to the parameters β_1 and β_2 are

$$b_1 = \frac{\Sigma x_j y_{1j}}{\Sigma x_j^2} \quad ; \quad b_2 = \frac{\Sigma x_j y_{2j}}{\Sigma x_j^2} \quad .$$

By G and H, the relation

$$d = b_1 + b_2$$

follows immediately.

Example 1.3.1.2 involves x'', y', z and u. The four vector functions are as follows, where in each case $j = 1, \ldots, n$.

$$\begin{aligned}
&G: && Z_j = X_{1j} + X_{2j} \; ; \; z_j = x_{1j} + x_{2j} \, , \\
&H: && U_j = Y_j \; ; \; u_j = y_j \, , \\
&\Phi: && Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \epsilon_j \, , \; E(\epsilon_j) = 0 , \\
&\Psi: && U_j = \gamma + \delta Z_j + \eta_j \, , \; E(\eta_j) = 0 .
\end{aligned}$$

From the macro data, the regression of U upon Z is computed. The regression coefficient d is as in the preceding example.

From the micro data, the regression of Y upon X_1 and X_2 is computed. The regression coefficients b_1 and b_2 corresponding to the parameters β_1 and β_2 are solved from the following pair of normal equations.

$$\begin{aligned}
(\Sigma x_{1j}^2) b_1 + (\Sigma x_{1j} x_{2j}) b_2 &= \Sigma x_{1j} y_j \, , \\
(\Sigma x_{2j} x_{1j}) b_1 + (\Sigma x_{2j}^2) b_2 &= \Sigma x_{2j} y_j \, .
\end{aligned}$$

By G and H, addition of the two normal equations produces

$$(\Sigma z_j x_{1j}) b_1 + (\Sigma z_j x_{2j}) b_2 = \Sigma z_j u_j \, .$$

When the LHS is substituted in the numerator of d, the relation

$$d = \left(\frac{\Sigma z_j x_{1j}}{\Sigma z_j^2} \right) b_1 + \left(\frac{\Sigma z_j x_{2j}}{\Sigma z_j^2} \right) b_2$$

follows immediately.

Example 1.3.1.3 involves x'', y'', z and u. The four vector functions are as follows, where in each case $j = 1, \ldots, n$.

$$G: \quad Z_j = X_{1j} + X_{2j} \; ; \; z_j = x_{1j} + x_{2j} \; ,$$

$$H: \quad U_j = Y_{1j} + Y_{2j} \; ; \; u_j = y_{1j} + y_{2j} \; ,$$

$$\Phi: \begin{cases} Y_{1j} = \alpha_1 + \beta_1 X_{1j} + \varepsilon_{1j} \; , \; E(\varepsilon_{1j}) = 0 \; , \\ Y_{2j} = \alpha_2 + \beta_2 X_{2j} + \varepsilon_{2j} \; , \; E(\varepsilon_{2j}) = 0 \; , \end{cases}$$

$$\Psi: \quad U_j = \gamma + \delta Z_j + \eta_j \; , \; E(\eta_j) = 0$$

The stochastic specification of the micro relation implies for $j = 1, \ldots, n$

$$\Phi^*: \begin{cases} E(y_{1j}) = \beta_1 x_{1j} \; , \\ E(y_{2j}) = \beta_2 x_{2j} \end{cases}$$

From the macro data, the regression of U upon Z is computed. The regression coefficient d is as in the preceding examples.

By the aggregating function H,

$$d = \frac{\Sigma z_j y_{1j}}{\Sigma z_j^2} + \frac{\Sigma z_j y_{2j}}{\Sigma z_j^2}$$

By application of Φ^*, there follows the relation

$$E_\Phi(d) = \left(\frac{\Sigma z_j x_{1j}}{\Sigma z_j^2} \right) \beta_1 + \left(\frac{\Sigma z_j x_{2j}}{\Sigma z_j^2} \right) \beta_2 \; ,$$

where E_Φ denotes expectation according to Φ.

1.3.2 Interpreting macro statistics in micro terms

A potential source of aggregation problems is the following situation. A theory is to be investigated empirically. Data that give sufficiently detailed information are not available, perhaps for reasons of cost. Instead, there are less detailed

data. The investigator must try to extract information from the data available, even though that information is incomplete.

In favourable cases, the situation can be formalized. A class of micro relations Φ is given or assumed, but the actual parameter values are not known. Micro data (x, y) are not available. A set of macro data (z, u) is available, and the aggregating functions G and H are known. From the macro data there is computed a vector of macro statistics

$$T(z, u).$$

The macro statistics are intended to shed light on the parameters of the micro relation Φ, or at least on the micro data (x, y).

Precisely what micro information do the mactro statistics T (z, u) give? This quite open-ended question will be called the interpretation problem for the aggregation and statistics in question.

The interpretation problem makes no reference to the macro relation Ψ of the aggregation. Thus, no macro relation need be assumed. If one is assumed, its only role is to serve as a motivation for the macro statistics T (z, u).

The three examples of 1.3.1 illustrate different kinds of answers to the interpretation problem. In all three examples, there is a single macro statistic, the regression coefficient $T(z, u) = d$.

In examples 1.3.1.1 and 1.3.1.2 there is considered also a vector of micro statistics

$$t(x, y)$$

that could be computed from the micro data, if these were available. There are in these examples two micro statistics, the regression coefficients $t(x, y) = (b_1, b_2)$. The macro statistic T (z, u) is interpreted in terms of the micro statistics t (x, y). The interpretation makes no reference to the micro relation Φ of the aggregation. Thus, no micro relation need be assumed. If one is assumed, as in these examples, its only role is to serve as a motivation for the micro statistics t (x, y).

In the first two examples, the interpretation problem is answered by a relation between the macro and micro statistics. The relation is not based on the macro and micro relations, only on the aggregating functions G and H. Such

answers to the interpretation problem will be called model-free relations.

In example 1.3.1.1, the model-free relation is $d = b_1 + b_2$. No unknown quantities are involved except the micro statistics. Loosely speaking, this is atypical.

In example 1.3.1.2, the model-free relation can be formulated

$$d = w_1 b_1 + w_2 b_2 ,$$

where w_1 and w_2 are functions of the independent micro data x, and are thus unknown. But whatever the micro data, it is known that

$$w_1 + w_2 = 1 ,$$

although w_1 or w_2 may be negative. Thus in a generalized sense, d is a weighted average of b_1 and b_2. This may be a useful interpretation if approximate values of w_1 and w_2 can be guessed.

In example 1.3.1.3, the interpretation problem is answered by a relation of a different kind. Its LHS is the expectation, according to the micro relation, of the macro statistics,

$$E_{\Phi}(T(z, u)) .$$

Its RHS involves the parameters of the micro relation Φ. The relation is based on the aggregating functions G and H and on the micro relation Φ. Such answers to the interpretation problem will be called expectational relations.

The expectational relation in example 1.3.1.3 is

$$E_{\Phi}(d) = w_1 \beta_1 + w_2 \beta_2 ,$$

where the weights w_1 and w_2 are as in the preceding example.

The model-free relation in example 1.3.1.2 implies a similar expectational relation, since by the micro relation $E(b_i) = \beta_i$, $i = 1, 2$.

The expectational relation in example 1.3.1.3 does not imply a similar model-free relation. Here it is not the case that $d = w_1 b_1 + w_2 b_2$, where b_1 and b_2 are the simple regression coefficients corresponding to β_1 and β_2.

A particular class of model-free relations will be introduced in 4.3, and a related class of expectational relations in 4.4. On the terminology in 1.3, see 1.4.2.

1.3.3 The macro parameters derived by Theil

Perhaps the most natural approach to aggregation is to look for conditions under which there is consistency. More often than not, the conditions are found to be unrealistically restrictive.

In his pioneering work on linear aggregation of economic relations, Theil [1954] took a more constructive approach. That approach will now be described, using the concepts and terminology that have been introduced earlier in this chapter.

The situation considered is of the kind where the interpretation problem is relevant. The macro statistics $T(z, u)$ are regression coefficients d_h, $h = 1, \ldots, q$. The micro relation Φ consists of one or more linear (regression) models with parameters β_i, $i = 1, \ldots, p$, where p is the total number of such parameters. The aggregating functions G and H are linear.

Theil's approach is in two steps. The first step is to answer the interpretation problem by means of expectational relations of the form

$$E_{\Phi}(d_h) = \sum_{i=1}^{p} w_{hi} \beta_i , \quad h = 1, \ldots, q ,$$

where w_{hi} are certain functions of the independent micro data x. Example 1.3.1.3 illustrates the first step.

The second step of Theil's approach is to postulate a macro relation Ψ which is a linear (regression) model with parameters δ_h, and to identify the expected macro regression coefficients with these parameters. "These estimates are, as usual, postulated to belong to certain macroparameters; and these are the parameters in which we are interested. More precisely, we interpret the macroparameters as the expectations of their estimates." (Theil [1954], moment 2.2.1.) Thus,

$$\delta_h = \sum_{i=1}^{p} w_{hi} \beta_i , \quad h = 1, \ldots, q .$$

In example 1.3.1.3, Theil would write $\delta = w_1 \beta_1 + w_2 \beta_2$.

In summary, Theil proceeds as follows. To begin with, there is an incomplete aggreation $\{G, H, \Phi\}$. Then a macro relation Ψ is postulated, and its parameters are derived from the given G, H and Φ. The result is a complete aggregation $\{G, H, \Phi, \Psi\}$.

It turns out that the complete aggregation thus established is not expectationally consistent. The semi-aggregated micro relation $H\Phi$ and the semi-disaggregated macro relation ΨG give, in general, different forecasts of the dependent macro data vector for a given independent micro data vector x. "Suppose, for instance, that for some λ all microvariables $x_{\lambda i}$ increase by one unit, so that the macrovariable x_{λ} must increase by I units. If we assume all other exogenous variables and the disturbances to be constant, we can deduce from the macroequation that y must be increase by β_{λ} I ..... units. On the other hand, the microtheory will tell us that under these assumptions each microvariable y_i will increase by $\beta_{\lambda i}$ units, so that the macrovariable y must increase by $\overline{\beta}_{\lambda}$ I units. This is clearly not identical with the prediction according to the macrotheory. So we see that there may be contradictions between conclusions from the macrotheory and those from the microtheory." (Theil [1954], moment 2.3.4.)

The purpose of the formal analysis later in this study is to try to sort out this confusing situation. Conclusions will be stated in 8.1 and 8.2.

1.4 Some further topics and references

1.4.1 Seven topics not covered in this study

This study does not attempt to cover the whole field of aggregation theory and applications. Seven of the more important omissions are listed below.

First, index-number theory is a part of the theory of aggregation. This is so whether the approach is that of I. Fisher [1922], that of Konyus [1924; 1939], or that of Theil [1960].

Second, the concept of functional separability. Economists sometimes wish to simplify differentiable functions by aggregating the arguments into a smaller number of aggregate arguments, each representing a subset of the original arguments. A necessary condition for consistency is that the original function has the property of functional separability. The concept was formulated independently by Sono [1945; 1961] and by Leontief [1947]; see also Morishima [1961]. Different kinds of functional separability have later been distinguished, as reviewed by Goldman and Uzawa [1964], and by W. Fisher [1969], section 4.3 and appendix E.

Third, the decision-theoretic approach to aggregation. Aggregation is sometimes unavoidable, perhaps for reasons of cost, although consistency is unattainable. The practical problem is then perhaps how to aggregate, and certainly how to use the macro data, so as to minimize the loss due to aggregation. The concepts of decision theory are applicable. How they can be applied was shown by Malinvaud [1956], section VI. This approach has been pursued by W. Fisher [1969], who gives further references; also i.a. by Schneeweiss [1965].

Fourth, what may perhaps be called aggregation by integration. In the examples of 1.2.1 and 1.3.1, the aggregation was by summation over a finite number of entities. Sometimes, instead, the aggregation is by integration over a statistical distribution, which is usually continuous. Examples are de Wolff [1941], and the "stratification approach" in market demand theory described by Wold [1952], section 7.4.

Fifth, the applications in economic theory are numerous. Some are in demand theory, e.g. Rajaoja [1958]. Many are concerned with production functions and the concept of capital, e.g. Solow [1956]. Some are in input-output analysis, e.g. Hatanaka [1952]. Aggregation is reviewed from the economist's point of view by Green [1964], who gives many references.

Sixth, there are applications in sociology and related behavioural sciences. Many are concerned with ecological correlation, e.g. Robinson [1950]; cf. 8.3.3 below. Aggregation is reviewed from the sociologist's point of view by Hannan [1971], who gives many references. The methodological traditions of economists and sociologists differ. The shift of perspective between Green [1964] and Hannan [1971] is notable.

Seventh, the somewhat unstructured topic or set of topics labelled classification, clustering, or taxonomy. A general reference giving many further references is Cormack [1971].

1.4.2 On the terminology introduced

The terms "micro/macro", "aggregation/disaggregation", "micro/macro relation", and "consistency" are fairly current usage.

The terms "segregated aggregation", "semi-aggregated micro / semi-disaggregated macro relation", "point-consistency", "expectational consistency",

"consistency problem", "direct/indirect/articulated consistency problem", "critical/discretionary parameter", "interpretation problem", and "model-free/expectational relation" are all innovations.

Instead of "micro" and "macro", Malinvaud [1956] and W. Fisher [1969] use detailed and simplified.

Instead of "aggregating function for independent/dependent data", Ijiri [1971] uses active/passive aggregation function.

Instead of "semi-aggregated micro / semi-disaggregated macro relation", Ijiri [1971] uses principal/surrogate function.

Instead of "expectational consistency", Theil [1954] uses perfection, and Hannan [1972] stochastic consistency.

A few more alternative terms will be mentioned in 1.4.3.

1.4.3 Some general references to literature

In the literature on aggregation, the following five works as outstanding.

Theil [1954] considers micro and macro relations that consist of linear (regression) models, and linear aggregation. Several different types of aggregation are distinguished; cf. 3.2.5 below. The primary purpose is to express the parameters of the macro relation in terms of those of the micro relation; cf. 1.3.3. A secondary purpose is to find out how to achieve perfect aggregation, i.e. expectational consistency. The results are discussed in economic terms, but the emphasis is on statistical concepts and computations.

Malinvaud [1956] is primarily concerned with the logic of aggregation and aggregation problems. Aggregation is defined as the representation of a detailed model by a simplified model. The use of the simplified model may involve disaggregating functions. A concept nearly equivalent to consistency is formulated as follows: Sometimes one can find intrinsic aggregates and an intrinsic (simplified) model such that operation with these involves no loss whatever of information essential to the user. A concept of representative aggregates is formulated to cover i.a. the representation of a great number of micro data by a statistical distribution. Finally, it is shown in detail how the problem of optimal aggregation can be formulated as a problem in decision theory. Malinvaud's exposition is strongly recommended.

Green [1964] studies a number of types of aggregation occurring in economics. The micro relation is given by economic theory. The set of independent micro data vectors admitted may be restricted by economic theory. Aggregating functions and macro relation must be chosen from certain classes. All functional relations are assumed to be deterministic and differentiable. The primary purpose is to find conditions for consistency. Except for a chapter on Theil's approach, aggregation is studied as a branch of mathematical economics.

W. Fisher [1969] studies aggregation as a branch of decision theory. The primary purpose is to find the optimal simplification. Consistency, called exact simplification, is but a rare limiting case. The problem formulation involves a loss function. A number of concrete numerical problems of widely different kinds are reformulated and studied as problems of optimal simplification.

Ijiri [1971] distinguishes four types of queries in aggregation theory. Two are concerned with different consistency concepts. The third type is concerned with the errors caused by aggregation, the fourth type with the selection of an optimal aggregation. The exposition covers a very wide field in a somewhat abstract manner.

Particularly numerous references to the literature on aggregation are given in Green [1964], W. Fisher [1969], Ijiri [1971], and Hannan [1971]. Some further references are Bentzel [1956], Thionet [1960], Nataf [1964], Lancaster [1966], Thionet [1967], Alstadheim [1968], F. Fisher [1969], Moriguchi [1970], Zellner and Montmarquette [1971], Pokropp [1972], Lütjohann [1972], and Wu [1973].

2 SOME ALGEBRA AND THEORY OF LEAST-SQUARES

2.1 Some matrix algebra

2.1.1 Some concepts in matrix algebra

Matrix algebra will be used to express the theory of generalized least-squares regression. A few notions of vector geometry will also be referred to.

Matrices will be denoted by capital letters like M, column vectors by lower-case letters like v. A matrix of r rows and c columns will be called an $r \times c$ matrix. A (column) vector of n elements will be called an n-vector. The transposes of M and v will be denoted M' and v'. The inverse of a square non-singular matrix will be denoted M^{-1}. The unit matrix of order n will be denoted I or I_n. The n-vector, each element of which is unity, will be denoted j or j_n, or sometimes i.

The following definitions and simple properties will be referred to without explicit quotation. A general reference is Graybill [1969].

An $n \times n$ matrix Q is positive-definite if and only if it is symmetric and $v'Qv > 0$ for any n-vector $v \neq 0$. A positive-definite matrix Q has an inverse Q^{-1} which is positive-definite.

Each positive-definite $n \times n$ matrix Q defines variants of the concepts of distance and orthogonality as follows. Let v_1 and v_2 be two n-vectors. The positive square root of the non-negative quantity

$$(v_1 - v_2)' Q (v_1 - v_2)$$

will be called the Q-distance between v_1 and v_2. Further, v_1 and v_2 will be called Q-orthogonal if and only if

$$v_1' Q v_2 = 0 .$$

Ordinary vector-geometric distance and orthogonality are the special variants where $Q = I_n$.

The covariance matrix of a set of non-degenerate random variables that are not linearly dependent, is positive-definite. (See e.g. Wilks [1962], section 3.5.)

A set S of n-vectors is a vector space if and only if it has the following property. If $v_1 \in S$ and $v_2 \in S$, then $a_1 v_1 + a_2 v_2 \in S$ for any two scalars a_1 and a_2.

Consider a vector space S of n-vectors. The set of all n-vectors v such that $w'v = 0$ for all $w \in S$ is another vector space, the orthogonal complement $S^{\perp}$ of S. The orthogonal complement of $S^{\perp}$ is S.

Let M be a given $n \times m$ matrix. The set of all n-vectors v such that $v = Mt$ for some m-vector t is a vector space, the column space of M. The set of all m-vectors w such that $w' = s'M$ for some n-vector s is also a vector space, the row space of M. The set of all m-vectors h such that $Mh = 0$ is a vector space too, the null space of M.

The row space of M and the null space of M are each other's orthogonal complements. Each determines the other uniquely.

Let M be a given $n \times m$ matrix, v a given n-vector, and h an m-vector of unknowns. The equation system $Mh = v$ has a solution if and only if v is in the column space of M.

2.1.2 Some propositions in matrix algebra

The following propositions will be useful later in the chapter.

Proposition 2.1.2.A

Let Q be a given positive-definite $n \times n$ matrix. Let M be a given $n \times m$ matrix.

The three matrices of m columns

$$M,\ QM,\ M'QM$$

have identical row spaces. ⚬

If $Mt = 0$, then $QMt = 0$. If $QMt = 0$, then $M'QMt = 0$. If $M'QMt = 0$, then $t'M'QMt = 0$, and then, since Q is positive-definite, $Mt = 0$. Thus the three matrices have identical null spaces. Consequently they have identical row spaces. □

Proposition 2.1.2.B

Let M be a given $n \times m$ matrix, v a given n-vector, and h an m-vector of unknowns. Let M and v be such that the equation system

$$Mh = v$$

is consistent. Let λ be a given m-vector.

A necessary and sufficient condition for the quantity $\lambda'h$ to be uniquely deter-

mined by the equation system is that λ is in the row space of M. ⁘

Let h_o be a particular solution, i.e. $Mh_o = v$.

Necessity. Assume that $\lambda'h$ is uniquely determined. Then $\lambda'(h_o + d) = \lambda'h_o$ for all d such that $M(h_o + d) = Mh_o$. Otherwise expressed, $\lambda'd = 0$ for all d such that $Md = 0$. Thus λ is orthogonal to the null space of M. Consequently λ is in the row space of M.

Sufficiency. Assume that λ is in the row space of M, i.e. there is an n-vector s such that $\lambda' = s'M$. Then for any solution h of the equation system, $\lambda'h = s'Mh = s'v = s'Mh_o = \lambda'h_o$. Thus $\lambda'h$ is uniquely determined by the equation system. □

Proposition 2.1.2.C

Let Q be a positive-definite $n \times n$ matrix. Let M be a given $n \times m$ matrix and v a given n-vector. Let h be an m-vector of unknowns.

The equation system

$$M'QMh = M'Qv$$

has a solution. ⁘

The vector M'Qv is in the column space of M'Q. Thus by P.2.1.2.A it is in the column space of M'QM. Consequently the equation system has a solution. □

Statisticians will no doubt interpret P.2.1.2.C in terms of normal equations. P.2.1.2.A and P.2.1.2.B are also likely to remind statisticians of certain well-known theorems on least-squares regression. The formulations have been chosen so as to bring out the purely algebraical content of those theorems. This is done in the interest of the distinction that will be made, between model-free properties in 2.2 and properties based on the linear model in 2.3 and 2.4.

The three propositions with their demonstrations were built from material found in Graybill [1961], Rao [1965], Seber [1966], and Searle [1971]. Detailed references are hardly necessary, and would be quite tedious.

2.2 Model-free properties of least-squares regression

2.2.1 The pattern of regression data

Linear regression analysis can be regarded as a systematic comparison of units of analysis. The comparison over a given set of units of analysis.

The units of analysis are observed with respect to a common set of regression variables. One of these, the regressand, is to be "explained" by a linear combination of the other regression variables, the regressors.

For each combination of one unit of analysis and one regression variable there is one datum. The total set of regression data can be partitioned into the regressand data and the regressor data.

The regression data can be arranged into a matrix, where each row is associated with one unit of analysis, and each column with one regression variable. The data matrix will be denoted

$$[y \mid X]$$

where y is the regressand vector and X is the regressor matrix. Each column of X is a regressor vector.

More often than not, the first regressor vector represents a dummy variable, the intercept regressor, which is equal to unity for every unit of analysis.

The term "unit of analysis" is taken from Blalock [1964].

2.2.2 Regression data versus observational data

The regression data need not be the data originally observed. Usually, the observational data can also be arranged into a matrix, where each row is associated with one unit of observation, and each column with one observational variable. One of these, the dependent variable, is to be "explained" in terms of the other observational variables, the independent variables. The intercept regressor is not usually counted as an observational variable.

An example will illustrate the distinction between observational data and regression data. The demand for cheese in a country is to be "explained" in terms of population, income and price. The units of observation are the twenty years 1951, ..., 1970. The dependent variable is the quantity of cheese sold Q. There

are three independent variables, the population N, the disposable income M, and a price index for cheese P. The observational data form a 20×4 matrix.

The demand function assumed is constant-elastic in per capita variables,

$$\left(\frac{Q}{N}\right)_t = \alpha \left(\frac{M}{N}\right)_t^{\beta} P_t^{\gamma} , \quad t = 1951, \ldots, 1970 .$$

The function is made linear in the parameters by taking logarithms. Then, a strong positive residual autocorrelation is eliminated by differencing. The transformed demand function is

$$Y_t = \beta X_{1t} + \gamma X_{2t} , \quad t = 1952, \ldots, 1970 ;$$

$$Y_t = \log \left(\frac{Q}{N}\right)_t - \log \left(\frac{Q}{N}\right)_{t-1} ,$$

$$X_{1t} = \log \left(\frac{M}{N}\right)_t - \log \left(\frac{M}{N}\right)_{t-1} ,$$

$$X_{2t} = \log P_t - \log P_{t-1} .$$

This is the relation fitted by linear regression analysis.

The units of analysis are, strictly speaking, the nineteen pairs of years $(1951,1952), \ldots, (1969,1970)$. Conventionally, the nineteen years $1952, \ldots, 1970$ are regarded as the units of analysis. The regressand is the differenced logarithmic per capita demand quantity Y. There are two regressors, the differenced logarithmic per capita income X_1 and the differenced logarithmic cheese price index X_2. In this example, there is no intercept regressor. The regression data form a 19×3 data matrix.

Henceforth, the terms "data" and "variable" refer to regression data and variables.

2.2.3 Least-squares approximation

Consider a given $n \times p$ regressor matrix X and a given regressand n-vector y. Any p-vector h defines a linear combination Xh of the regressor vectors. If that n-vector is used as an approximation of the regressand vector, the approximation errors committed are the elements of the vector $y - Xh$.

The purpose of model-free linear regression is to find that linear approximation Xh which is in some sense best. Each positive-definite $n \times n$ matrix Q

defines a variant of generalized least-squares regression, say GLS(Q)-regression for short. GLS(Q)-regression chooses that weight vector h which minimizes the Q-distance between the regressand vector y and its approximation Xh. Ordinary least-squares is the special variant where $Q = I_n$.

As is well known, GLS(Q)-regression leads to the system

$$X'QXh = X'Qy$$

of normal equations.

Proposition 2.2.3

Let Q be a given positive-definite n × n matrix. Let X be a given n × p regressor matrix and y a given regressand n-vector. Consider GLS(Q)-regression.

(i) The normal equations are satisfied by a least one p-vector h = b.

(ii) Let b satisfy the normal equations. Then

$$X'Q(y - Xb) = 0 .$$

In words, the approximation error vector y - Xb is Q-orthogonal to the regressor vectors.

(iii) Let b satisfy the normal equations, while h is any other p-vector. Then

$$(y - Xh)'Q(y - Xh) \geq (y - Xb)'Q(y - Xb) .$$

In words, the approximation Xh is not better than Xb in terms of Q-distance. ○○

(i) This follows from P.2.1.2.C.

(ii) This follows immediately from the normal equations.

(iii) Decompose $y - Xh = (y - Xb) + X(b - h)$. By (ii),

$$(y - Xh)'Q(y - Xh) = (y - Xb)'Q(y - Xb) + (b - h)'X'QX(b - h).$$

Since Q is positive-definite, the last term is non-negative. □

Any p-vector satisfying the normal equations will be denoted b. The normal equations will be written

$$X'QXb = X'Qy ,$$

that is with b instead of h. Each element of a solution vector b is associated with one regressor, and is called a regression coefficient. The intercept is the regression coefficient associated with the intercept regressor.

Any vector of regression coefficients b defines two n-vectors

$$\hat{y} = Xb \quad \text{and} \quad e = y - Xb ,$$

the approximation vector and the residual vector, respectively.

2.2.4 Unique coefficient functions

Consider again the model-free GLS(Q)-regression of y upon X, where the order of X is $n \times p$. Any p-vector λ defines a linear coefficient function $\lambda'b$ of the regression coefficients b. A coefficient function $\lambda'b$ will be called unique if and only if the system of normal equations $X'QXb = X'Qy$ determines it uniquely.

Proposition 2.2.4.A

Let Q be a given positive-definite $n \times n$ matrix. Let X be a given $n \times p$ regressor matrix and y a given regressand n-vector. Consider GLS(Q)-regression. Let λ be a p-vector.

(i) A necessary and sufficient condition for the coefficient function $\lambda'b$ to be unique is that λ is in the row space of X.

(ii) The regression determines the approximation vector $\hat{y}$ and the residual vector e uniquely. ◦◦

(i) By P.2.1.2.B, a necessary and sufficient condition for $\lambda'b$ to be uniquely determined by the normal equations is that λ is in the row space of $X'QX$. By P.2.1.2.A, the row spaces of $X'QX$ and of X are identical.

(ii) Let x_j' be the j'th row vector of X. By (i), the coefficient functions $x_j'b$, $j = 1, \ldots, n$, are unique. Thus $\hat{y} = Xb$ and $e = y - Xb$ are uniquely determined. □

Whether a coefficient function $\lambda'b$ is unique depends on λ and the regressor matrix X only. It does not depend on the variant Q of GLS-regression, nor on the regressand vector y.

If the rank of X is p, the normal equations have the single solution

$$b = (X'QX)^{-1} X'Qy ,$$

and all coefficient functions are unique.

The following proposition will be useful later in 2.4 and 4.3.

Proposition 2.2.4.B

Let Q be a given positive-definite $n \times n$ matrix. Let X be a given $n \times p$

regressor matrix. Let y be a variable regressand n-vector. Consider GLS(Q)-regression. Let λ be a given p-vector such that the coefficient function $\lambda'b$ is unique.

(i) There exists at least one p-vector r such that

$$X'QXr = \lambda .$$

(ii) The n-vector

$$\tau = QXr$$

is uniquely determined by the equation system $X'QXr = \lambda$.

(iii) The coefficient function $\lambda'b$ is the linear function

$$\lambda'b = \tau'y$$

of the variable regressand vector y. ⁘

(i) Since $\lambda'b$ is unique, by P.2.1.2.B λ is in the row space of X'QX.

(ii) By P.2.1.2.B a quantity $k'r$ is uniquely determined if the p-vector k is in the row space of X'QX. By P.2.1.2.A each row vector of QX is in the row space of X'QX.

(iii) By (i), the normal equations, and (ii),

$$\lambda'b = r'X'QXb = r'X'Qy = \tau'y . \quad \square$$

Consider fixed Q, X and λ such that $\lambda'b$ is unique. Then the coefficient function $\lambda'b$ is the fixed-weight linear function $\tau'y$ of the regressand vector y.

2.2.5 Some references to literature

Model-free properties of regression are clearly distinguished from properties based on the linear model in Goldberger [1964], section 4.2, for full-rank ordinary least-squares. A study devoted exclusively to model-free properties is Leander [1963]. Even the standard errors of the regression coefficients are given a model-free interpretation by Goldberger [1968].

As is well known, GLS can be reduced to ordinary least-squares by a suitable linear transformation of the units of analysis. An alternative is to proceed as in P.2.2.3. This approach is cited also by Theil [1971], section 6.1.

The standard approach to least-squares with possibly less than full rank seems

to be to assume the linear model and proceed in terms of estimability. Model-free properties are then not easily distinguished. When all references to a model were weeded out from Graybill [1961], section 11.2, the material remaining was sufficient to build up P.2.2.4.A and P.2.2.4.B. Another source of inspiration was Rao [1965], section 4a.

The concept of a generalized inverse matrix is used in least-squares theory by i.a. Rao [1965], Pringle and Rayner [1971], and Searle [1971]. The consequent gain in elegance will be forgone in this study.

2.3 The linear model and the regression model

2.3.1 The incomplete and complete linear models

Least-squares regression is often motivated as a method of estimation. The regression data $[\, y \mid X \,]$ are then assumed to have been generated in accordance with a probabilistic model. One alternative is known as the Linear Model, another one as the Regression Model. The latter will be described in 2.3.4.

The linear model involves the following six components.

(i) X is an $n \times p$ regressor matrix.

(ii) y is an $n \times 1$ regressand vector.

(iii) β is a $p \times 1$ parameter vector.

(iv) ϵ is an $n \times 1$ disturbance vector.

(v) W is an $n \times n$ standardized covariance matrix.

(vi) σ^2 is a scalar average disturbance variance.

The incomplete linear model consists of equations (1) and (2) and the explanations following them.

(1) $y = X\beta + \epsilon$.

(2) $E(\epsilon) = 0$.

The regressor matrix X is non-stochastic and known. It is identical with the regressor data matrix actually observed. The parameter vector β is non-stochastic and unknown. The disturbance vector ϵ is stochastic and unobservable. The regressand vector y of the model is stochastic. The regressand data vector

actually observed is one realization of it.

The complete linear model adds equation (3) and the explanations following it.

(3) $\quad E(\varepsilon \varepsilon') = \sigma^2 W.$

The standardized covariance matrix W is positive-definite, has (for example) tr (W) = n, and is known. The average disturbance variance σ^2 is unknown.

The incomplete linear model amounts to a partical specification of the first order moments of the distribution of the regressand vector y. The expected regressand vector is some vector $E(y) = X\beta$ in the column space of the regressor matrix.

The complete linear model adds a partial specification of the second order moments of the distribution of y. The covariance matrix of y is proportional to the known matrix W. A special case is when $W = I_n$.

The complete linear model is the traditional Linear Model. The incomplete linear model is sufficient for the purposes of the formal analysis of chapters 4 to 8.

2.3.2. Identified parameter functions

The linear model can be reformulated in terms of a seventh component.

(vii) μ is an $n \times 1$ mean vector.

Equations (1) and (2) in 2.3.1 are replaced by the following two equations.

(1*) $\quad \mu = X\beta.$

(2*) $\quad E(y) = \mu.$

The mean vector μ is non-stochastic. It is known to belong to the column space of X but is otherwise unknown.

When least-squares regression is used as a method of estimation, the purpose is to estimate, as far as possible, the parameter vector β. The relation between β and the stochastic data vector y is in two steps. First, by (1*) the parameter vector β determines the mean vector μ. Second, by (2*) the mean vector μ partly determines the distribution of the regressand vector y.

The second step (2*) can be reversed in the sense that μ can be estimated from y. In fact, the approximation vector $\hat{y}$ is an unbiased estimator of μ. But

the first step (1*) cannot always be reversed in the sence that if μ were known, then β would be known too. Several different parameter vectors β may correspond to a given mean vector μ. Granted that β does not influence the distribution of y except through μ, several different vectors β are then observationally equivalent. There is this a problem of identification.

The concept of identification is well known to econometricians since Koopmans [1953]. It is usually formulated and studied for models of the type known as Simultaneous Equations (see e.g. Johnston [1963], section 9-2). The parameters β and the means μ in the linear model correspond to the structural and reduced-form parameters, respectively, in the simultaneous equations model. That the concept of identification is applicable in this way to the linear model is indicated in passing by Seber [1966], section 3.4.

Any p-vector λ defines a linear parameter function $\lambda'\beta$ of the parameters β of the linear model. A parameter function $\lambda'\beta$ will be called identified if and only if the consistent equation system $X\beta = \mu$ determines it uniquely.

Proposition 2.3.2

Let X be a given $n \times p$ matrix. Let μ be a given n-vector such that $\mu = X\beta$ for at least one p-vector β. Assume a linear model where X is the regressor matrix and μ is the mean vector. Let λ be a p-vector.

A necessary and sufficient condition for the parameter function $\lambda'\beta$ to be identified is that λ is in the row space of X. ⁸₈

This follows from P.2.1.2.B. □

Whether a parameter function $\lambda'\beta$ is identified depends on λ and the regressor matrix X only. It does not depend on the standardized covariance matrix W of the linear model. The linear model may even be incomplete.

If the rank of X is p, the equation system (1*) has the single solution

$$\beta = (X'QX)^{-1}X'Q\mu,$$

where Q is an arbitrary positive-definite $n \times n$ matrix. Then all parameter functions are identified.

2.3.3 Designed and observed regressor data

The elements of the regressor matrix X can be of two kinds. Some have no empirical content. They are artifacts used in the construction of a linear model to fit given assumptions. Such regressor data will be called designed. Others record an observation of the real world. If the real world were sufficiently different, they would be numerically different. Such regressor data will be called observed.

The vectors of the regressor matrix X can be of three kinds. If all the elements of a regressor vector are designed data, that regressor will be called designed. If all the elements of a regressor vector are different observed data, that regressor will be called observed. The intercept regressor is designed. All regressors used in the analysis of variance are designed. Most regressors used in regression analysis are observed. Regressors that are neither designed nor purely observed will be called mixed.

The following example illustrates the concepts. The macro-economic consumption function makes consumption C a linear function of gross national product P. Data are available for half-years indexed t = 1, 2, 3, ... The relation assumed is as follows.

$$
\begin{aligned}
&C_t = \alpha + \delta P_t + \varepsilon_t && \text{when t is odd,}\\
&C_t = \varkappa + \theta P_t + \varepsilon_t && \text{when t is even,}\\
&E(\varepsilon_t) = 0,\\
&E(\varepsilon_t \varepsilon_{t+s}) = \sigma^2 \rho^s && \text{for } s = 0, 1, 2, \ldots
\end{aligned}
$$

The autocorrelation is $\rho = 0.8$ approximately.

For reasons of efficiency (cf. Zellner [1962a]) the two lines are estimated jointly by GLS regression. The linear model $y = X\beta + \varepsilon$ is constructed as follows.

$$
\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & P_1 & 0 \\ 1 & 1 & P_2 & P_2 \\ 1 & 0 & P_3 & 0 \\ 1 & 1 & P_4 & P_4 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \alpha \\ \varkappa - \alpha \\ \delta \\ \theta - \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \vdots \end{bmatrix}
$$

There are equivalent alternatives.

The first regressor in the example is the intercept regressor. The second regressor is a dummy variable indicating the even-numbered half-years. Both these regressors are designed. The third regressor records gross national product. This regressor is observed. The fourth regressor is mixed.

The even-numbered elements of the fourth, mixed, regressor are replicates of the corresonding elements of the third regressor. Arrangements with dummy variables and mixed regressors usually involve such replication of observed regressor data.

The distinction between designed and observed regressor data is relevant to the formulation of a consistency problem in 4.2. It will also be referred to in the discussion of one type of aggregation in 3.1.5.

2.3.4 The regression model

The linear model treats any element of the regressor matrix as a known parameter. For designed regressor data this is natural, e.g. in the analysis of variance. It is less natural for observed regressor data, which are thought of as variable. The linear model can be modified as follows, so as to take account of the variability of the regressor matrix.

The regression model is like the linear model except for the characterization of X. The regressor matrix X of the regression model is stochastic. The regressor data matrix actually given, at least partly from observation, is one realization of it. The distribution of the disturbance vector ε is independent of that of X. Thus equations (1), (2) and (3) of 2.3.1 hold for any realized X.

The formal analysis in chapters 4 to 8 will be based on linear models. Regression models will occur in 9.2.

2.4 Least-squares estimation in the linear model

2.4.1 Estimable parameter functions

Consider a linear model where the order of the regressor matrix X is $n \times p$. Consider a given linear function $\lambda'\beta$ of the p parameters of the model. Any n-vector t defines a linear function $t'y$ of the regressand data, a linear estimator.

The parameter function $\lambda'\beta$ will be called unbiasedly estimable if and only if there exists an estimator $t'y$ such that $E(t'y) = \lambda'\beta$ identically in β.

Proposition 2.4.1.A

Let X be a given $n \times p$ matrix. Let β be a given p-vector. Assume a linear model where X is the regressor matrix and β is the parameter vector. Let λ be a p-vector.

A necessary and sufficient condition for the parameter function $\lambda'\beta$ to be estimable is that λ is in the row space of X. ○○

Necessity. Assume that $\lambda'\beta$ is estimable. Then for a certain n-vector t, $E(t'y) = t'X\beta = \lambda'\beta$ identically in β. Thus $t'X = \lambda'$, i.e. λ is in the row space of X.

Sufficiency. Assume that $\lambda' = t'X$ for some n-vector t. Then $E(t'y) = t'X\beta =$ $= \lambda'\beta$ identically in β. □

The above proposition is well known in least-squares theory. It holds even if the linear model assumed is incomplete.

The next proposition states a connection between the concepts of uniqueness, identification, and estimability.

Proposition 2.4.1.B

Let X be a given $n \times p$ matrix. Let λ be a p-vector.

Let y be any n-vector. Let Q be any positive-definite $n \times n$ matrix. Let b be a p-vector of unknowns. Consider the normal equation system $X'QXb = X'Qy$. Call this system NE.

Let β be any p-vector. Consider any linear model where X is the regressor matrix and β the parameter vector. Call this model LM.

The model LM may be incomplete. No relation is assumed to hold between Q and the covariance matrix of LM.

Consider the following three statements.

(i) The coefficient function $\lambda'b$ in NE is unique.

(ii) The parameter function $\lambda'\beta$ in LM is identified.

(iii) The parameter function $\lambda'\beta$ in LM is estimable.

The three statements are either all true or all false. ○○

By P.2.2.4.A, P.2.3.2, and P.2.4.1.A, a necessary and sufficient condition for each of the three statements to be true is that λ is in the row space of X. □

The concept of uniqueness does not refer to a model The concepts of identification and estimability do refer to a linear model, which may be incomplete. For a given regressor matrix X, the three concepts are co-extensive in the sense that $\lambda'b$ is unique if and only if $\lambda'\beta$ is identified, and if and only if $\lambda'\beta$ is estimable.

Unless the rank of X is p, there are coefficient functions $\lambda'b$ that are not unique and parameter functions $\lambda'\beta$ that are not identified (or estimable). A remedy sometimes applied is to add suitable linear constraints $R\beta = 0$ and $Rb = 0$. This approach is discussed i.a. in Scheffé [1959] and Searle [1971].

2.4.2 Best linear unbiased estimation

Assume the complete linear model. Consider a parameter function $\lambda'\beta$.

Any estimator $t'y$ will be called an unbiased estimator of $\lambda'\beta$ if and only if $E(t'y) = \lambda'\beta$ identically in β.

An unbiased estimator $t_o'y$ of $\lambda'\beta$ will be called a best linear unbiased estimator, abbreviated BLUE, of $\lambda'\beta$ if and only if it has the following property: For any other unbiased linear estimator $t'y$ of $\lambda'\beta$, $\mathrm{Var}(t'y) \geq \mathrm{Var}(t_o'y)$.

Only if $\lambda'\beta$ is estimable does it make sense to look for a BLUE. The answer turns out to involve the GLS(W^{-1})-regression of y upon X, where W is the standardized covariance matrix of the linear model.

Proposition 2.4.2

Assume the complete linear model (1), (2), (3). Consider the normal equations (4). The only stochastic components are ε, y and b.

(1) $y = X\beta + \varepsilon$,

(2) $E(\varepsilon) = 0$,

(3) $E(\varepsilon\varepsilon') = \sigma^2 W$

(4) $X'W^{-1}Xb = X'W^{-1}y$.

Let $\lambda'\beta$ be an estimable (identified) parameter function. Consider the analogous coefficient function $\lambda'b$.

(i) $\lambda'b = \tau'y$, where $\tau = W^{-1}Xr$,
where r is any vector such that $X'W^{-1}Xr = \lambda$;
and τ is determined uniquely by λ, X and W.

(ii) $E(\lambda'b) = \lambda'\beta$.

(iii) $\lambda'b$ is the only BLUE of $\lambda'\beta$. ⁸

(i) Since $\lambda'\beta$ is estimable, by P.2.4.1.B $\lambda'b$ is a unique coefficient function in (4). Then P.2.2.4.B is applicable with $Q = W^{-1}$.

(ii) By (i) and the linear model, $E(\lambda'b) = \tau'X\beta = r'X'W^{-1}X\beta = \lambda'\beta$.

(iii) Let $t'y = (\tau + d)'y$ be an unbiased estimator of $\lambda'\beta$. Then by (ii) and the linear model, $d'X\beta = 0$ identically in β. Thus $d'X = 0$, which by (i) implies $d'W\tau = 0$. Hence and by the linear model, $\mathrm{Var}(t'y) = \sigma^2 t'Wt = \sigma^2\tau'W\tau + \sigma^2 d'Wd$. Since W is positive-definite, the variance is minimized if and only if $d = 0$. □

The above is a version of the well-known Gauss-Markov theorem on least-squares. It says that $\lambda'b$, which may be called the least-squares estimator of $\lambda'\beta$, is the BLUE of $\lambda'\beta$, given two conditions. First, $\lambda'\beta$ must be estimable. Second, the GLS variant must be that defined by W^{-1}.

The standard least-squares theory developed in 2.4 is too well-known for detailed references to be required. The presentation chosen owes most to Graybill [1961]. Other sources of inspiration were Rao [1952], Rao [1965], and Searle [1971].

2.5 Regression in deviation form

2.5.1 Means and deviations

In generalized least-squares regression involving an intercept, certain generalized means and deviations are useful. Before these are introduced, notation will be slightly modified. Throughout, j is an n-vector of unit elements, $q = p - 1$ and Q is an arbitrary positive-definite $n \times n$ matrix.

The regressand n-vector is still written y. The $n \times p$ regressor matrix, earlier written X, is partitioned and re-written

$$[j \mid X],$$

where j is the intercept regressor vector, while X is the remaining n × q regressor submatrix. The p-vector of regression coefficients, earlier written b, is partitioned conformably, and the approximation vector is written

$$\hat{y} = ja + Xb ,$$

where a is the regression intercept, while b is the remaining subvector of q regression coefficients.

The scalar $\bar{y}$ and the row q-vector $\bar{x}'$ defined as follows

$$(\bar{y} \mid \bar{x}') = (j'Qj)^{-1} j'Q [\, y \mid X \,]$$

will be called the regressand Q-mean and the vector of regressor Q-means. The regressand n-vector and the n × q regressor matrix of Q-deviations are defined as follows,

$$[\, y - j\bar{y} \mid X - j\bar{x}' \,] .$$

Ordinary means and deviations are the special variants where $Q = I_n$.

The regression data $[\, y - j\bar{y} \mid X - j\bar{x}' \,]$ will be said to be the regression data $[\, y \mid X \,]$ in deviation form. The deviation form data can be produced as follows

$$[\, y - j\bar{y} \mid X - j\bar{x}' \,] = D [\, y \mid X \,],$$

where D is the n × n deviation-producing matrix

$$D = I_n - j(j'Qj)^{-1} j'Q .$$

Because $j'QD = 0$, any vector of data in Q-deviation form is Q-orthogonal to the intercept regressor vector. Thus any vector of Q-deviation data has zero Q-mean.

The deviation-producing matrix D for $Q = I_n$ is formulated and studied in Kloek [1961].

2.5.2 The deviation form normal equations

Let y be a regressand n-vector and X an n × q regressor matrix. Let Q be a positive-definite n × n matrix. Let $\bar{y}$ and $\bar{x}'$ be the Q-means of y and x .

The GLS(Q)-regression of y upon X and the intercept regressor will be called the raw form regression. The GLS(Q)-regression of $(y - j\bar{y})$ upon $(X - j\bar{x}')$ with no intercept added will be called the corresponding deviation form regression.

The normal equations of the raw form regression are

$$\text{Ra: } j'Qja + j'QXb = j'Qy ,$$

$$\text{Rb: } X'Qja + X'QXb = X'Qy ,$$

where Ra is a single equation, while Rb is the remaining subsystem of q equations.

The normal equations of the deviation form regression are

$$\text{Db: } (X - j\bar{x}')'Q(X - j\bar{x}')b = (X - j\bar{x}')'Q(y - j\bar{y}) .$$

In order to determine the intercept, they are complemented by

$$\text{Da: } a = \bar{y} - \bar{x}'b ,$$

a well-known computing formula.

The content of the following proposition is well known, at least for full rank ordinary least-squares regression.

Proposition 2.5.2.

Consider the GLS(Q)-regression of y upon X and the intercept regressor, the raw form regression, whose normal equation system is Ra and Rb. Consider also the corresponding deviation form regression, whose complemented normal equation system is Da and Db.

The equation systems (Ra, Rb) and (Da, Db) are equivalent. ∞

The equation system (Ra, Rb) can be written

$$\begin{bmatrix} j'Qj & j'QX \\ X'Qj & X'QX \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} j'Qy \\ X'Qy \end{bmatrix}$$

Premultiplication by the $p \times p$ matrix

$$\begin{bmatrix} (j'Qj)^{-1} & O \\ -\bar{x} & I_q \end{bmatrix}$$

produces the equation system

$$\begin{bmatrix} 1 & \bar{x}' \\ (X - j\bar{x}')'Qj & (X - j\bar{x}')'QX \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \bar{y} \\ (X - j\bar{x}')'Qy \end{bmatrix}$$

Since $(X - j\bar{x}')'Qj = 0$, this is equivalent to

$$\begin{bmatrix} 1 & \bar{x}' \\ 0 & (X - j\bar{x}')' Q (X - j\bar{x}') \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \bar{y} \\ (X - j\bar{x}')' Q (y - j\bar{y}) \end{bmatrix}$$

This is the equation system (Da, Db).

Since $j'Q(X - j\bar{x}') = 0$ and $j'Q(y - j\bar{y}) = 0$, the equation system (Da, Db) is equivalent to

$$\begin{bmatrix} 1 & \bar{x}' \\ 0 & X'Q(X - j\bar{x}') \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \bar{y} \\ X'Q(y - j\bar{y}) \end{bmatrix}$$

Premultiplication by the $p \times p$ matrix

$$\begin{bmatrix} j'Qj & O \\ X'Qj & I_q \end{bmatrix}$$

produces the equation system

$$\begin{bmatrix} j'Qj & j'QX \\ X'Qj & X'QX \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} j'Qy \\ X'Qy \end{bmatrix}$$

This is the equation system (Ra, Rb). □

A unique coefficient function in the raw form regression can be computed from any solution of the complemented normal equations of the corresponding deviation form regression.

When the intercept (or, expressed differently, its regressor) is eliminated from (Ra, Rb), the outcome is Db. More regressors can be eliminated step-wise in an analogous manner. For full rank ordinary least-squares, this is discussed in Lütjohann [1970b].

2.5.3 The coefficient of determination

Consider again a raw form GLS(Q)-regression and the corresponding deviation form regression. By P.2.2.4.A, both produce unique approximation and residual vectors. By P.2.5.2., the two residual vectors are equal, for

$$(y - j\bar{y}) - (X - j\bar{x}')b = y - j(\bar{y} - \bar{x}'b) - Xb = y - ja - Xb .$$

The deviation form regression decomposes its regressand vector into an approximation vector and a residual vector as follows.

$$(y - j\bar{y}) = (X - j\bar{x}')b + e .$$

By P.2.2.3, e is Q-orthogonal to each of the regressor vectors $(X - j\bar{x}')$. Consequently, the deviation form regression decomposes the total sum of squares as follows.

$$(y - j\bar{y})' Q (y - j\bar{y}) = b' (X - j\bar{x}')' Q (X - j\bar{x}')b + e'Qe$$

The two parts are called the sum of squares due to regression and the residual sum of squares, respectively. Since Q is positive-definite, each of the three terms is non-negative.

The coefficient of determination R^2 of the deviation form regression is defined as follows.

$$R^2 = 1 - \frac{e'Qe}{(y - j\bar{y})' Q (y - j\bar{y})} .$$

In the degenerate case where $y = j\bar{y}$, R^2 is not defined. By the decomposition of the sum of squares, $0 \leq R^2 \leq 1$.

The above defines the coefficient of determination for a deviation form regression. By convention, R^2 for a raw form regression is (usually) defined to be identical to R^2 for the corresponding deviation form regression. The positive square root of R^2 is called the coefficient of multiple correlation between the raw form regressand y and the set of raw form regressors X.

The coefficient of determination and related subjects are discussed at greater length e.g. by Theil [1971], chapter 4. The outline given will be sufficient for present purposes. The explicit generalization above to GLS(Q)-regression was formulated independently by Buse [1973] and by Lütjohann [1970a].

2.5.4 A decomposition of a moment matrix

In the analysis of variance of a completely randomized design, the total sum of squares is decomposed into one sum of squares between the treatments and another sum of squares within them. A multivariate generalization will now be given.

Consider an $n \times p$ data matrix X. The p variables can be regressors, regressands, or both. Let the set of n units of analysis, and the rows of X, be

partitioned into $m \geq 2$ disjoint, exhaustive subsets. The number of units in the h'th subset is denoted n_h, $h = 1, \ldots, m$.

Let Q be a positive-definite $n \times n$ matrix. Let Q be such that, if it is partitioned conformably with X, Q is block-diagonal.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_h \\ \vdots \\ X_m \end{bmatrix} \quad ; \quad Q = \begin{bmatrix} Q_1 & \cdots & O & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & Q_h & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & O & \cdots & Q_m \end{bmatrix}$$

Since Q is positive-definite, so is each Q_h.

Let j_n, j_m and i_h, $h = 1, \ldots, m$, be column vectors of n, m and n_h unit elements, respectively. Let K be the $n \times m$ matrix formed by arranging block-diagonally the vectors i_h.

$$K = \begin{bmatrix} i_1 & \cdots & O & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & i_h & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & O & \cdots & i_m \end{bmatrix} .$$

The relation $Kj_m = j_n$ follows from the definition.

Let $\overline{Q}$ be the $m \times m$ diagonal matrix defined as follows.

$$\overline{Q} = K'QK .$$

The h'th diagonal element of $\overline{Q}$ is $i_h' Q_h i_h > 0$, $h = 1, \ldots, m$.

The Q-means over the n units of analysis are the row p-vector

$$\bar{\bar{x}}' = (j_n' Q j_n)^{-1} j_n' QX .$$

The Q_h-means over the n_h units of analysis that form the h'th subset of units, are the row p-vector

$$\bar{x}_h' = (i_h' Q_h i_h)^{-1} i_h' Q_h X_h .$$

For $h = 1, \ldots, m$, $\bar{x}_h'$ is the h'th row vector of the $m \times p$ matrix

$$\bar{X} = (K'QK)^{-1} K'QX \ .$$

The matrix $\bar{X}$ can be regarded as a data matrix where each of the m subsets is represented by its Q_h-means. Since $j_n = Kj_m$, $K'QK = \bar{Q}$, and $K'QX = K'QK\bar{X}$,

$$\bar{\bar{x}}' = (j_m' \bar{Q} j_m)^{-1} j_m' \bar{Q} \bar{X} \ .$$

Thus the Q-means over the n units of analysis are also the $\bar{Q}$-means over the m subsets.

The data X in Q-deviation form are the $n \times p$ matrix

$$X - j_n \bar{\bar{x}}' \ .$$

The data X_h in Q_h-deviation form are the $n_h \times p$ matrix

$$X_h - i_h \bar{x}_h' \ .$$

For $h = 1, \ldots, m$, $X_h - i_h \bar{x}'$ is the h'th submatrix of the row-partitioned $n \times p$ matrix

$$X - K\bar{X} \ .$$

The subset data $\bar{X}$ in $\bar{Q}$-deviation form are the $m \times p$ matrix.

$$\bar{X} - j_m \bar{\bar{x}}' \ .$$

A related matrix of order $n \times p$ is

$$K(\bar{X} - j_m \bar{\bar{x}}') \ ,$$

where the data for each unit of analysis are replaced by those for the subset to which the unit belongs.

Since $Kj_m = j_n$, the deviation form data matrices are related as follows.

$$(X - j_n \bar{\bar{x}}') = (X - K\bar{X}) + K(\bar{X} - j_m \bar{\bar{x}}') \ .$$

Since $K'Q(X - K\bar{X}) = 0$ and $K'QK = \bar{Q}$, this implies the following decomposition of the $p \times p$ <u>total</u> moment matrix.

$$(X - j_n \bar{\bar{x}}')' Q (X - j_n \bar{\bar{x}}') =$$

$$= (X - K\bar{X})' Q (X - K\bar{X}) + (\bar{X} - j_m \bar{\bar{x}}')' \bar{Q} (\bar{X} - j_m \bar{\bar{x}}') \ .$$

The two components can be called the <u>within subsets</u> and <u>between subsets</u> moment matrices. Since Q is block-diagonal, the former can be further decomposed as follows.

$$(X - K\bar{X})' Q (X - K\bar{X}) = \sum_{h=1}^{m} (X_h - i_h \bar{x}_h')' Q_h (X_h - i_h \bar{x}_h') \ .$$

The h'th subcomponent is the moment matrix <u>within the h'th subset</u> of units of analysis.

<u>Proposition 2.5.4</u>.

Under the assumptions and in the notation of <u>2.5.4.</u>,

$$(X - j_n \bar{\bar{x}}')' Q (X - j_n \bar{\bar{x}}') = \sum_{h=1}^{m} (X_h - i_h \bar{x}_h')' Q_h (X_h - i_h \bar{x}_h') +$$

$$+ (\bar{X} - j_m \bar{\bar{x}}')' \bar{Q} (\bar{X} - j_m \bar{\bar{x}}') \ . \quad \circ\circ$$

The demonstration has been given in the text. □

The analysis of covariance in a completely randomized design uses the above decomposition of the total moment matrix. For the case where $p = 2$ and $Q = I_n$, this can be seen from Kendall [1948], section 24.28 et seqq.

3 FIVE TYPES OF AGGREGATION OF REGRESSION DATA

3.1. Four elementary types of aggregation

3.1.1 Aggregation of regression data

Regression analysis requires data that conform to a certain pattern, which was described in 2.2.1. Any set of data conforming to that pattern can be called a set of regression data. The pattern characterizing regression data can be said to have four "elements": the regressors, the regressand, the units of analysis, and the set of units of analysis.

Regression data, like other data, can be aggregated. Four elementary types of aggregation of regression data will be introduced in 3.1. Each of the four elementary types of aggregation affects one the four "elements" of the pattern of regression data, but leaves the other three "elements" unaffected, as far as logically possible. A fifth, more complex, type of aggregation of regression data will be introduced in 3.2.

The five types of aggregation are applied to micro data that form k sets of regression data, where for some types $k = 1$ while for others $k \geq 2$. The h'th set of micro regression data will be denoted

$$[\, y_h \,\vdots\, X_h \,] \ ,$$

where y_h is the h'th micro regressand vector and X_h is the h'th micro regressor matrix. (When $k = 1$, the index h is dropped.)

Each of the five types of aggregation produces macro data that form one set of regression data. The macro regression data will be denoted

$$[\, u \,\vdots\, Z \,] \ ,$$

where u is the macro regressand vector and Z is the macro regressor matrix.

The five types of aggregation are all segregated in the sense of 1.1.3. For brevity, the two aggregating functions are re-named. The aggregating function for independent data G, which relates Z to X_h, $h = 1, \ldots, k$, will be called the regressor aggregating function. The aggregating function for dependent data H, which relates u to y_h, $h = 1, \ldots, k$, will be called the regressand aggregating function. The two aggregating functions need not be linear.

The five types of aggregation are aggregations in the narrower sense indicated in 1.1.4. No explicit reference will be made to micro or macro relations. The following is however tacitly understood. If there is a micro relation Φ, then for any j and h it relates the j'th element of y_h to the j'th row vector of X_h. Similarly, if there is a macro relation Ψ, the it relates the j'th element of u to the j'th row vector of Z. The micro and macro relations, if any, need not be linear.

The five types of aggregation of regression data will not be defined rigorously. Rather, five characteristic patterns of data and relations will be indicated and given names.

3.1.2 Six related examples

Each of the four elementary types of aggregation of regression data will be introduced by means of an example. The fifth type of aggregation will be introduced by means of two examples. The six examples select micro data from a common pool of micro data.

Examples 3.1.2, micro data. Consider a fictitious cross-section study of how fertility in sheep is affected by more or less damp weather. Data are collected covering one gestation period, say from October to March.

The units of analysis are small geographical regions, say parishes.

There are two sets of units of analysis. One consists of all the rural parishes of all the counties of southern Scotland. The other one consists of all the rural parishes of all the counties of northern England.

There are four regressands. One is the number of black-faced ewe lambs born alive in the parish in Spring. The other three are the numbers of black-faced ram lambs, white-faced ewe lambs, and white-faced ram lambs born alive in the parish in Spring.

There are four regressors. One is the number of black-faced sheep alive in the parish after the Autumn slaughter. Another one is the number of white-faced sheep alive in the parish after the Autumn slaughter. The third regressor is the amount of rain fallen in the county to which the parish belongs, during the period of investigation. The fourth regressor is the amount of snow fallen in the county during the same period, converted into inches of rain.

The parishes are indexed j. The counties are indexed i. For any i, the set of integers P_i is such that $j \in P_i$ if and only if the j'th parish belongs to the i'th county. The sets P_S and P_E are such that $j \in P_S$ or $j \in P_E$ if and only if the j'th parish is in Scotland or in England, respectively. The sets C_S and C_E are such that $i \in C_S$ or $i \in C_E$ if and only if the i'th county belongs to Scotland or England, respectively.

The regressands are denoted y^{BF}, y^{BM}, y^{WF} and y^{WM}. The letters B, W, F and M indicate black-faced, white-faced, ewe (female) and ram (male) lambs, respectively. The regressors are denoted x^B, x^W, x^R and x^S. The letters B, W, R and S indicate black-faced sheep, white-faced sheep, rain and snow, respectively.

The rain and snow data are for counties. Thus if for a given integer i, $j_1 \in P_i$ and $j_2 \in P_i$, then x_j^R and x_j^S for $j = j_1$ are identical to those for $j = j_2$.

Consider the regression of the number of lambs upon the number of sheep, the amount of rain, and the amount of snow. If the regression is concerned with black-faced ewe lambs in southern Scotland, the regression data matrix consists of the row vectors

$$\left(y_j^{BF} \;\vdots\; 1 \quad x_j^B \quad x_j^R \quad x_j^S \right) \qquad \text{for all } j \in P_S .$$

There are eight such micro regression data matrices, one for each combination of black-faced or white-faced, female or male, Scotland or England.

In the examples, a macro datum that is identical to a single micro datum keeps its micro notation in the macro data.

Example 3.1.2.1, macro data. Consider the micro regression data matrix for black-faced ewe lambs in southern Scotland. Let the distinction between rain and snow be dropped. Then the two micro regressors x^R and x^S are replaced by a single macro regressor z^P, the amount of precipitation in the county to which the parish belongs. Formally,

$$z_j^P = x_j^R + x_j^S .$$

The row vectors

$$\left(y_j^{BF} \;\vdots\; 1 \quad x_j^B \quad z_j^P \right) \qquad \text{for all } j \in P_S$$

form a macro regression data matrix.

Example 3.1.2.2, macro data. Consider the two micro regression data matrices for black-faced ewe lambs and black-faced ram lambs in southern Scotland. These two regression data matrices have a common set of regressors but different regressands. Let the distinction between ewe lambs and ram lambs be dropped. Then the two micro regressands y^{BF} and y^{BM} are replaced by a single macro regressand u^{B}, the number of black-faced lambs born. Formally,

$$u_j^B = y_j^{BF} + y_j^{BM} \quad .$$

The row vectors

$$\left(u_j^B \;\vdots\; 1 \quad x_j^B \quad x_j^R \quad x_j^S \right) \qquad \text{for all } j \in P_S$$

form a macro regression data matrix.

Example 3.1.2.3, macro data. Consider the micro regression data matrix for black-faced ewe lambs in southern Scotland. Let the parishes be exchanged for the counties as units of analysis. The number of black-faced ewe lambs born in a county, u^{BF}, is the sum of the numbers of black-faced ewe lambs born in those parishes that constitute the county. Similarly, the number of black-faced sheep in a county, z^{B}, is a sum over the parishes of the county. Formally,

$$u_i^{BF} = \sum_{j \in P_i} y_j^{BF} \quad ,$$

$$z_i^B = \sum_{j \in P_i} x_j^B \quad .$$

The amounts of rain and snow fallen are for counties already. They are equal for all parishes within a given county. These data are retained unchanged, but notation is changed as follows.

$$z_i^R = x_j^R \qquad \text{for any } j \in P_i \quad ,$$

$$z_i^S = x_j^S \qquad \text{for any } j \in P_i \quad .$$

The row vectors

$$\left(u_i^{BF} \;\vdots\; 1 \quad z_i^B \quad z_i^R \quad z_i^S \right) \qquad \text{for all } i \in C_S$$

form a macro regression data matrix.

Example 3.1.2.4, macro data. Consider the two micro regression data matrices for black-faced ewe lambs in southern Scotland and in northern England. These two regression data matrices are for the same regressand and regressor variables but for different sets of units of analysis. Let the distinction between Scotland and England be dropped. Then the two sets of index values P_S and P_E are replaced by a single set, their union $P_S \cup P_E$. The row vectors

$$\left(y_j^{BF} \;\vdots\; 1 \quad x_j^B \quad x_j^R \quad x_j^S \right) \quad \text{for all } j \in (P_S \cup P_E)$$

form a macro regression data matrix.

Example 3.1.2.5, macro data. Consider the two micro regression data matrices for black-faced ewe lambs and white-faced ewe lambs in southern Scotland. These two regression data matrices are for the same units of analysis and for analogous regressand and regressor variables. Let the distinction between black-faced and white-faced ewe lambs and sheep be dropped. Then the two analogous regressands y^{BF} and y^{WF} are replaced by a single regressand u^F, the number of ewe lambs born. Similarly, the two analogous regressors x^B and x^W are replaced by a single regressor z^N, the number of sheep. Formally,

$$u_j^F = y_j^{BF} + y_j^{WF} \quad ,$$

$$z_j^N = x_j^B + x_j^W \quad .$$

The two regressors for rain and snow x^R and x^S are common to both micro regression data matrices. These data are retained unchanged. The row vectors

$$\left(u_j^F \;\vdots\; 1 \quad z_j^N \quad x_j^R \quad x_j^S \right) \quad \text{for all } j \in P_S$$

form a macro regression data matrix.

Example 3.1.2.6, macro data. Consider the four micro regression data matrices for black-faced ewe lambs, black-faced ram lambs, white-faced ewe lambs, and white-faced ram lambs in southern Scotland. These four regression data matrices are for the same units of analysis. Let the distinction between black-faced and white-faced lambs and sheep be dropped. Let the distinction between ewe lambs and ram lambs be dropped. Let the distinction between rain and snow

be dropped. Then the four micro regressands are replaced by a single macro regressand u, the total number of lambs born. Formally

$$u_j = y_j^{BF} + y_j^{BM} + y_j^{WF} + y_j^{WM} \quad .$$

Further, macro regressors z^N and z^P are formed as in examples 3.1.2.5 and 3.1.2.1, respectively. The row vectors

$$\left(u_j \;\vdots\; 1 \quad z_j^N \quad z_j^P \right) \qquad \text{for all } j \in P_s$$

form a macro regression data matrix.

3.1.3 Aggregation of regressors

Aggregation of regressors starts from a single micro regression data matrix $[\, y \,\vdots\, X \,]$. Macro regressors are formed as functions of the micro regressors. The regressand, the individual units of analysis, and the set of units of analysis, are not changed. The outcome is a macro regression data matrix $[\, u \,\vdots\, Z \,]$.

Example 3.1.2.1 is a case of linear aggregation of regressors. Each macro regressor is a linear function of some of the micro regressors.

Linear aggregation of p micro regressors into q macro regressors is described by the vector and matrix equations

$$u = y \;,$$
$$Z = XG \;,$$

where G is some $p \times q$ matrix of constants.

Whether any $p \times q$ matrix G should be considered to define a linear aggregation of regressors, is a matter of convention. Perhaps it is preferable to accept only those matrices G that satisfy the following two restrictions. First, no row vector of G is a zero vector. Second, the rank of G is q.

A special class of linear aggregations of regressors, which satisfy the two restrictions, will be introduced in 3.1.7. Aggregation of regressors is the subject of 5.1 and 5.2.

3.1.4 Aggregation of regressands

Aggregation of regressands starts from $k \geq 2$ micro regression data matrices $[\, y_i \,\vdots\, X_i \,]$, $i = 1, \ldots, k$, each of which refers to the same units of analysis. A macro regressand is formed as a function of the micro regressands. The regressors are not changed, in the sense that the union of the k sets of micro regressors is taken as the set of macro regressors. The individual units of analysis, and the set of units of analysis, are not changed. The outcome is a macro regression data matrix $[\, u \,\vdots\, Z \,]$.

Example 3.1.2.2 is a case of linear aggregation of regressands. The macro regressand is a linear function of the micro regressands. Further, in the example the micro regressor matrices X_i are equal by definition.

More generally, let two micro regressor vectors be considered identical if and only if they are equal by definition, not merely by numerical accident. Let the union of the k sets of p_i micro regressors X_i form the total micro regressor matrix X. Let the k micro regressand vectors y_i form the micro regressand matrix Y. Linear aggregation of regressands is described by the vector and matrix equations

$$u = Yh \quad ,$$
$$Z = X \quad ,$$

where h is some k-vector of constants.

Whether any k-vector h should be considered to define a linear aggregation of regressands, is a matter of convention. Perhaps it is preferable to accept only those vectors h that have no element equal to zero.

Aggregation of regressands is the subject of 5.3.

3.1.5 Aggregation of units of analysis

Aggregation of units of analysis starts from a single micro regression data matrix $[\, y \,\vdots\, X \,]$. Macro units of analysis are formed as functions of the micro units of analysis, in the sense that the data for any macro unit are formed as functions of the data of some given micro units. The regressors and regressand are not changed, in the sense that any macro datum for a given variable is a function of micro data for that variable only. The set of units of analysis is not

changed, in the sense that the macro data are functions of the micro data only, while (usually) each micro unit is "represented" in at least one macro unit. The outcome is a macro regression data matrix $[\, u \,\vdots\, Z \,]$.

The distinction between designed, observed and mixed regressors introduced in 2.3.3 may be relevant in aggregation of units of analysis. Let the micro and macro regressors be partitioned

$$X = [\, X^{\Delta} \,\vdots\, X^{M} \,\vdots\, X^{O} \,]$$
$$Z = [\, Z^{\Delta} \,\vdots\, Z^{M} \,\vdots\, Z^{O} \,],$$

where X^{Δ}, X^{M} and X^{O} are the designed, mixed, and observed micro regressors, while Z^{Δ}, Z^{M} and Z^{O} are the corresponding subsets of macro regressors. Aggregation of units of analysis often treats the purely observed regressors X^{O} in the same way as the regressand y, but may treat the designed and mixed regressors X^{Δ} and X^{M} differently.

Example 3.1.2.3 is a case of linear aggregation of units of analysis. Each macro datum is a linear function of some of the micro data. In the example, the intercept regressor is designed, the regressor x^{B}, the number of sheep, is purely observed, and the regressors x^{R} and x^{S}, the amounts of rain and snow, are mixed because they contain replicated observed regressor data. In the example, further, the purely observed regressor x^{B} is treated like the regressand y^{BF}, and the other regressors differently.

Linear aggregation of n micro units of analysis into m macro units of analysis is (incompletely) described by the matrix equation

$$[\, u \,\vdots\, Z^{O} \,] = T' \, [\, y \,\vdots\, X^{O} \,],$$

where T is some $n \times m$ matrix of constants. The treatment of the designed and mixed regressors can sometimes be described by the matrix equation

$$[\, Z^{\Delta} \,\vdots\, Z^{M} \,] = S' \, [\, X^{\Delta} \,\vdots\, X^{M} \,],$$

where S is some $n \times m$ matrix of constants. The matrix S is not necessarily uniquely determined by the process of aggregation. For example, if the intercept regressor is the only not purely observed micro and macro regressor, S is any matrix whose column sums are unity.

Whether any $n \times m$ matrix T, and any treatment of the designed and mixed regressors, should be considered to define a linear aggregation of units of analysis, is a matter of convention. Perhaps it is preferable to accept only those matrices T that satisfy the following two conditions. First, no row vector of T is a zero vector. Second, the rank of T is m. Whether the set of purely observed regressors X^O should be permitted to be empty is another matter of convention.

A special class of aggregations of units of analysis, which satisfy the two restrictions on T, will be introduced in 3.1.8. Aggregation of units of analysis is the subject of 6.1 and 6.2.

3.1.6 Aggregation of sets of units

Aggregation of sets of units starts from $k \geq 2$ micro regression data matrices $[\, y_h \,|\, X_h \,]$, $h = 1, \ldots, k$, each of which refers to the same variables. A macro set of units of analysis is formed as a function of the micro sets of units of analysis. The regressand and regressors are not changed. The individual units of analysis are not changed. The outcome is a macro regression data matrix $[\, u \,|\, Z \,]$.

Example 3.1.2.4 is a case of aggregation of sets of units. It may perhaps be characterized as linear, but it is difficult to imagine what a non-linear aggregation of sets of units might be like. The linear operation involved is not addition of scalars or vectors, but addition of sets.

Aggregation of sets of units is described by the matrix equation

$$[\, u \,|\, Z \,] = \left[\begin{array}{c|c} y_1 & X_1 \\ \hline \vdots & \vdots \\ \hline y_h & X_h \\ \hline \vdots & \vdots \\ \hline y_k & X_k \end{array}\right] .$$

The macro set of units is the union of the micro sets of units.

The micro sets of units have been taken to be, and to be treated as, disjoint. Generalization so as to permit overlapping micro sets is conceivable, but will not be discussed.

Aggregation of sets of units of analysis is the subject of 6.3.

3.1.7 Partitioned aggregation of regressors

An important special class of aggregation of p micro regressors into q macro regressors is as follows. Perhaps after a reordering, the micro regressors are partitioned into q disjoint families of micro regressors,

$$X = [\, X_1 \;\vdots\; \cdots\cdots \;\vdots\; X_i \;\vdots\; \cdots\cdots \;\vdots\; X_q \,] .$$

For $i = 1, \ldots, q$, the i'th family has $p_i \geq 1$ members. If $p_i = 1$, the single micro regressor forming the i'th family will be called a bachelor regressor. The partitioning is exhaustive, $p_1 + \ldots + p_q = p$.

For $i = 1, \ldots, q$, the i'th macro regressor z_i is a function of the members of the corresponding family of micro regressors only. The members of non-corresponding families do not influence z_i. This class of aggregations of regressors will be called partitioned aggregation of regressors.

Example 3.1.2.1 is a case of partitioned linear aggregation of regressors. The intercept regressor and the regressor x^B, the number of sheep, are bachelor regressors. The regressors x^R and x^S, for rain and snow, form a third family.

In partitioned linear aggregation of regressors, the transformation matrix G of 3.1.3 is block-diagonal.

$$G = \begin{Bmatrix} g_1 & \cdots\cdots & 0 & \cdots\cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots\cdots & g_i & \cdots\cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots\cdots & 0 & \cdots\cdots & g_q \end{Bmatrix}$$

For $i = 1, \ldots, q$, the order of the i'th diagonal block g_i is $p_i \times 1$. The matrix equation $Z = XG$ decomposes into the vector equations

$$z_i = X_i g_i , \quad i = 1, \ldots, q .$$

The conventional restrictions on G suggested in 3.1.3 require that no element of any g_i is zero.

The term "partitioned aggregation" is due to W. Fisher [1969], page 10.

3.1.8 Partitioned aggregation of units of analysis

An important special class of aggregations of n micro units of analysis into m macro units of analysis is as follows. Perhaps after a reordering, the micro units are partitioned into m disjoint subsets of micro units,

$$[\, y \mid X^{\Delta} \mid X^{M} \mid X^{O} \,] = \left[\begin{array}{c|c|c|c} y_1 & X_1^{\Delta} & X_1^{M} & X_1^{O} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline y_h & X_h^{\Delta} & X_h^{M} & X_h^{O} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline y_m & X_m^{\Delta} & X_m^{M} & X_m^{O} \end{array}\right]$$

For $h = 1, \ldots, m$, the h'th subset has $n_h \geq 1$ members. The partitioning is exhaustive, $n_1 + \ldots + n_m = n$.

For $h = 1, \ldots, m$, the h'th macro unit is a function of the members of the corresponding h'th subset of micro units only. This class of aggregations of units will be called partitioned aggregation of units of analysis.

Example 3.1.2.3 is a case of partitioned linear aggregation of units of analysis. The subset of micro units corresponding to the h'th macro unit consists of those parishes that belong to the h'th county.

In partitioned linear aggregation of units of analysis, the transformation matrix T of 3.1.5 is block-diagonal.

$$T = \begin{bmatrix} t_1 & \cdots & O & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & t_h & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & O & \cdots & t_m \end{bmatrix}$$

For $h = 1, \ldots, m$, the order of the h'th diagonal block t_h is $n_h \times 1$. The matrix equation $[\, u \,\vdots\, Z^O \,] = T' [\, y \,\vdots\, X^O \,]$ decomposes into the vector equations

$$[\, u_h \,\vdots\, z_h^{O'} \,] = t_h' [\, y_h \,\vdots\, X_h^O \,], \quad h = 1, \ldots, m \ .$$

The conventional restrictions on T suggested in 3.1.5 require that no element of any t_h is zero.

If a linear aggregation of units of analysis is such that T is block-diagonal while the transformation matrix S of 3.1.5 cannot be chosen to be block-diagonal conformably with T, the term "partitioned" is not appropriate.

Example 3.1.2.3 illustrates a particular kind of mixed regressors that may occur in partitioned aggregation of units of analysis. The macro regressor submatrix Z^M is logically prior to the micro regressor submatrix X^M, which is constructed from Z^M by repeating the h'th row n_h times, for $h = 1, \ldots, m$. Then

$$X^M = KZ^M$$

where the matrix K is as in 2.5.4. Mixed regressors of this kind may be called repetitive. The intercept regressor, although purely designed, can be counted as a repetitive mixed regressor, because $j_n = Kj_m$.

3.2 Aggregation of aspects

3.2.1 Simple aggregation of aspects

Simple aggregation of aspects starts from $k \geq 2$ micro regression data matrices. Each matrix refers to the same n units of analysis and has the same number $p + q$ of regressors. Each of the k matrices is concerned with a different aspect

of the units of analysis. Usually, the meaning of the micro regressand, or of the i'th micro regressor, is analogous in all aspects.

The regressors of any aspect are of two kinds, $p \geq 0$ common regressors and $q \geq 0$ specific regressors. If the i'th regressor is common, then the i'th micro regressors of all k aspects are equal by definition. If the i'th regressor is specific, then the i'th micro regressors of the aspects are all different, except perhaps by numerical accident. The micro regressands are specific to the aspects.

Perhaps after a reordering of the regressors, the micro regression data matrices can be written

$$[\, y_h \,\vdots\, X^C \,\vdots\, X_h^S \,] , \quad h = 1, \ldots, k ,$$

where X^C are the p common regressors, and X_h^S are the q regressors specific to the h'th aspect. The aspect index h has been omitted from the regressor submatrices X_h^C, which are all equal by definition.

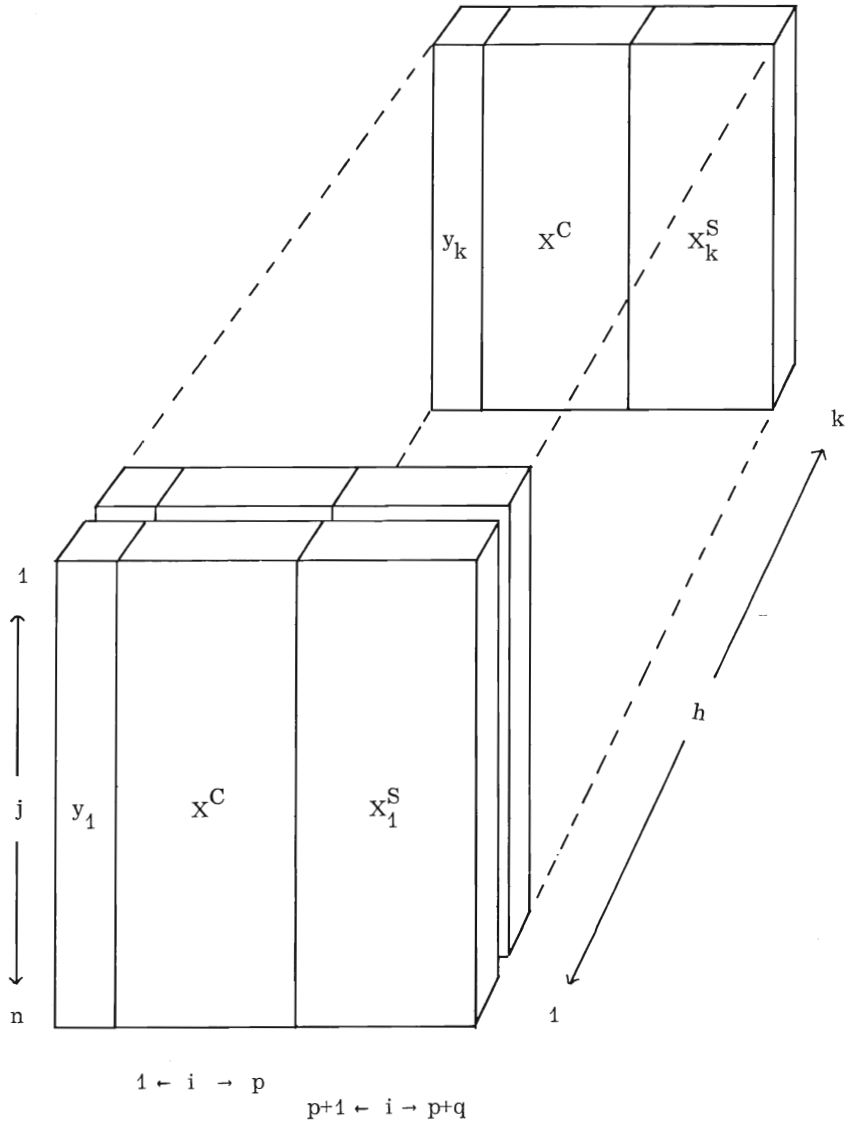
The k micro regression data matrices can be arranged into the three-dimensional data block shown in Figure 3.2.1. Rows and columns correspond to units of analysis and variables, respectively. The layers of the data block correspond to the aspects.

A macro regressand is formed as the function f of the micro regressands of the k aspects. If the i'th regressor is common, it is retained as the i'th macro regressor. If the i'th regressor of every aspect is specific, the i'th macro regressor is formed as the function f of the i'th micro regressors of the k aspects. The individual units of analysis, and the set of units, are not changed. The outcome is a macro regression data matrix

$$[\, u \,\vdots\, Z^C \,\vdots\, Z^S \,] ,$$

partitioned in the same way as the micro data matrices.

Example 3.1.2.5 is a case of simple linear aggregation of aspects. The function f is linear. There are two aspects, the black-faced sheep and the white-faced sheep. The intercept regressor and the regressors x^R and x^S, rain and snow, are common to the aspects. The fourth regressor x^B or x^W, number of black-faced or white-faced sheep, is specific to the aspects. The specific regressor is aggregated in the same way as the regressand y^{BF} or y^{WF}, number of



Aspects: $h = 1, \ldots, k$.

Regressors: $i = 1, \ldots, p + q$.

Units of analysis: $j = 1, \ldots, n$.

Figure 3.2.1 The data block in simple aggregation of aspects

black-faced or white-faced ewe lambs, which is also specific. The specific variables have analogous meanings in the two aspects.

Simple linear aggregation of aspects is described by the vector and matrix equations

$$u = \sum_{h=1}^{k} f_h y_h ,$$

$$Z^C = X^C ,$$

$$Z^S = \sum_{h=1}^{k} f_h X_h^S ,$$

where

$$f = (f_1 \cdots f_h \cdots f_k)'$$

is some k-vector of constants.

Whether any k-vector f should be considered to define a simple linear aggregation of aspects, is a matter of convention. Perhaps it is preferable to accept only those vectors f that have no element equal to zero.

Simple and more general aggregation of aspects is the subject of chapter 7.

3.2.2 Analysis in terms of regressands and regressors

Simple aggregation of aspects can be decomposed into two consecutive steps, each of which is one of the elementary types of aggregation. First, certain data matrices and groups of variables must be defined.

The micro regressand vectors y_h, ordered $h = 1, \ldots, k$, form the micro regressand matrix Y of k columns.

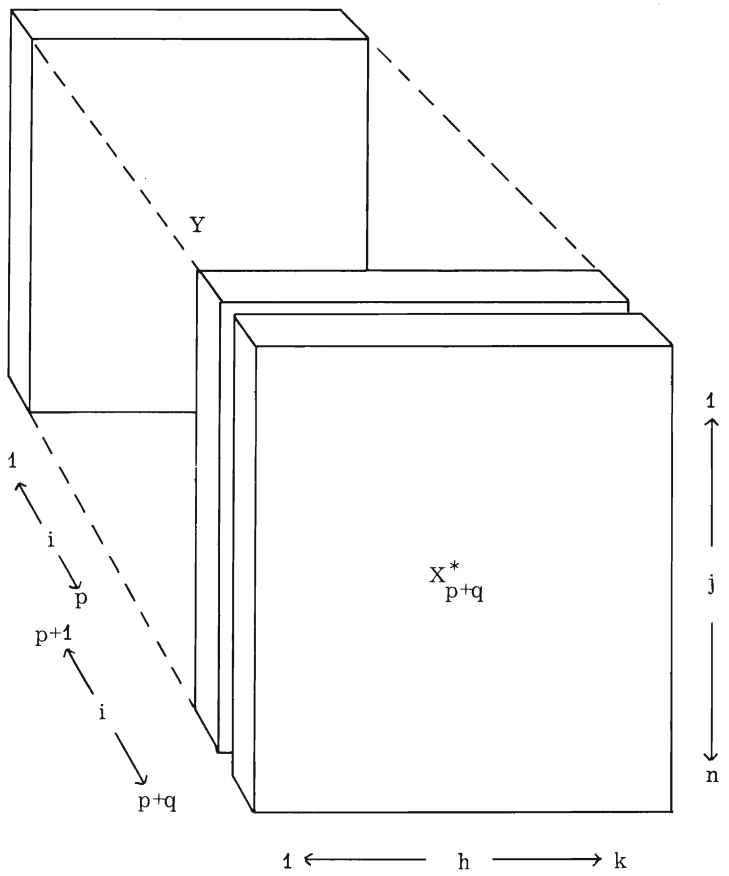
The common micro regressor vectors x_i, ordered $i = 1, \ldots, p$, form the common micro regressor submatrix X^C of p columns.

Consider the i'th specific regressor of each aspect. The k specific (usually analogous) micro regressor vectors x_{hi}, ordered $h = 1, \ldots, k$, form the i'th specific micro regressor submatrix X_i^* of k columns. There are q such matrices.

The total micro regressor matrix is

$$X^* = [X^C \mid X_{p+1}^* \mid \cdots\cdots \mid X_{p+q}^*]$$

and has $p + kq$ columns.



Aspects: $h = 1, \ldots, k$.

Regressors: $i = 1, \ldots, p + q$.

Units of analysis: $j = 1, \ldots, n$.

Figure 3.2.2 The data block seen from the right

The matrix X^C is a submatrix of every layer of the data block of Figure 3.2.1. The matrices Y and X_i^*, $i = p + 1, \ldots, p + q$, are columns of the data block, as shown in Figure 3.2.2.

The $p + kq$ micro regressors of X^* are partitioned into $p + q$ families as follows. Each common micro regressor x_i, $i = 1, \ldots, p$, is a bachelor regressor, and forms a single-member family. For each $i = p + 1, \ldots, p + q$, the k specific (usually analogous) micro regressors of X_i^* form a family.

The first step of simple aggregation of aspects is an aggregation of regressands as described in 3.1.4. It starts from the total micro data matrix $[\, Y \mid X^* \,]$. A macro regressand u is formed as a function f of the k members of the micro regressand matrix Y. If the aggregation is linear, then

$$u = Yf \, .$$

The outcome is a semi-aggregated regression data matrix $[\, u \mid X^* \,]$, where the regressand vector is macro while the regressor matrix is micro.

The second step of simple aggregation of aspects is a partitioned aggregation of regressors as described in 3.1.7. It starts from the semi-aggregated regression data matrix $[\, u \mid X^* \,]$. A macro regressor is formed from each family of micro regressors. For $i = 1, \ldots, p$, the i'th macro regressor z_i is identical to the i'th common micro regressor x_i, i.e. $Z^C = X^C$. For $i = p + 1, \ldots, p + q$, the i'th macro regressor z_i is formed as the function f of the k members of the i'th family of specific (usually analogous) micro regressors X_i^*. If the aggregation is linear, then

$$z_i = X_i^* f \, , \quad i = p + 1, \ldots, p + q \, .$$

The outcome is a macro regression data matrix

$$[\, u \mid Z \,] = [\, u \mid Z^C \mid Z^S \,]$$

which is analogous to the layers of the data block. If the aggregation is linear, the second step can be written $Z = XG$, where

$$G = \begin{bmatrix} I_p & O & O & \cdots & O \\ O & f & O & \cdots & O \\ O & O & f & \cdots & O \\ \vdots & \vdots & \vdots & & \vdots \\ O & O & O & \cdots & f \end{bmatrix}$$

is a block-diagonal $(p + kq) \times (p + q)$ matrix.

Simple aggregation of aspects is an aggregation of regressands followed by a particular partitioned aggregation of regressors.

3.2.3 Analysis in terms of sets and units

Simple aggregation of aspects can be decomposed into two consecutive steps different from those of 3.2.2. Again, each step is one of the elementary steps of aggregation. First, sub-units of analysis, and certain sets of sub-units, must be defined.

There are k aspects and n units of analysis. Each combination of one aspect h and one unit of analysis j defines a sub-unit of analysis, double-indexed hj. The data for sub-unit hj are the row data vector

$$(y \mid x_1 \cdots x_p \mid x_{p+1} \cdots x_{p+q})_{hj} .$$

This vector is found at the intersection of layer h and row j in the data block. There are kn sub-units of analysis.

For $h = 1, \ldots, k$, a sub-unit belongs to the set A_h if and only if it belongs to the h'th aspect. For $j = 1, \ldots, n$, a sub-unit belongs to the set U_j if and only if it belongs to the j'th unit of analysis. The total set of sub-units is denoted S. The sub-units and sets are indicated by Figure 3.2.3 which is, in a sense, the right hand surface of the data block.

The first step in the alternative decomposition of simple aggregation of aspects is an aggregation of sets of units as described in 3.1.6. It starts from the micro sets of sub-units A_h, $h = 1, \ldots, k$. A macro set of units is formed as the union of the micro sets. The outcome is the total set S of sub-units of analysis.

unit \ aspect	1 ..... h ..... k	sets
1	11 ..... h1 ..... k1	U_1
j	1j ..... hj ..... kj	U_j
n	1n ..... hn ..... kn	U_n
sets	A_1 ..... A_h ..... A_k	S

Figure 3.2.3 Sub-units and sets in simple aggregation of aspects

The second steps is a partitioned aggregation of units of analysis as described in 3.1.8. It starts from the total set S of sub-units of analysis. The total set is partitioned into the equally large subsets U_j, $j = 1, \ldots, n$. For any j, let $[\, y_j \mid X_j^C \mid X_j^S \,]$ denote the matrix formed by the k row data vectors for the sub-units that belong to U_j. Then the partitioned regression data matrix is

$$[\, y_* \mid X_*^C \mid X_*^S \,] = \left[\begin{array}{c|c|c} y_1 & X_1^C & X_1^S \\ \hline \vdots & \vdots & \vdots \\ \hline y_j & X_j^C & X_j^S \\ \hline \vdots & \vdots & \vdots \\ \hline y_n & X_n^C & X_n^S \end{array}\right]$$

The common regressors of X_*^C are equal for all sub-units within any one subset U_j. Thus, these regressors are repetitive mixed regressors, and X_*^C corresponds to X^M of 3.1.8.

The j'th macro unit of analysis is formed from the k members of the subset U_j of sub-units as follows. The regressand and the specific regressors are aggregated by means of a function f. For the common regressors, data are taken

from any one member of U_j. The outcome is a macro data matrix $[u \quad Z^C \quad Z^S]$ of n rows. If the aggregation is linear, then

$$[u \mid Z^S] = T' [y_* \mid X_*^S],$$

where

$$T = \begin{bmatrix} f & \cdots & O & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & f & \cdots & O \\ \vdots & & \vdots & & \vdots \\ O & \cdots & O & \cdots & f \end{bmatrix}$$

is a block-diagonal $kn \times n$ matrix.

Simple aggregation of aspects can be regarded as an aggregation of sets of units followed by a particular partitioned aggregation of units of analysis.

3.2.4 General aggregation of regressands and regressors

Simple aggregation of k aspects is, by 3.2.2, an aggregation of regressands followed by a partitioned aggregation of regressors. The micro regressors are of two kinds. Some are common to all aspects, while the others are specific, each to one aspect. The families of micro regressors are also of two kinds. Some consist of one common micro regressor. The others consist of k specific micro regressors, one from each aspect.

Example 3.1.2.6 is a case of what might perhaps be called general linear aggregation of aspects. There are four aspects, the four kinds of lambs. The intercept regressor and the rain and snow regressors, x^R and x^S, are common to all four aspects. Each of the two regressors for numbers of sheep, x^B and x^W, occurs in two out of the four aspects. No regressor is specific to one aspect.

Example 3.1.2.6 can be decomposed into an aggregation of regressands followed by a partitioned aggregation of regressors. There are three families of micro regressors. One consists of the intercept regressor alone. The second family consists of x^B and x^W. The third family consists of the two micro regressors x^R and x^S, which are common to all aspects.

A modified example, say 3.1.2.6*, is obtained if the snow regressor x^S is excluded a priori from the two micro regression data matrices for ram lambs.

Then two of the aspects have four regressors, while the other two aspects have only three regressors.

Whether example 3.1.2.6, or 3.1.2.6*, should be called an aggregation of aspects, is a matter of convention. The term general aggregation of regressands and regressors is perhaps to be preferred.

3.2.5 Types of aggregation in the literature

The five types of aggregation are not in themselves an innovation, but the names given to them in chapter 3 are new. Each of the five types is likely to be found in the literature under some different name, or perhaps without a name.

In aggregation of regressors, of regressands, or of aspects — unlike aggregation of units, or of sets — no information on the structure of the aggregation is gained by considering more than one unit of analysis. These three types of aggregation are often studied in the following context rather than that of regression analysis. Interest is focused on the micro and macro relations as theoretical models valid for any unit of analysis. The generic unit of analysis is considered.

Outside the context of regression analysis, aggregation of regressors and regressands are better called aggregation of independent and dependent variables.

Partitioned aggregation of independent variables has been studied by many economists. A typical name used is "grouping of variables in a single utility or production function". Green [1964]. The concept of functional separability mentioned in 1.4.1 is concerned with this type of aggregation.

Partitioned aggregation of units of analysis in a context of regression analysis is studied by Prais and Aitchison [1954] under the name of "grouping of observations", and by Cramer [1964]. As pointed out by Blalock [1964], the sociological concept of "ecological correlation", which will be touched upon in 8.3.3, is concerned with this type of aggregation.

Aggregation of sets of units is often called "pooling". If the sets are generated by different models, the aggregation entails what Wold [1940], paragraph 23, calls a "stratification effect". Given standard assumptions, the homogeneity of the micro models can be tested by analysis of covariance as shown by Malinvaud [1964], section 7.7.

A closely related complex of problems is raised by the alternative or joint use of cross section and time series data, or of a cross section followed over time. An uncommonly thorough empirical investigation of a cross section followed over time is Kuh [1963], chapter 5. The relevance of the problem complex to econometric model construction is briefly indicated by Christ [1966], several sections of Part Two. Its relevance to econometric estimation is evident also from Kimbell [1970]. The connection with aggregation is stated explicitly by Ringstad [1971]; cf. also Misra [1972]. In view of the great practical importance of this complex of problems, it is remarkable how little most text-books of econometrics have to say about it.

Simple aggregation of aspects for the generic unit of analysis is probably the type of aggregation most studied by economists. Often, the unit of analysis understood is a period of time, the aspects are households or firms, the micro relation is a micro-economic theory, and the macro data are macro-economic aggregates. A name often used is "aggregation of economic relations", e.g. Green [1964]. Quite naturally, this is the type of aggregation primarily considered in a text-book of macro-economics, Ackley [1961], chapter XX, or in a textbook of econometrics, Theil [1971], section 11.3.

In a context of regression analysis, Theil [1954] distinguishes several types of aggregation. One of these is concerned with systems of simultaneous equations, and is very complex. Theil's other types of aggregation are as follows.

"Aggregation over one set of individuals", Theil [1954], chapter II, is simple aggregation over aspects. "Aggregation over several sets of individuals or commodities", chapter III, is more general aggregation of regressands and regressors.

"Aggregation over time periods", chapter IV, is aggregation of units of analysis, if there are no lagged variables. The presence of time lags makes it much mor complex.

"Aggregation in a changing economy", Theil [1954], section 6.1, is concerned with a time series study of macro-economic data. The total set of time periods is partitioned into disjoint subsets of consecutive periods, say epochs. Each epoch is characterized by a different number of micro-economic entities and by different parameters. First, the micro-economic entities are aggregated within

each epoch. Second, the epochs are aggregated. In other words, this is a set of separate general aggregations of regressands and regressors, followed by an aggregation of sets of units.

4 A FORMAL ANALYSIS OF LINEAR AGGREGATION IN REGRESSION: ASSUMPTIONS AND PROBLEMS

4.1 Outline and linearity assumptions

4.1.1 The formal analysis in outline

Chapters 4 to 8 of this study apply a common scheme of formal analysis to each of the five types of aggregation of regression data introduced in chapter 3. The concepts involved in the scheme of analysis have been introduced in chapters 1 and 2. The assumptions and problems involved will be specified in the present chapter 4. The analysis is performed in chapters 5, 6 and 7. Conclusions are stated in chapter 8, where certain related questions are also discussed.

For each of three types of aggregation, the scheme of analysis will be applied to two different variants of the type. The first variant is quite general, while the second variant is a special case of particular interest. The three types of aggregation occurring in two variants are aggregation of regressors (5.1 and 5.2), aggregation of units of analysis (6.1 and 6.2), and aggregation of aspects (7.1 and 7.2).

For each of the remaining two types of aggregation, the scheme of analysis will be applied to a single variant. These two types are aggregation of regressands (5.3) and aggregation of sets of units (6.3).

The scheme of analysis is concerned with two different problems. The first problem is a certain aggregation problem directly related to consistency as discussed in 1.2. In particular, it is an articulated consistency problem in the sense of 1.2.4. This consistency problem will be introduced in 4.2.

The second problem is less precise. It is about how to interpret the macro regression coefficients in micro terms as discussed in 1.3. In particular, wherever possible the interpretation will be by means of a model-free relation, and otherwise by means of an expectational relation, as discussed in 1.3.2. These two variants of interpretation will be introduced in 4.3 and 4.4.

4.1.2 Two linearity assumptions

As a preliminary, four subjects are briefly recapitulated. First, by 1.1.4 and 1.1.5 an aggregation is defined by four vector functions $\{G, H, \Phi, \Psi\}$, and the latter two may be stochastic. Second, by 3.1.1 the data in each of the five types of aggregation conform to the following pattern. The micro data form k regression data matrices $[\, y_h \,\vdots\, X_h \,]$, $h = 1, \ldots, k$, where in some types $k = 1$ and in others $k \geq 2$. The macro data form a single regression data matrix $[\, u \,\vdots\, Z \,]$. Third, in 2.3.3 a distinction was made between designed and observed regressor data. Fourth, in 2.3.1 the incomplete linear model was defined.

The formal analysis in chapters 4 to 8 is based throughout on two assumptions of linearity. The first linearity assumption is in two parts as follows. The regressor aggregating function

$$G: \quad Z = g(X_1, \ldots, X_h, \ldots, X_k)$$

is linear in the sense that each observed macro regressor datum is a linear function of some observed micro regressor data. Similarly, the regressand aggregating function

$$H: \quad u = h(y_1, \ldots, y_h, \ldots, y_k)$$

is linear in the sense that each macro regressand datum is a linear function of some micro regressand data. (All regressand data are observed.)

The second linearity assumption is in two parts as follows. Whenever a micro relation is assumed, or is otherwise considered, it is taken to consist of k incomplete linear models, one for each micro regression data matrix.

$$\Phi: \left\{ \begin{array}{l} y_h = X_h \beta_h + \varepsilon_h \,; \\ E(\varepsilon_h) = 0 \,, \end{array} \right\} \quad h = 1, \ldots, k \,.$$

Similarly, whenever a macro relation is considered, it is taken to consist of an incomplete linear model

$$\Psi: \left\{ \begin{array}{l} u = Z\delta + \eta \,; \\ E(\eta) = 0 \,. \end{array} \right.$$

The covariance matrices of ε_h and η are always left unspecified.

Each of the k linear models of Φ will be called a micro model, and its parameter vector β_h a vector of micro parameters. The vector

$$\beta = (\beta_1' \mid \ldots \mid \beta_h' \mid \ldots \mid \beta_k')'$$

will be called the total micro parameter vector, also when $k = 1$. The linear model of Ψ will be called the macro model, and its parameter vector δ the vector of macro parameters.

4.1.3 On simultaneous linearity

Consider the following example of simple aggregation of aspects. The aspects $h = 1, \ldots, k$ are different households. The generic unit of analysis considered (cf. 3.2.5) is a time period. The observed variables are

x_h = the disposable income of the h'th household,
y_h = the demand for tea of the h'th household,
z = the joint disposable income of the k households,
u = the joint demand for tea of the k households.

The regressor and regressand aggregating functions are

$$G: \quad z = \sum_{h=1}^{k} x_h ;$$

$$H: \quad u = \sum_{h=1}^{k} y_h .$$

The micro and macro relations are

$$\Phi: \quad y_h = \alpha_h x_h^{\beta_h}, \quad h = 1, \ldots, k ;$$

$$\Psi: \quad u = \gamma z^{\delta} .$$

The micro and macro relations consist of constant-elastic demand functions where prices are implicitly held constant.

Alternatively, the data can be converted into natural logarithms,

$$x_h^* = \log x_h \ ,$$

$$y_h^* = \log y_h \ ,$$

$$z^* = \log z \ ,$$

$$u^* = \log u \ .$$

The regressor and regressand aggregating functions, and the micro and macro relations can then be expressed as follows.

$$G^* : \ z^* = \log \sum_{h=1}^{k} \exp x_h^* \ ;$$

$$H^* : \ u^* = \log \sum_{h=1}^{k} \exp y_h^* \ ;$$

$$\Phi^* : \ y_h^* = \log \alpha_h + \beta_h x_h^* \ , \quad h = 1, \ldots, k \ ;$$

$$\Psi^* : \ u^* = \log \gamma + \delta z^* \ .$$

The transformed micro and macro relations are linear.

The aggregation $\{G, H, \Phi, \Psi\}$ can be described as a linear aggregation of non-linear relations. The equivalent aggregation $\{G^*, H^*, \Phi^*, \Psi^*\}$ can be described as, on the contrary, a non-linear aggregation of linear relations.

A first linearity assumption for $\{G, H\}$ and a second linearity assumption for $\{\Phi, \Psi\}$ were stated in 4.1.2. Both are meant to apply simultaneously to specified data. As the example indicates, the simultaneity is not unimportant. (Cf. also 2.2.2.)

4.2 A consistency problem

4.2.1 The status of the four vector functions

Section 4.2 introduces a consistency problem. Whenever this consistency problem is studied in chapters 4 to 8, the following applies.

The regressor and regressand aggregating functions $\{G, H\}$ are assumed to be given linear vector functions with numerically specified coefficients.

Certain classes of micro and macro relations $\{\Phi, \Psi\}$ are considered, but there is no assumption that any micro or macro relation is valid.

The micro relations Φ considered consist of incomplete linear micro models in the sense of 4.1.2. These micro relations are identified by their total micro parameter vectors β.

The macro relations Ψ considered consist of an incomplete linear macro model. These macro relations are identified by their macro parameter vectors δ.

4.2.2 A limitation

Section 4.2 is the outcome of a conflict, where the desire for simplicity partly defeated the desire for generality. The fundamental ideas invoked are applicable in a wider context.

The notation and the technical details are tailor-made for the five types of aggregation as delimited in chapters 5, 6 and 7. Even so, a certain modification will be needed in aggregation of units of analysis, in 6.1.3.

Recall the concepts introduced in 2.3.3. A regressor vector is either designed, observed, or mixed. This distinction is relevant to the formulation of the consistency problem (but not to the interpretation problem). For simplicity, the following limitation is maintained throughout. Whenever the consistency problem is considered, there are assumed to be no mixed micro regressor vectors. A partial exception is however made in partitioned aggregation of units, in 6.2.2 and 6.2.4.

4.2.3 The semi-aggregated and semi-disaggregated models

The regressand aggregating function H is given. Any particular micro relation Φ considered implies a semi-aggregated micro relation $H\Phi$ in the sense of 1.1.5. Because of the linearity assumptions of 4.1.2, $H\Phi$ can be written as an incomplete linear model. This can be done in more than one way, but for each type of aggregation a particular model formulation will be chosen.

$$H\Phi: \begin{cases} u = X_A \beta_A + \epsilon_A , \\ E(\epsilon_A) = 0 . \end{cases}$$

This model $H\Phi$ will be called the semi-aggregated micro model. The matrix X_A will be called the semi-aggregated (micro) regressor matrix. The vector β_A

will be called the semi-aggregated micro parameter vector. The elements of β_A are known linear functions of some elements of the total micro parameter vector β of 4.1.2.

The regressor aggregating function G is given. Any particular macro relation Ψ considered implies a semi-disaggregated macro relation ΨG in the sense of 1.1.5. Because of the linearity assumptions of 4.1.2, ΨG can be written as an incomplete linear model. For each type of aggregation, a particular model formulation will be chosen, which involves the semi-aggregated (micro) regressor matrix X_A.

$$\Psi G: \begin{cases} u = X_A \delta_D + \eta , \\ E(\eta) = 0 . \end{cases}$$

This model ΨG will be called the semi-disaggregated macro model. The vector δ_D will be called the semi-disaggregated macro parameter vector. The elements of δ_D are known linear functions of some elements of the macro parameter vector δ.

The semi-aggregated regressor matrix X_A will always be defined so as to have the following properties. It can be partitioned

$$X_A = [\, X_A^{\Delta} \,\vdots\, X_A^{O} \,] .$$

Each element of the semi-aggregated designed regressor matrix X_A^{Δ} is designed. Some elements of the semi-aggregated observed regressor matrix X_A^{O} may be designed to be zero, but no column of X_A^{O} consists exclusively of such elements. Each element of X_A^{O} that is not designed to be zero, is a known linear function of some observed micro regressor data.

The parameter vectors of $H\Phi$ and ΨG are partitioned conformably with the semi-aggregated regressor matrix as follows

$$[\, \beta_A \,\vdots\, \delta_D \,] = \left[\begin{array}{c:c} \beta_A^{\Delta} & \delta_D^{\Delta} \\ \hdashline \beta_A^{O} & \delta_D^{O} \end{array}\right] .$$

The subvectors β_A^{Δ} and δ_D^{Δ} will be called the semi-aggregated micro and semi-disaggregated macro shift vectors. The subvectors β_A^{O} and δ_D^{O} will be called the

semi-aggregated micro and semi-disaggregated macro slope vectors. Further, the two vectors α_A and γ_D defined as follows

$$[\alpha_A \vdots \gamma_D] = X_A^{\Delta} [\beta_A^{\Delta} \vdots \delta_D^{\Delta}]$$

will be called the semi-aggregated micro and semi-disaggregated macro intercept vectors.

The semi-aggregated micro and semi-disaggregated macro models can be reformulated as follows.

$$H\Phi: \quad E_{H\Phi}(u) = \alpha_A + X_A^O \beta_A^O \ .$$

$$\Psi G: \quad E_{\Psi G}(u) = \gamma_D + X_A^O \delta_D^O \ .$$

These model formulations will be used whenever the consistency problem is studied in chapters 4 to 8.

4.2.4 A set of independent micro data vectors

Let the number of macro units of analysis be denoted m. The observed semi-aggregated (micro) regressor matrix X_A^O has m rows. From the j'th row vector of X_A^O, omit any elements designed to be zero. Let the remaining subvector, a row vector of order q_j, be denoted x_j'.

The independent micro data vector x of 1.1 and 1.2 will be identified with the following vector of $1 + q_1 + \ldots + q_j + \ldots + q_m$ elements.

$$x = (1 \vdots x_1' \vdots \ldots \vdots x_j' \vdots \ldots \vdots x_m')' .$$

The first element of x is designed to be unity.

The concept of consistency formulated in 1.2.3 refers to a set Ξ of independent micro data vectors x. Such a set Ξ will be said to be of full subvector rank if and only if it satisfies the following incomplete specification. For every $j = 1, \ldots, m$ one can select from Ξ $q_j + 1$ vectors x such that the $q_j + 1$ subvectors of order $q_j + 1$

$$(1 \vdots x_j')'$$

are linearly independent.

Whenever the consistency problem is studied in chapters 4 to 8, a set Ξ of full subvector rank is understood. This means two things. First, the designed micro regressor data remain fixed, while the observed micro regressor data are variable. Second, the observed micro regressor data are subject to no restriction favourable to consistency; that this is so will be shown in 4.2.6.

In four out of the five types of aggregation, it is easy to see what the specification of Ξ means in terms of the micro regressor matrices X_h, $h = 1, \ldots, k$. In the remaining type, aggregation of units of analysis, this matter must be given special attention; see 6.1.2.

4.2.5 An indirect consistency problem

An indirect consistency problem in the sense of 1.2.2 will now be formulated. This crude problem will be refined in 4.2.7.

The four vector functions $\{G, H, \Phi, \Psi\}$ are as stated in 4.2.1. The independent micro data vector x is as in 4.2.4. There is a set Ξ of independent micro data vectors, and Ξ is of full subvector rank as defined in 4.2.4. In conformity with 1.2.3, the term "consistency" is understood to mean expectational consistency for all $x \in \Xi$.

The total parameter vector p consists of the parameters of the micro and macro relations Φ and Ψ. Thus (cf. 4.1.2),

$$p = (\beta' \vdots \delta')' ,$$

where β is the total micro parameter vector and δ is the macro parameter vector. The coefficients of the linear regressor and regressand aggregating functions G and H are not included in p, because they are assumed to be given and fixed.

To the given set Ξ of independent micro data vectors x there corresponds a set $\Pi(\Xi)$ of total parameter vectors p such that the aggregation is consistent if and only if $p \in \Pi(\Xi)$. The set $\Pi(\Xi)$ may be empty.

The indirect consistency problem asks: What is the set $\Pi(\Xi)$? In other words, what micro and macro models, identified by β and δ, combine into a consistent aggregation?

4.2.6 A general proposition

The semi-aggregated micro model $H\Phi$ and the semi-disaggregated macro model ΨG both make the expected macro regressand vector $E(u)$ a function of the semi-aggregated observed (micro) regressor matrix X_A^O. The aggregation is consistent if and only if

$$E_{H\Phi}(u) = E_{\Psi G}(u)$$

whenever X_A^O is admissible according to the set Ξ of independent micro data vectors.

The following proposition uses the notation and terminology introduced in 4.2.3, 4.2.4 and 4.2.5.

Proposition 4.2.6

Consider an aggregation where the semi-aggregated micro relation $H\Phi$ and the semi-disaggregated macro relation ΨG are the following incomplete linear models.

$$H\Phi: \quad E_{H\Phi}(u) = \alpha_A + X_A^O \beta_A^O .$$

$$\Psi G: \quad E_{\Psi G}(u) = \gamma_D + X_A^O \delta_D^O .$$

Let no column vector of the semi-aggregated observed micro regressor matrix X_A^O consist exclusively of elements designed to be zero. The aggregation is characterized by the semi-aggregated micro and semi-disaggregated macro intercept and slope vectors $\{\alpha_A, \beta_A^O, \gamma_D, \delta_D^O\}$.

Consider a set Ξ of independent micro data vectors x, where x consists of all elements of X_A^O not designed to be zero, plus a single element designed to be unity. Let the set Ξ be of full subvector rank.

A necessary and sufficient condition for consistency is

$$\begin{bmatrix} \alpha_A \\ \hline \beta_A^O \end{bmatrix} = \begin{bmatrix} \gamma_D \\ \hline \delta_D^O \end{bmatrix} \quad . \quad \circ$$

Let the j'th elements of u, α_A and γ_D be denoted u_j, α_j and γ_j. Let the subvector of x deriving from the j'th row of X_A^O be denoted x_j. Let the subvectors

of β_A^O and δ_D^O associated with x_j be denoted β_j and δ_j. The pair $\{H\Phi, \Psi G\}$ of m-vector equations consists of the m pairs of scalar equations

$$\left.\begin{aligned} E_{H\Phi}(u_j) &= \alpha_j + x_j' \beta_j \\ E_{\Psi G}(u_j) &= \gamma_j + x_j' \delta_j \end{aligned}\right\} \quad j = 1, \ldots, m .$$

The aggregation is point-consistent for a certain X_A^O if and only if

$$(1 \mid x_j') \left[\begin{bmatrix} \alpha_j \\ \hline \beta_j \end{bmatrix} - \begin{bmatrix} \gamma_j \\ \hline \delta_j \end{bmatrix} \right] = 0$$

for $j = 1, \ldots, m$.

Necessity. Assume consistency.

Consider a particular j. Since Ξ is of full subvector rank, there exist $q_j + 1$ linearly independent admissible subvectors $(1 \mid x_j')$ of order $q_j + 1$. Let these row vectors form a square matrix M_j. Because of the consistency,

$$M_j \left[\begin{bmatrix} \alpha_j \\ \hline \beta_j \end{bmatrix} - \begin{bmatrix} \gamma_j \\ \hline \delta_j \end{bmatrix} \right] = 0 .$$

Thus since M_j is non-singular, $\alpha_j = \gamma_j$ and $\beta_j = \delta_j$.

Repeat for $j = 1, \ldots, m$. Since no column vector of X_A^O is designed to be a zero vector, every element of $\beta_A^O - \delta_D^O$ occurs in at least one $\beta_j - \delta_j$. Thus $\alpha_A = \gamma_D$ and $\beta_A^O = \delta_D^O$.

Sufficiency. Assume that $\alpha_A = \gamma_D$ and $\beta_A^O = \delta_D^O$. Then

$$E_{H\Phi}(u) = E_{\Psi G}(u)$$

for any matrix X_A^O of the appropriate order. □

The elements of the semi-aggregated micro and semi-disaggregated macro intercept and slope vectors $\{\alpha_A, \beta_A^O, \gamma_D, \delta_D^O\}$ are known functions of the elements of the total micro and macro parameter vectors $\{\beta, \delta\}$. Consequently, P.4.2.6 provides the answer of the indirect consistency problem formulated in 4.2.5.

The set Ξ of independent micro data vectors x has been specified to be of full subvector rank. If this specification were favourable to consistency, the aggrega-

tion could be expectationally point-consistent for every $x \in \Xi$, and expectationally point-inconsistent for at least one x outside Ξ. The demonstration of P.4.2.6 shows that this cannot happen. If the aggregation is point-consistent for every $x \in \Xi$, the "necessity" part of the demonstration shows that the micro and macro parameters have certain properties. The "sufficiency" part shows that these properties imply expectational point-consistency for every x. Thus, as was stated in 4.2.4, the specification of Ξ is not favourable to consistency.

4.2.7 An articulated consistency problem

The indirect consistency problem of 4.2.5 will now be refined into an articulated consistency problem in the sense of 1.2.4.

Consider the indirect consistency problem and a given total micro parameter vector β. It may or may not be possible to find a macro parameter vector δ such that the aggregation is consistent. If this is possible, δ must be chosen so as to fit β.

The total parameter vector p is partitioned into a subvector p_1 of critical parameters and a subvector p_2 of discretionary parameters. In particular,

$$p_1 = \beta \quad \text{and} \quad p_2 = \delta .$$

The following two questions are asked.

Principal question. For what class of total micro parameter vectors β is concistency at all attainable?

Corollary question. When consistency is attainable, how should the macro parameter vector δ be chosen in order to attain it?

Throughout chapters 4 to 8, this indirect and articulated consistency problem will be called the consistency problem. Indirectly, P.4.2.6 provides the answer of the consistency problem. Attention will be focused on the principal question, because the corollary question will always be found to be rather trivial.

4.2.8 References to literature

The consistency problem formulated in 4.2 is inspired by two similar problems formulated by Theil [1954] and Ijiri [1968].

Theil [1954] formulates in chapter 7 a Rule of Perfection, which is equivalent

to the concept of expectational consistency. He then formulates the following consistency problem. The aggregating functions G and H are linear and given. The micro and macro relations Φ and Ψ consist of linear models. Does the aggregation satisfy the Rule of Perfection?

For several different types of aggregation, Theil [1954] formulates necessary and sufficient conditions for the Rule of Perfection to be satisfied. The conditions are restrictions on the total micro parameter vector β. They do not mention the macro parameter vector δ.

Ijiri [1968] formulates the following consistency problem. Let u, x and z be as e.g. in Figure 1.1.3, and let A and B be coefficient matrices of appropriate orders. Ijiri considers a general deterministic linear semi-aggregated micro relation and a general linear aggregating function for independent data,

$$H\Phi: \quad u = Ax ;$$
$$G: \quad z = Bx .$$

Is it possible to find a deterministic macro relation

$$\Psi: \quad u = \psi(z)$$

such that the aggregation is consistent?

Ijiri [1968] formulates a necessary and sufficient condition for a consistent macro relation to exist. The condition is a restriction on the coefficient matrices A and B. The derivation of the condition is made more explicit in Ijiri [1971], section 2.4.

Ijiri's problem and solution can be generalized so as to permit stochastic vector functions $H\Phi$ and Ψ. Simply interpret u as the expected dependent macro data vector, and "consistency" as expectational consistency.

When "stochasticized" as indicated, Ijiri's problem and solution include the consistency problem of 4.2.7 and its solution by P.4.2.6 as the following special case. The aggregating functions are given. Thus the matrix B is given, and the matrix A is a function of the total micro parameter vector β only. The consistency condition in terms of A and B is read as a restriction on β.

In a cryptical footnote, Ijiri [1968] acknowledges the connection with Theil [1954].

4.3 Interpretation by a model-free relation

4.3.1 The status of the four vector functions

Section 4.3 introduces an interpretation of macro regression coefficients by means of a model-free relation. Whenever this model-free interpretation is studied in chapters 4 to 8, the following applies.

The regressor and regressand aggregating functions $\{G, H\}$ are assumed to be given linear vector functions with numerically specified coefficients.

No micro or macro relations $\{\Phi, \Psi\}$ are assumed.

4.3.2 The macro and micro statistics considered

Recall the concepts of model-free least-squares regression introduced in 2.2.

The macro data form an m-rowed regression data matrix $[\, u \mid Z \,]$. The macro statistics considered are the GLS(W) macro regression coefficients d, where W is a positive-definite $m \times m$ matrix. The vector d satisfies the system of macro normal equations

$$Z'WZd = Z'Wu \,,$$

but d need not be unique.

The micro data form n_h-rowed regression data matrices $[\, y_h \mid X_h \,]$, $h = 1, \ldots, k$. The micro statistics considered are the GLS(Q_h) micro regression coefficients b_h, where Q_h is a positive-definite $n_h \times n_h$ matrix. The vector b_h satisfies the h'th system of micro normal equations

$$X_h'Q_hX_hb_h = X_h'Q_hy_h \,,$$

but b_h need not be unique. If $k = 1$, the index h is dropped throughout.

Whenever the model-free interpretation is studied, the positive-definite matrix W is arbitrary, unless otherwise stated. The positive-definite matrices Q_h, on the contrary, are always chosen so as to fit W.

4.3.3 The semi-aggregated and auxiliary statistics

The semi-aggregated (micro) regressor matrix introduced in 4.2.3 and the macro regressand vector form an m-rowed regression data matrix $[\, u \mid X_A \,]$. The

model-free interpretation is primarily in terms of the GLS(W) semi-aggregate regression coefficients b_A, where W is the positive-definite m × m matrix involved in the macro regression coefficients d. The vector b_A satisfies the system of semi-aggregate normal equations

$$X_A'WX_Ab_A = X_A'Wu ,$$

but b_A need not be unique.

The model-free interpretation involves the concept of auxiliary regression. The auxiliary regressions are the GLS(W) regressions of each of the p semi-aggregated regressors X_A upon the q macro regressors Z, where W is the positive-definite m × m matrix involved in the macro and semi-aggregate regression coefficients d and b_A. The q × p matrix of auxiliary regression coefficients C satisfies the p systems of q auxiliary normal equations

$$Z'WZC = Z'WX_A ,$$

but C need not be unique.

The concept of auxiliary regression is due to Theil [1954], Theil [1957].

4.3.4 A general proposition

The macro regression coefficients d are now to be interpreted in semi-aggregate terms. But the macro normal equations may not determine d uniquely. Therefore, attention is limited to those macro coefficient functions $\mu'd$ that are unique in the sense defined in 2.2.4.

The following proposition confronts regressions of the macro regressand u upon two different sets of regressors. One set is the macro regressors Z, and the other set is the semi-aggregated micro regressors X_A. The proposition interprets $\mu'd$ in terms of a semi-aggregate regression coefficient function $\lambda'b_A$.

Proposition 4.3.4

Let u be a regressand m-vector. Let Z be an m × q regressor matrix. Let X_A be an m × p regressor matrix. Let W be a positive-definite m × m matrix.

Let d be any q-vector satisfying the normal equations for the GLS(W) regression of u upon Z, the macro regression.

Let b_A be any p-vector satisfying the normal equations for the GLS(W) regression of u upon X_A, the semi-aggregate regression.

Let C be any $q \times p$ matrix satisfying the normal equations for the GLS(W) regressions of X_A upon Z, the auxiliary regressions.

Let μ be a q-vector such that $\mu'd$ is a unique coefficient function in the macro regression.

(i) The p-vector $\lambda = C'\mu$ is uniquely determined by the auxiliary normal equations.

(ii) $\lambda'b_A$ is a unique coefficient function in the semi-aggregate regression.

(iii) If there exists a $p \times q$ matrix M such that $X_A M = Z$, then

$$\mu'd = \lambda'b_A .$$

(iv) If there exists a $p \times q$ matrix M such that $X_A M = Z$, then

$$M'\lambda = \mu . \quad \circ\circ$$

(i) Let λ_i be the i'th element of λ. Let C_i and X_i be the i'th column vectors of C and X_A. Since $\mu'd$ is a unique coefficient function in the macro regression, by P.2.2.4.A μ is in the row space of Z. Therefore by P.2.2.4.A $\lambda_i = \mu'C_i$ is a unique coefficient function in the auxiliary regression of X_i upon Z. The argument is repeated for $i = 1, \ldots, p$.

(ii) The demonstration is divided into two steps.

(ii:1) Since $\mu'd$ is a unique coefficient function in the macro regression, by P.2.2.4.B there exists at least one q-vector s such that $Z'WZs = \mu$. For any such vector s, by the auxiliary normal equations, $\lambda' = \mu'C = s'ZWZC = s'Z'WX_A$. Thus $\lambda' = t'X_A$ for $t = WZs$, i.e. λ is in the row space of X_A.

(ii:2) Therefore by P.2.2.4.A $\lambda'b_A$ is a unique coefficient function in the semi-aggregate regression.

(iii) The demonstration is divided into three steps.

(iii:1) Since $\mu'd$ is a unique coefficient function in the macro regression, by P.2.2.4.B there exists at least one q-vector s such that $Z'WZs = \mu$, and for any such vector s, $\mu'd = s'Z'Wu$. Further, for any such vector s, by the auxiliary normal equations, $s'Z'WX_A = s'Z'WZC = \mu'C = \lambda'$.

(iii:2) Since $X_A M = Z$ and by the semi-aggregated normal equations,

$Z'Wu = M'X_A'Wu = M'X_A'WX_Ab_A = Z'WX_Ab_A$.

(iii:3) Therefore, $\mu'd = s'ZWu = s'Z'WX_Ab_A = \lambda'b_A$.

(iv) Since $\mu'd$ is a unique coefficient function in the macro regression, by P.2.2.4.B there exists at least one q-vector s such that $Z'WZs = \mu$. For any such vector s, by the auxiliary normal equations and because $X_AM = Z$, $M'\lambda = M'C'\mu = M'C'Z'WZs = M'X_A'WZs = Z'WZs = \mu$. □

An analogous proposition is found in 4.4.3 below.

4.3.5 The weight-sum relations

If the aggregating functions G and H imply that $X_AM = Z$ for some matrix M, then by P.4.3.4 any unique macro coefficient function $\mu'd$ can be interpreted as a unique semi-aggregate coefficient function $\lambda'b_A$, where $\lambda = C'\mu$. The vector λ is also related to the vector μ by the equation $M'\lambda = \mu$.

In a practical situation, the macro regression data $[\,u \mid Z\,]$ and the matrix M may be all that is known. Then the auxiliary regression coefficients C cannot be computed, and the vector λ corresponding to a given vector μ cannot be determined in practice. In such situations, the relation $M'\lambda = \mu$ is helpful. It cannot in general be solved for λ, but it provides the following partial information.

Let M_i be the i'th column vector of M. Let μ_i be the i'th element of μ. Then for $i = 1, \ldots, q$,

$$M_i'\lambda = \mu_i .$$

These restrictions on the weight vector λ will be called the weight-sum relations.

4.3.6 An interpretation in two steps

Three kinds of regressions were introduced in 4.3.2 and 4.3.3, macro, semi-aggregate, and micro. There are three corresponding kinds of coefficient functions, macro $\mu'd$, semi-aggregate $\lambda'b_A$, and micro $\lambda_h'b_h$, $h = 1, \ldots, k$.

The desired model-free interpretation will be achieved, if at all, in two steps. The first step is as follows. A uniqe macro coefficient function $\mu'd$ is shown to be equal to a unique semi-aggregate coefficient function $\lambda'b_A$, where λ is determined by μ. The second step is as follows. A unique semi-aggregate coefficient function $\lambda'b_A$ is shown to be equal to the sum of micro coefficient functions

$\lambda_h' b_h$, $h = 1, \ldots, k$, where λ_h are determined by λ . Thus

$$\mu' d = \lambda' b_A = \sum_{h=1}^{k} \lambda_h' b_h ,$$

and the weight vectors λ_h of the implied micro coefficient functions are ultimately determined by the macro weight vector μ .

A simple first step of the interpretation requires a simple relation between the macro data $[\, u \,|\, Z \,]$ and the semi-aggregated data $[\, u \,|\, X_A \,]$. A fairly general class of such helpful relations has been found. If there exists a matrix M such that $Z = X_A M$, then P.4.3.4 provides the first step of the interpretation.

A simple second step of the interpretation requires a simple relation between the semi-aggregate data $[\, u \,|\, X_A \,]$ and the micro data $[\, y_h \,|\, X_h \,]$, $h = 1, \ldots, k$. No general class of such helpful relations has been found. Two very particular relations will be used in 5.1.3 and 6.3.3.

Whether the implied micro coefficient functions are unique is discussed in 4.4.5 below.

4.4 Interpretation by an expectational relation

4.4.1 The status of the four vector functions

Section 4.4 introduces an interpretation of macro regression coefficients by means of an expectational relation. Whenever this expectational interpretation is studied in chapters 4 to 8, the following applies.

The regressor and regressand aggregating functions $\{G, H\}$ are assumed to be given linear vector functions with numerically specified coefficients.

The micro relation Φ is assumed to consist of incomplete linear micro models in the sense of 4.1.2. The total micro parameter vector β is not assumed to be known.

No macro relation Ψ is assumed.

4.4.2 The macro statistics and the semi-aggregated model

The statistics to be interpreted are the GLS(W) macro regression coefficients d introduced in 4.3.2. The vector d satisfies the macro normal equations

$$Z'WZd = Z'Wu$$

but need not be unique. The positive-definite matrix W is arbitrary.

The interpretation is based on the semi-aggregated micro model introduced in 4.2.3, which is now written as follows.

$$H\Phi: \quad E_{H\Phi}(u) = X_A\beta_A \ .$$

The semi-aggregated micro parameter vector β_A is a known function of the total micro parameter vector β.

4.4.3 A general proposition

The macro regression coefficients d are now to be interpreted in semi-aggregated (micro) model terms. As in 4.3.4, attention is limited to unique macro coefficient functions $\mu'd$.

The following proposion confronts a regression and a linear model both involving the macro regressand vector u. The regression is of u upon the macro regressors Z. The model relates u to the semi-aggregated micro regressors X_A. The proposition interprets $\mu'd$ in terms of a semi-aggregated micro parameter function $\lambda'\beta_A$.

Proposition 4.4.3

Let u be a regressand m-vector. Let Z be an $m \times q$ regressor matrix. Let X_A be an $m \times p$ regressor matrix. Let W be a positive-definite $m \times m$ matrix.

Let d be any q-vector satisfying the normal equations for the GLS(W) regression of u upon Z, the macro regression.

Let u and X_A be related by the semi-aggregated micro model

$$E_{H\Phi}(u) = X_A\beta_A \ ,$$

whose parameter p-vector β_A is not assumed to be known.

Let C be any $q \times p$ matrix satisfying the normal equations for the GLS(W) regressions of X_A upon Z, the auxiliary regressions.

Let μ be a q-vector such that $\mu'd$ is a unique coefficient function in the macro regression.

(i) The p-vector $\lambda = C'\mu$ is uniquely determined by the auxiliary normal equations.

(ii) $\lambda'\beta_A$ is an estimable (or identified) parameter function in the semi-aggregated micro model.

(iii) The semi-aggregated micro model implies that

$$E_{H\Phi}(\mu'd) = \lambda'\beta_A .$$

(iv) If there exists a $p \times q$ matrix M such that $X_A M = Z$, then

$$M'\lambda = \mu . \quad \circ\circ$$

(i) The same demonstration as for P.4.3.4 (i).

(ii) The demonstration is divided into two steps.

(ii:1) As in (ii:1) of the demonstration for P.4.3.4 (ii), λ is in the row space of X_A.

(ii:2) Therefore by P.2.4.1.A (or P.2.3.2) $\lambda'\beta_A$ is an estimable (or identified) parameter function in the semi-aggregated micro model.

(iii) The demonstration is divided into three steps.

(iii:1) As in (iii:1) of the demonstration for P.4.3.4 (iii), for certain vectors s, $\mu'd = s'Z'Wu$ and $s'Z'WX_A = \lambda'$.

(iii:2) By the semi-aggregated micro model, $E_{H\Phi}(Z'Wu) = Z'WX_A\beta_A$.

(iii:3) Therefore, $E_{H\Phi}(\mu'd) = E_{H\Phi}(s'Z'Wu) = s'Z'WX_A\beta_A = \lambda'\beta_A$.

(iv) The same demonstration as for P.4.3.4 (iv). □

By the semi-aggregated micro model, $E_{H\Phi}(\mu'd) = \lambda'\beta_A$. If, in addition, the aggregating functions G and H are such that $X_A M = Z$ for some matrix M, then there are also the weight-sum relations $M'\lambda = \mu$. As indicated in 4.3.5, these may be helpful.

Essentially, P.4.4.3 is a generalization of the analysis of specification errors due to Theil [1957]. If the macro regressor vectors are linearly independent, every element of the macro coefficient vector d is unique. The auxiliary regression coefficients are also unique, $C = (Z'WZ)^{-1}Z'WX_A$. If further $W = I$, then by P.4.4.3 (iii),

$$E_{H\Phi}(d) = (Z'Z)^{-1}Z'X_A\beta_A .$$

This is the central result of Theil [1957].

4.4.4 An interpretation in two steps

The desired expectational interpretation is achieved in two steps. The first step is as follows. The expectation of a unique macro coefficient function $\mu'd$ is shown to be equal to an estimable semi-aggregated micro parameter function $\lambda'\beta_A$, where λ is determined by μ. The second step is as follows. An estimable semi-aggregated micro parameter function $\lambda'\beta_A$ is shown to be equal to the sum of micro parameter functions $\lambda_h'\beta_h$, $h = 1, \ldots, k$, where λ_h are determined by λ. Thus,

$$E_{H\Phi}(\mu'd) = \lambda'\beta_A = \sum_{h=1}^{k} \lambda_h'\beta_h ,$$

and the weight vectors λ_h of the implied micro parameter functions are ultimately determined by the macro weight vector μ.

The first step of the interpretation is always provided by P.4.4.3. It is important to note that the first step evaluates the expectation according to the semi-aggregated micro model $H\Phi$ of a macro coefficient function $\mu'd$. No macro model has been assumed.

The second step of the interpretation runs as follows. The elements of the semi-aggregated micro parameter vector β_A are known linear functions of the elements of the total micro parameter vector β, i.e. of the elements of the micro parameter vectors β_h, $h = 1, \ldots, k$. Therefore any semi-aggregated micro parameter function $\lambda'\beta_A$ is easily "separated" into a sum of k micro parameter functions $\lambda_h'\beta_h$.

Whether the implied micro parameter functions are estimable is discussed in 4.4.5.

4.4.5 On uniqueness and estimability

The model-free interpretation of 4.3.6 produces implied micro coefficent functions $\lambda_h'b_h$. The expectational interpretation of 4.4.4 produces implied micro parameter functions $\lambda_h'\beta_h$. It remains to discuss whether these are, respectively, unique and estimable (or identified).

First consider the expectational interpretation. By P.2.2.4.B $\mu'd = \pi'u$, where the vector π is determined by μ. The elements of the macro regressand vector u are known linear functions of the elements of the micro regressand

vectors y_h, $h = 1, \ldots, k$. Therefore any macro "estimator" $\pi'u$ is easily "separated" into a sum of k micro estimators $\tau_h'y_h$, where τ_h are determined by π. Thus,

$$\mu'd = \pi'u = \sum_{h=1}^{k} \tau_h'y_h ,$$

and the weight vectors τ_h of the implied micro estimators are ultimately determined by the macro weight vector μ. The implied micro estimators are not necessarily GLS estimators.

The expectational interpretation holds identically in the total micro parameter vector β. Therefore, it can be "separated" into components as follows.

$$E_{\Phi}(\tau_h'y_h) = \lambda_h'\beta_h , \quad h = 1, \ldots, k .$$

For each h, the h'th implied micro estimator is an unbiased estimator of the h'th implied micro parameter function. Thus by the definition in 2.4.1, the implied parameter functions are estimable.

Now consider the model-free interpretation, which assumes no models. For the sake of the argument, consider hypothetical linear micro models Φ. These imply a (hypothetical) expectational interpretation. The two interpretations are written as follows.

$$\mu'd = \sum_{h=1}^{k} \lambda_h^{M'} b_h .$$

$$E_{H\Phi}(\mu'd) = \sum_{h=1}^{k} \lambda_h^{E'} \beta_h .$$

By the argument above, the hypothetical implied micro parameter functions $\lambda_h^{E'} \beta_h$ are (hypothetically) estimable. Therefore, if

$$\lambda_h^{M} = \lambda_h^{E} , \quad h = 1, \ldots, k ,$$

then by P.2.4.1.B the implied micro coefficient functions $\lambda_h^{M'} b_h$ are unique.

It is not self-evident that the two sets of weight vectors λ_h^{M} and λ_h^{E} are identical. In the two cases where model-free interpretations are given below, in 5.1.3 and 6.3.3, they are identical. Further, in these two cases the implied micro "estimators" are GLS estimators.

4.4.6 Three critical assumptions

The following three assumptions are of critical importance for the interpretation in micro terms of a unique macro coefficient function $\mu'd$.

Assumption M. The macro and semi-aggregated micro regressor matrices are related by an equation $X_A M = Z$.

Assumption P. The semi-aggregate and micro regression data are so simply related that there is a known relation $b_A = \Sigma P_h b_h$ between the semi-aggregate and micro regression coefficients (if Q_h are chosen to fit W).

Assumption Φ. There are linear micro models such that $E_{H\Phi}(u) = X_A \beta_A$.

Assumption M guarantees a model-free interpretation $\lambda' b_A$ in semi-aggregate terms, and weight-sum relations $M'\lambda = \mu$.

Assumptions M and P jointly guarantee a model-free interpretation $\Sigma \lambda_h' b_h$ in micro terms. In favourable cases, $M'\lambda = \mu$ can be read as weight-sum relations for the micro weight vectors λ_h.

Assumption Φ guarantees an expectational interpretation $\Sigma \lambda_h' \beta_h$ in micro terms, but no weight-sum relations.

Assumptions Φ and M jointly produce $\Sigma \lambda_h' \beta_h$ and $M'\lambda = \mu$. In favourable cases, the latter relation can be read as weight-sum relations for the micro weight vectors λ_h.

4.4.7 References to literature

The larger part of Theil [1954] is concerned with expectational interpretation of the kind formulated in 4.4.4. This is however not clearly explained in the text, where it is said that macro parameters are "interpreted" as the expected macro regression coefficients. That Theil [1954] actually defines the macro parameters as the (micro-) expected macro regression coefficients is stated explicitly in Kloek [1961], Theil [1962] and Misra [1967], but only as it were in passing. It is emphasized in Lütjohann [1972]; cf. also Wu [1973].

That Theil [1954] is actually concerned with expectational interpretation is easier to see from the summary given in Theil [1971], section 11.3. The notation used there gives a clear indication of the fact. The term "macro parameter" is never mentioned.

The expectational interpretation assumes no macro model. The model-free interpretation formulated in 4.3.6 is the outcome of an attempt to eliminate the micro models too. Both kinds of interpretation were described in Lütjohann [1971].

5 A FORMAL ANALYSIS: AGGREGATION OF REGRESSORS OR OF REGRESSANDS

5.1 Linear transformation of regressors

5.1.1 Definition and notation

The micro data form the $n \times (1 + p)$ regression data matrix $[\, y \mid X \,]$, where $n \geq 1$ and $p \geq 1$. The macro data form the $n \times (1 + q)$ regression data matrix $[\, u \mid Z \,]$, where $q \geq 1$. For $j = 1, \ldots, n$, the j'th macro unit of analysis is identical to the j'th micro unit of analysis.

The regressor and regressand aggregating functions are

$$G: \quad Z = XG \,,$$

$$H: \quad u = y \,,$$

where G is a given $p \times q$ matrix of constants. Each macro regressor is a given linear function of the micro regressors.

Linear transformation of regressors includes cases that it may not be natural to call aggregations; cf. 3.1.3.

The micro and macro relations

$$\Phi: \quad y = X\beta + \epsilon \,; \; E(\epsilon) = 0 \,,$$

$$\Psi: \quad u = Z\delta + \eta \,; \; E(\eta) = 0$$

are sometimes considered.

Designed and observed regressors are sometimes distinguished; cf. 4.2.2. There are $p^{\Delta} \geq 0$ designed and $p^{O} \geq 0$ observed micro regressors $X = [\, X^{\Delta} \mid X^{O} \,]$, where $p^{\Delta} + p^{O} = p$. There are $q^{\Delta} \geq 0$ designed and $q^{O} \geq 0$ observed macro regressors $Z = [\, Z^{\Delta} \mid Z^{O} \,]$, where $q^{\Delta} + q^{O} = q$.

The regressor aggregating function is partitioned accordingly.

$$G: \quad [\, Z^{\Delta} \mid Z^{O} \,] = [\, X^{\Delta} \mid X^{O} \,] \begin{bmatrix} G^{\Delta} & G^{T} \\ \hline O & G^{O} \end{bmatrix} \,.$$

By definition, the designed macro regressors Z^{Δ} are not affected by the observed micro regressors X^{O}. The observed macro regressors Z^{O} may be affected (translated) by the designed micro regressors X^{Δ}.

The micro and macro parameter vectors are partitioned accordingly.

$$\Phi: \quad E(y) = X^{\Delta}\beta^{\Delta} + X^{O}\beta^{O} ,$$
$$\Psi: \quad E(u) = Z^{\Delta}\delta^{\Delta} + Z^{O}\delta^{O} .$$

The semi-aggregated regressor matrix is defined to be identical to the micro regressor matrix; cf. 4.2.3.

$$X_A = [\, X_A^{\Delta} \,\vdots\, X_A^{O} \,] = [\, X^{\Delta} \,\vdots\, X^{O} \,] = X .$$

The designed submatrices are identical. The observed submatrices are identical.

5.1.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are

$$H\Phi: \quad E_{H\Phi}(u) = \alpha_A + X_A^{O}\beta_A^{O} ,$$

$$\Psi G: \quad E_{\Psi G}(u) = \gamma_D + X_A^{O}\delta_D^{O} ,$$

where

$$\begin{bmatrix} \alpha_A \\ \hline \beta_A^{O} \end{bmatrix} = \begin{bmatrix} X^{\Delta}\beta^{\Delta} \\ \hline \beta^{O} \end{bmatrix} ,$$

$$\begin{bmatrix} \gamma_D \\ \hline \delta_D^{O} \end{bmatrix} = \begin{bmatrix} X^{\Delta}(G^{\Delta}\delta^{\Delta} + G^{T}\delta^{O}) \\ \hline G^{O}\delta^{O} \end{bmatrix} .$$

The following proposition answers the consistency problem. Note that the micro parameters β are critical and the macro parameters δ discretionary; cf. 4.2.7.

Proposition 5.1.2

Consider linear transformation of regressors, where the transformation matrix G is partitioned as above.

A necessary and sufficient condition for consistency to be attainable is that the following two statements are both true.

(i) The micro parameter subvector β^{O} associated with the p^{O} observed micro regressors is such that

$$\beta^O = G^O \delta^O$$

for at least one q^O-vector δ^O.

(ii) The designed micro regressor submatrix X^Δ and the micro parameter subvector β^Δ associated with the p^Δ designed micro regressors are such that, for some δ^O satisfying (i),

$$X^\Delta \beta^\Delta = X^\Delta (G^\Delta \delta^\Delta + G^T \delta^O)$$

for at least one q^Δ-vector δ^Δ. ∘∘

This is a special case of P.4.2.6. □

Statement (i) taken alone is a necessary condition for consistency to be attainable. The micro parameter subvector β^O must be in the column space of G^O.

5.1.3 A model-free interpretation

The macro regression data $[\, u \mid Z \,]$ and the semi-aggregate regression data $[\, u \mid X_A \,]$ are related by the equation $X_A G = Z$.

The semi-aggregate regression data $[\, u \mid X_A \,]$ and the micro regression data $[\, y \mid X \,]$ are identical. Thus any vector b_A of GLS(W) semi-aggregate regression coefficients is also a vector b of GLS(W) micro regression coefficients, and conversely; cf. 4.3.6.

The following proposition provides a model-free interpretation. Note that, by P.4.3.4, the semi-aggregate coefficient function $\lambda' b_A$ is unique.

Proposition 5.1.3

Consider linear transformation of regressors. Let d, b_A and b by any vectors of GLS(W) macro, semi-aggregate, and micro regression coefficients, respectively. Let C be any matrix of GLS(W) auxiliary regression coefficients.

Let μ be a q-vector such that $\mu' d$ is a unique macro coefficient function. Define the p-vector $\lambda = C' \mu$. Then:

(i) $\mu' d = \lambda' b_A = \lambda' b$.

(ii) $G' \lambda = \mu$. ∘∘

This is a special case of P.4.3.4, except for the second equation of (i), which follows because $b_A = b$. □

In the terms of 4.3.2, this model-free interpretation chooses $Q = W$.

5.1.4 An expectational interpretation

The micro model $E(y) = X\beta$ implies the semi-aggregated micro model

$$H\Phi: \quad E_{H\Phi}(u) = X_A\beta_A ,$$

where $X_A = X$ and $\beta_A = \beta$. As before, $X_A G = Z$.

The following proposition provides an expectational interpretation. Note that, by P.4.4.3, the semi-aggregated micro parameter function $\lambda'\beta_A$ is estimable.

Propositon 5.1.4

Consider linear transformation of regressors. Let d be any vector of GLS(W) macro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

Let μ be a q-vector such that $\mu'd$ is a unique macro coefficient function. Define the p-vector $\lambda = C'\mu$. Then:

(i) $E_{H\Phi}(\mu'd) = \lambda'\beta_A = \lambda'\beta$.

(ii) $G'\lambda = \mu$. ∘∘

This is a special case of P.4.4.3, except for the second equation of (i), which follows because $\beta_A = \beta$. □

The implied micro parameter function $\lambda'\beta$ in P.5.1.4 has the same weight vector λ as the implied micro coefficient function $\lambda'b$ in P.5.1.3; cf. 4.4.5.

5.1.5 An example

Let the micro and macro regression data matrices be as follows.

$$[\,y \mid X\,] = [\,y \mid X^{\Delta} \mid X^{O}\,] = [\,y \mid j \mid x_1 \;\; x_2\,] ,$$

$$[\,u \mid Z\,] = [\,u \mid Z^{\Delta} \mid Z^{O}\,] = [\,u \mid j \mid z\,] .$$

The intercept regressor is the only designed micro and macro regressor. There are two observed micro regressors and one observed macro regressor. The regressand aggregating function H is $u = y$. Let the regressor aggregating function G be as follows.

$$[\,j \mid z\,] = [\,Z^{\Delta} \mid Z^{O}\,] = [\,X^{\Delta} \mid X^{O}\,] \begin{bmatrix} G^{\Delta} & G^{T} \\ O & G^{O} \end{bmatrix} =$$

$$= [\,j \mid x_1 \;\; x_2\,] \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} .$$

For simplicity, let the rank of Z be $q = 2$. Then all macro and auxiliary regression coefficients are unique.

The micro and macro parameters and GLS(W) regression coefficients are denoted as follows.

$$[\,\beta \mid b\,] = \begin{bmatrix} \beta^{\Delta} & b^{\Delta} \\ \beta^{O} & b^{O} \end{bmatrix} = \begin{bmatrix} \beta^{\Delta} & b^{\Delta} \\ \beta_1 & b_1 \\ \beta_2 & b_2 \end{bmatrix} ,$$

$$[\,\delta \mid d\,] = \begin{bmatrix} \delta^{\Delta} & d^{\Delta} \\ \delta^{O} & d^{O} \end{bmatrix} .$$

The GLS(W) auxiliary regression coefficients are denoted as follows.

$$C = \begin{bmatrix} 1 & c_{1j} & c_{2j} \\ 0 & c_{1z} & c_{2z} \end{bmatrix} .$$

The auxiliary regressions are of X upon Z.

Consider the consistency problem. By P.5.1.2 consistency is attainable if and only if there exist scalars δ^{O} and δ^{Δ} such that

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta^{O} = G^{O}\delta^{O} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \delta^{O} ,$$

$$j\beta^{\Delta} = X^{\Delta}\beta^{\Delta} = X^{\Delta}\,(G^{\Delta}\delta^{\Delta} + G^{T}\delta^{O}) = j\delta^{\Delta} .$$

The critical condition is the first one, which requires that $\beta_2 = 2\beta_1$.

Consider the problem of interpreting the macro coefficient function

$$\mu' d = [\,0 \quad 1\,] \begin{bmatrix} d^{\Delta} \\ d^{O} \end{bmatrix} = d^{O} ,$$

i.e. the macro slope. The implied micro coefficient and parameter functions are denoted as follows.

$$\lambda' [\, b \mid \beta \,] = [\, \lambda^{\Delta'} \mid \lambda^{O'} \,] \left[\begin{array}{c|c} b^{\Delta} & \beta^{\Delta} \\ \hline b^{O} & \beta^{O} \end{array}\right] = [\, \lambda^{\Delta} \mid \lambda_1 \lambda_2 \,] \left[\begin{array}{c|c} b^{\Delta} & \beta^{\Delta} \\ \hline b_1 & \beta_1 \\ b_2 & \beta_2 \end{array}\right] .$$

The weight vector λ considered is $\lambda = C'\mu$, i.e.

$$\begin{bmatrix} \lambda^{\Delta} \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c_{1j} & c_{1z} \\ c_{2j} & c_{2z} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ c_{1z} \\ c_{2z} \end{bmatrix} .$$

By P.5.1.3 and P.5.1.4,

$$d^{O} = c_{1z} b_1 + c_{2z} b_2 ,$$

$$E_{H\Phi}(d^{O}) = c_{1z}\beta_1 + c_{2z}\beta_2 .$$

Further, there are the weight-sum relations $G'\lambda = \mu$, i.e.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} \lambda^{\Delta} \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \lambda^{\Delta} \\ \lambda_1 + 2\lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

That $\lambda^{\Delta} = 0$ is seen from $\lambda = C'\mu$ even if the observed micro data are unknown. The second weight-sum relation says that

$$\lambda_1 + 2\lambda_2 = c_{1z} + 2c_{2z} = 1$$

whatever the observed micro data.

5.1.6 A special case: Omission of regressors

Omission of regressors is a special linear transformation of regressors. For simplicity, let there be a single designed regressor, the intercept regressor. The first $q^{O} < p^{O}$ observed micro regressors X_Z^O are retained as macro regressors. The remaining micro regressors X_O^O are simply omitted. The regressor aggregating function G is as follows,

$$[\,j \mid Z^O\,] = [\,j \mid X_Z^O \;\; X_O^O\,] \begin{bmatrix} 1 & 0 \\ 0 & I^O \\ 0 & 0 \end{bmatrix} ,$$

where I^O is the unit matrix of order q^O. Omission of regressors is hardly a case of aggregation, but is of interest nevertheless.

The micro parameter subvectors associated with X_Z^O and X_O^O are denoted β_Z^O and β_O^O. The GLS(W) micro regression coefficient subvectors are indexed analogously, and so are the subvectors of λ. The matrix of GLS(W) auxiliary regression coefficients can be taken to be

$$C = \begin{bmatrix} 1 & 0 & c'_{Oj} \\ 0 & I^O & C_{OZ} \end{bmatrix} ,$$

also when C is not unique.

Consider the consistency problem. By P.5.1.2, a necessary condition for consistency to be attainable is that there exists a q^O-vector δ^O such that

$$\begin{bmatrix} \beta_Z^O \\ \beta_O^O \end{bmatrix} = \begin{bmatrix} I^O \\ 0 \end{bmatrix} \delta^O .$$

The condition requires that $\beta_O^O = 0$; the omission of relevant regressors causes inconsistency.

Consider the problem of interpreting the q^O macro GLS(W) regression coefficient functions

$$M'd = [\,O \mid I^O\,] \begin{bmatrix} d^\Delta \\ \hline d^O \end{bmatrix} = d^O ,$$

i.e. the macro slopes. The weight vectors λ of the q^O implied macro coefficient and parameter functions are the q^O column vectors of the $p \times q^O$ matrix $\Lambda = C'M$.

$$\Lambda = \begin{bmatrix} \Lambda^\Delta \\ \Lambda_Z^O \\ \Lambda_O^O \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I^O \\ c_{Oj} & C'_{OZ} \end{bmatrix} \begin{bmatrix} 0 \\ I^O \end{bmatrix} = \begin{bmatrix} 0 \\ I^O \\ C'_{OZ} \end{bmatrix}$$

In this case, the weight-sum relations $G'\Lambda = M$ provide no additional information.

If the macro slopes are unique, then by P.5.1.3 and P.5.1.4,

$$d^O = \Lambda' b = b_Z^O + C_{OZ} b_O^O ,$$

$$E_{H\Phi}(d^O) = \Lambda' \beta = \beta_Z^O + C_{OZ} \beta_O^O .$$

The elements of the $q^O \times (p^O - q^O)$ matrix C_{OZ} are the slopes in the GLS(W) auxiliary regression of the omitted X_O^O on the retained X_Z^O.

Omission of regressors is an error of specification. Its consequences are shown by the two interpretative relations. A least the expectational relation is due to Theil [1957]. Both relations are given by Goldberger [1964], section 4.10. Some applications are found in Griliches [1957], Nerlove [1958b], Goldberger [1961], Haitovsky [1966], and Box [1966].

5.2 Unweighted partitioned aggregation of observed regressors

5.2.1 Definition and notation

Unweighted partitioned aggregation of observed regressors will now be defined; cf. 3.1.7. What was said in 5.1.1 - 5.1.4 applies, and is specialized as follows.

The p micro regressors are partitioned into q disjoint, exhaustive families. For $i = 1, \ldots q$, the i'th macro regressor is the sum of the members of the i'th family of micro regressors. Some families have a single member, a bachelor regressor. The p^Δ designed micro regressors X^Δ and the first r observed micro regressors X_B^O are bachelor regressors, and are "aggregated" into $q^\Delta = p^\Delta$ and r macro regressors $Z^\Delta = X^\Delta$ and $Z_B^O = X_B^O$. The remaining $p^O - r$ observed micro regressors form $s = q^O - r$ families. For $h = 1, \ldots, s$, the h'th family X_h^O has $p_h^O \geq 2$ members, and their sum is the macro regressor z_h^O.

Let I^Δ and I^B denote the unit matrices of orders p^Δ and r. For $h = 1, \ldots, s$, let i_h denote the p_h^O-vector of unit elements. The regressor aggregating function G is as follows.

$$[\,Z^{\Delta} \mid Z_B^O z_1^O \ldots z_s^O\,] = [\,X^{\Delta} \mid X_B^O X_1^O \ldots X_s^O\,] \left[\begin{array}{c|ccc} I^{\Delta} & O & O & \ldots\ldots\ O \\ \hline O & I^{B} & O & \ldots\ldots\ O \\ O & O & i_1 & \ldots\ldots\ O \\ \vdots & \vdots & \vdots & \vdots \\ O & O & O & \ldots\ldots\ i_s \end{array}\right]$$

The transformation matrix G is block-diagonal.

The subvectors of b associated with X_B^O and X_h^O are denoted b_B^O and b_h^O, and similarly for β and λ. The subvector and elements of d associated with Z_B^O and z_h^O are denoted d_B^O and d_h^O, and similarly for δ.

The matrix of auxiliary regression coefficients can always be taken to be

$$C = \begin{bmatrix} I^{\Delta} & O & C_1^{\Delta} & \ldots\ldots & C_s^{\Delta} \\ O & I^{B} & C_1^{B} & \ldots\ldots & C_s^{B} \\ O & O & c'_{11} & \ldots\ldots & c'_{s1} \\ \vdots & \vdots & \vdots & & \vdots \\ O & O & c'_{1s} & \ldots\ldots & c'_{ss} \end{bmatrix},$$

also when C is not unique. The submatrix c'_{hk} is a row vector of order p_h^o.

5.2.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro models are as in 5.1.2, where now

$$\begin{bmatrix} \alpha_A \\ \hline \\ \beta_A^O \\ \\ \\ \end{bmatrix} = \begin{bmatrix} X^{\Delta}\beta^{\Delta} \\ \hline \beta_B^O \\ \beta_1^O \\ \vdots \\ \beta_s^O \end{bmatrix}, \qquad \begin{bmatrix} \gamma_D \\ \hline \\ \delta_D^O \\ \\ \\ \end{bmatrix} = \begin{bmatrix} X^{\Delta}\delta^{\Delta} \\ \hline \delta_B^O \\ i_1\,\delta_1^O \\ \vdots \\ i_s\,\delta_s^O \end{bmatrix}.$$

Proposition 5.2.2

Consider unweighted partitioned aggregation of observed regressors.

A necessary and sufficient condition for consistency to be attainable is that for $h = 1, \ldots, s$, the micro parameter subvectors β_h^O are such that

$$\beta_h^O = i_h \delta_h^O$$

for some scalars δ_h^O. ∘∘

This is a special case of P.5.1.2. Some parts of the condition stated there are automatically fulfilled, and have been omitted. □

The consistency condition can be reformulated in words. For any non-bachelor family of observed micro regressors, the micro parameters associated with the members must all be equal.

5.2.3 Model-free and expectational interpretations

Interpretations will be given for two kinds of macro regression coefficients. The first kind d_k^O is associated with a macro regressor that is the sum of $p_k^O \geq 2$ observed micro regressors; $k = 1, \ldots, s$.

Proposition 5.2.3.A

Consider unweighted partitioned aggregation of observed regressors. Let d and b be any vectors of GLS(W) macro and micro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

If d_k^O is a unique coefficient function in the macro regression, the following three statements are true.

(i) $$d_k^O = \sum_{h=1}^{s} c_{hk}' b_h^O \quad .$$

(ii) $$E_{H\Phi}(d_k^O) = \sum_{h=1}^{s} c_{hk}' \beta_h^O \quad .$$

(iii) $$i_h' c_{hk} = \begin{cases} 1 & \text{if} \quad h = k \, , \\ 0 & \text{if} \quad h \neq k \, . \end{cases}$$ ∘∘

Let μ_k be a q-vector whose $(q^{\Delta} + r + k)$'th element is unity, while all other elements are zero. Then $d_k^O = \mu_k' d$, and $\lambda_k = C' \mu_k$ is the transposed

$(q^{\Delta} + r + k)$'th row vector of C .

Statements (i), (ii) and (iii) are special cases of P.5.1.3(i), of P.5.1.4(i), and of P.5.1.3(ii) = P.5.1.4(ii), respectively. □

The second kind of macro regression coefficient d_m is associated with a bachelor regressor. The corresponding micro regression coefficient and parameter are b_m and β_m . For any $m = 1, \ldots, p^{\Delta} + r$, let the m'th row vector of C be denoted

$$[t'_m \quad c'_{1m} \ldots\ldots\ldots c'_{sm}] ,$$

where t'_m is the m'th row vector of the unit matrix of order $p^{\Delta} + r$, while c'_{hm} is a row vector of order p_h^O . No distinction is made between designed and observed bachelor regressors.

Proposition 5.2.3.B

The same assumptions as in P.5.2.3.A.

If d_m is a unique coefficient function in the macro regression, the following three statements are true.

(i) $\quad d_m = b_m + \sum_{h=1}^{s} c'_{hm} b_h^O$.

(ii) $\quad E_{H\Phi}(d_m) = \beta_m + \sum_{h=1}^{s} c'_{hm} \beta_h^O$.

(iii) $\quad i'_h c_{hm} = 0 , \quad h = 1, \ldots, s .$ ○○

Let μ_m be a q-vector whose m'th element is unity, while all other elements are zero. Then $d_m = \mu'_m d$, and $\lambda_m = C'\mu_m$ is the transposed m'th row vector of C .

The same demonstrations as for P.5.2.3.A. □

The expectational interpretations will now be described in words. Recall the concepts of corresponding and non-corresponding families introduced in 3.1.7. The description is incomplete, but is valid for both P.5.2.3.A and P.5.2.3.B.

The micro expectation of a unique macro coefficient is the sum of q weighted sums of micro parameters, one for each family of micro regressors. The sum of the weights for the corresponding family is unity, also when this family has a

single member. The sum of the weights for any non-corresponding family is zero, also when the family has a single member. The micro parameters of non-corresponding bachelor regressors are therefore never effectively included.

5.2.4 On the connection with Theil's analysis

The expectational interpretation of 5.2.3 very much resembles Theil [1954], Theorem 1. The weights are auxiliary regression coefficients exactly as in Theil. The weight-sum relations are the same as in Theil.

An important difference is that 5.2 is about partitioned aggregation of regressors, while Theil's Theorem 1 is about simple aggregation of aspects. As was pointed out in 3.2.2, simple aggregation of aspects is an aggregation of regressands followed by a partitioned aggregation of regressors. As will be seen in 5.3 and 7.2, the aggregation of regressands is quite harmless, and it is the aggregation of regressors that causes all the difficulties.

A less important difference is that Theil [1954] treats all regressors alike, except the intercept regressor. As was seen above, as far as interpretation is concerned, all bachelor regressors behave analogously. The intercept regressor is the most frequently occurring bachelor regressor. Theil's formulation is correct but less informative.

5.3 Aggregation of regressands

5.3.1 Definition and notation

The micro data form $k \geq 2$ regression data matrices $[\, y_i \,\vdots\, X_i \,]$ of orders $n \times (1 + p_i)$, where $n \geq 1$ and $p_i \geq 1$, $i = 1, \ldots, k$. Let the i'th set of micro regressor vectors be called R_i. The union of the k sets R_i has $p \geq \max p_i$ members. These form the $n \times p$ total micro regressor matrix X; cf. 3.1.4. The k micro regressand vectors form the $n \times k$ micro regressand matrix Y.

The macro data form the $n \times (1 + p)$ regression data matrix $[\, u \,\vdots\, Z \,]$. For $j = 1, \ldots, n$, the j'th macro unit of analysis is identical to the j'th micro unit of analysis.

The regressor and regressand aggregating functions are

$$G: \quad Z = X ,$$

$$H: \quad u = Yh ,$$

where h is a given k-vector of constants. The macro regressand is a given linear function of the micro regressands. The set of macro regressors is the union of the sets of micro regressors.

The micro and macro relations

$$\Phi: \quad y_i = X_i\beta_i + \varepsilon_i , \; E(\varepsilon_i) = 0 , \; i = 1, \ldots, k ,$$

$$\Psi: \quad u = Z\delta + \eta , \quad E(\eta) = 0$$

are sometimes considered. Alternatively, the k micro models are expressed in terms of the common total micro regressor matrix as follows,

$$\Phi: \quad y_i = X\beta_i^* + \varepsilon_i , \; E(\varepsilon_i) = 0 , \; i = 1, \ldots, k .$$

The augmented micro parameter vectors β_i^* are defined as follows. Those elements of β_i^* that are associated with the set R_i of regressors, are equal to the corresponding elements of β_i. All other elements of β_i^* are defined to be zero.

The k augmented micro parameter vectors β_i^* form the $p \times k$ augmented micro parameter matrix B^*. The k micro models are written jointly

$$\Phi: \quad E(Y) = XB^* .$$

Sometimes, $p^\Delta \geq 0$ designed regressors $X^\Delta = Z^\Delta$ and $p^O \geq 0$ observed regressors $X^O = Z^O$ are distinguished, where $p^\Delta + p^O = p$. The augmented micro parameter matrix and the macro parameter vector are partitioned accordingly.

$$\Phi: \quad E(Y) = X^\Delta B^{*\Delta} + X^O B^{*O} ,$$

$$\Psi: \quad E(u) = Z^\Delta \delta^\Delta + Z^O \delta^O .$$

The semi-aggregated regressor matrix is defined to be identical to the total micro regressor matrix.

$$X_A = [\, X_A^\Delta \,\vdots\, X_A^O \,] = [\, X^\Delta \,\vdots\, X^O \,] = X .$$

The designed and observed submatrices are identical, respectively.

5.3.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are $E_{H\Phi}(u) = \alpha_A + X_A^O \beta_A^O$, and $E_{\Psi G}(u) = \gamma_D + X_A^O \delta_D^O$, where

$$\begin{bmatrix} \alpha_A \\ \hline \beta_A^O \end{bmatrix} = \begin{bmatrix} X^\Delta B^{*\Delta} h \\ \hline B^{*O} h \end{bmatrix} ,$$

$$\begin{bmatrix} \gamma_D \\ \hline \delta_D^O \end{bmatrix} = \begin{bmatrix} X^\Delta \delta^\Delta \\ \hline \delta^O \end{bmatrix} .$$

The following proposition answers the consistency problem.

Proposition 5.3.2

In aggregation of regressands, consistency is always attainable. ○○

P.4.2.6 applies. For any B^*, consistency is attained by choosing $\delta = B^* h$. □

5.3.3 An expectational interpretation

The micro models $E(Y) = XB^*$ imply the semi-aggregated micro model $E_{H\Phi}(u) = X_A \beta_A$, where $X_A = X$ and $\beta_A = B^* h$. Further, $X_A = Z$.

For any p-vector μ, let μ_i be the subvector corresponding to the set R_i of regressors. Let h_i be the i'th element of h. The following proposition provides an expectational interpretation.

Proposition 5.3.3

Consider aggregation of regressands. Let d be any vector of GLS(W) macro regression coefficients.

Let μ be a p-vector such that $\mu' d$ is a unique macro coefficient function. For $i = 1, \ldots, k$ define the p_i-vector $\lambda_i = h_i \mu_i$. Then:

$$E_{H\Phi}(\mu' d) = \mu' \beta_A = \sum_{i=1}^{k} \lambda_i' \beta_i . \quad ○○$$

Since $X_A = Z$, one can take the matrix of auxiliary regression coefficients to be $C = I_p$. Then by P.4.4.3, $E_{H\Phi}(\mu' d) = \mu' \beta_A$. Further, from $\beta_A = B^* h = \Sigma h_i \beta_i^*$

it follows that $\mu' \beta_A = \Sigma h_i \mu' \beta_i^* = \Sigma \lambda_i' \beta_i$. □

In particular, if $h = j_k$, then the weight vector λ_i of the i'th implied micro parameter function is simply the subvector of μ corresponding to the set R_i of regressors.

Except for special cases such as when all R_i are equal, there is no simple model-free interpretation.

6 A FORMAL ANALYSIS: AGGREGATION OF UNITS OF ANALYSIS OR OF SETS OF UNITS

6.1 Linear transformation of units of analysis

6.1.1 Definition and notation

The micro data form the $n \times (1 + p)$ regression data matrix $[\, y \,\vdots\, X \,]$, where $n \geq 1$ and $p \geq 1$. The macro data form the $m \times (1 + p)$ regression data matrix $[\, u \,\vdots\, Z \,]$, where $m \geq 1$. For $i = 1, \ldots, p$, the i'th macro regressor corresponds to the i'th micro regressor. Designed and observed regressors are distinguished. There are $p^{\Delta} \geq 0$ designed and $p^{O} \geq 0$ observed regressors, where $p^{\Delta} + p^{O} = p$; $X = [\, X^{\Delta} \,\vdots\, X^{O} \,]$ and $Z = [\, Z^{\Delta} \,\vdots\, Z^{O} \,]$.

For simplicity, in the general case of aggregation of units of analysis in 6.1 there are assumed to be no mixed regressors. A particular class of mixed regressors will be admitted in the special case in 6.2.

The regressor and regressand aggregating functions are

$$G: \begin{cases} Z^{\Delta} = g^{\Delta}(X^{\Delta}) \,, \\ Z^{O} = T' X^{O} \,, \end{cases}$$

$$H: \quad u = T' y \,,$$

where g^{Δ} is a given matrix function and T is a given $n \times m$ matrix of constants. As far as the regressand and the observed regressors are concerned, each macro unit of analysis is a given linear function of the micro units of analysis. The function g^{Δ} will be left unspecified, but is sometimes also linear; cf. 3.1.5.

Linear transformation of units of analysis includes cases that it may not be natural to call aggregations; cf. 3.1.5.

The micro and macro relations

$$\Phi: \quad y = X^{\Delta}\beta^{\Delta} + X^{O}\beta^{O} + \varepsilon \,, \quad E(\varepsilon) = 0 \,,$$

$$\Psi: \quad u = Z^{\Delta}\delta^{\Delta} + Z^{O}\delta^{O} + \eta \,, \quad E(\eta) = 0$$

are sometimes considered.

The semi-aggregated regressor matrix and its two submatrices are defined to be as follows.

$$X_A = [\, X_A^{\Delta} \,\vdots\, X_A^{O} \,] = [\, T'X^{\Delta} \,\vdots\, T'X^{O} \,] = T'X \,.$$

The observed semi-aggregated regressor submatrix X_A^O is thus identical to the observed macro regressor submatrix Z^O. The designed submatrices X_A^{Δ} and Z^{Δ} may differ.

6.1.2 On the meaning of full subvector rank

The consistency problem refers to a set Ξ of independent micro data vectors x. As explained in 4.2.4, the vector x is a function of the observed semi-aggregated regressor submatrix X_A^O. In aggregation of units of analysis, the latter matrix is the function $X_A^O = T'X^O$ of the observed micro regressor submatrix X^O. The set Ξ is specified to be of full subvector rank. The meaning of this specification in terms of X_A^O is stated in 4.2.4. Its meaning in terms of the observed micro regressor data X^O is not immediately obvious.

To begin with, a particular set of matrices X^O will be described. Next, the corresponding set of matrices X_A^O will be shown to satisfy that part of the specification of Ξ which refers to the first row vector of X_A^O. Finally, a generalization will be indicated.

Let $z_1^{O'}$ be the first row p^O-vector of $X_A^O = Z^O$. Let t_1 be the first column n-vector of T. Let t_1 and X^O be partitioned as follows.

$$z_1^{O'} = t_1' X^O = (t_{11} \,\vdots\, t_{12}') \begin{bmatrix} x_1^{O'} \\ \hdashline X_2^O \end{bmatrix} \,,$$

where $x_1^{O'}$ is the first row p^O-vector of X^O. Without loss of generality, assume that $t_{11} \neq 0$.

Consider $p^O + 1$ different matrices

$$X_h^O = \begin{bmatrix} x_{1h}^{O'} \\ \hdashline X_2^O \end{bmatrix} \,, \qquad h = 1, \ldots, p^O + 1$$

which differ as to their first rows only. Assume that the vectors

$$(1 \,\vdots\, x_{1h}^{O'}) \,, \qquad h = 1, \ldots, p^O + 1$$

are linearly independent. These row vectors form the $(p^O + 1) \times (p^O + 1)$ matrix

$$[\, j \,\vdots\, X_1 \,] \ ,$$

where j is the $(p^O + 1)$-vector of unit elements.

Each matrix X_h^O implies a vector $z_{1h}^{O'}$. Define $z_2' = t_{12}' X_2^O$. Then

$$(1 \,\vdots\, z_{1h}^{O'}) = (1 \,\vdots\, t_{11} x_{1h}^{O'} + z_2') \ , \quad h = 1, \ldots, p^O + 1 \ .$$

These row vectors form the $(p^O + 1) \times (p^O + 1)$ matrix

$$[\, j \,\vdots\, Z_1 \,] = [\, j \,\vdots\, t_{11} X_1 + j z_2' \,] \ ,$$

where j is as above.

Proposition 6.1.2

Notation and assumptions as in the text. In particular, $t_{11} \neq 0$, and the $p^O + 1$ row vectors of $[\, j \,\vdots\, X_1 \,]$ are linearly independent.

Then the $p^O + 1$ row vectors of $[\, j \,\vdots\, Z_1 \,]$ are linearly independent. □

Assume the opposite! Then there exists a vector $v \neq 0$ such that $v' [\, j \,\vdots\, Z_1 \,] = (0 \,\vdots\, 0)$. Consequently, $0 = v' Z_1 = t_{11} v' X_1 + v' j z_2' = t_{11} v' X_1$. Thus since $t_{11} \neq 0$, $v' [\, j \,\vdots\, X_1 \,] = (0 \,\vdots\, 0)$, and the row vectors of $[\, j \,\vdots\, X_1 \,]$ are linearly dependent contrary to assumption.

Thus the row vectors of $[\, j \,\vdots\, Z_1 \,]$ are linearly independent. □

The demonstration does not require that $z_2' \neq 0$ or $t_{12} \neq 0$.

The reasoning is easily generalized. For any given $j = 1, \ldots, m$, a set M_j of observed micro regressor submatrices X^O can be indicated such that the corresponding set of observed semi-aggregated regressor matrices X_A^O satisfies that part of the specification of Ξ which refers to the j'th row vector of X_A^O. The union of the sets M_j implies a set of matrices X_A^O of full subvector rank.

6.1.3 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are not quite as stated in 4.2.3. There, $H\Phi$ and ΨG express u as two different functions of $[\, X_A^\Delta \,\vdots\, X_A^O \,]$. In aggregation of units of analysis, $H\Phi$ and ΨG instead express

u as two different functions of $[\, X^{\Delta} \,\vdots\, X^{O}_{A} \,]$. The two relations are $E_{H\Phi}(u) = \alpha_A + X^{O}_{A}\beta^{O}_{A}$ and $E_{\Psi G}(u) = \gamma_D + X^{O}_{A}\delta^{O}_{D}$, where

$$\begin{bmatrix} \alpha_A \\ \hdashline \beta^{O}_{A} \end{bmatrix} = \begin{bmatrix} T'X^{\Delta}\beta^{\Delta} \\ \hdashline \beta^{O} \end{bmatrix} ,$$

$$\begin{bmatrix} \gamma_D \\ \hdashline \delta^{O}_{D} \end{bmatrix} = \begin{bmatrix} g^{\Delta}(X^{\Delta})\,\delta^{\Delta} \\ \hdashline \delta^{O} \end{bmatrix} .$$

The following proposition answers the consistency problem.

Proposition 6.1.3

Consider linear transformation of units of analysis.

A necessary and sufficient condition for consistency to be attainable is as follows. The designed micro regressor submatrix X^{Δ} and the associated micro parameter subvector β^{Δ} are such that

$$T'X^{\Delta}\beta^{\Delta} = g^{\Delta}(X^{\Delta})\,\delta^{\Delta}$$

for at least one p^{Δ}-vector δ^{Δ}. ⚇

This is a special case of P.4.2.6. □

If there are no designed micro or macro regressors, consistency is always attained by choosing $\delta^{O} = \beta^{O}$. Cf. the end of 6.1.4.

6.1.4 An expectational interpretation

The micro model $E(y) = X^{\Delta}\beta^{\Delta} + X^{O}\beta^{O}$ implies the semi-aggregated micro model $E_{H\Phi}(u) = X^{\Delta}_{A}\beta^{\Delta}_{A} + X^{O}_{A}\beta^{O}_{A}$, where $\beta^{\Delta}_{A} = \beta^{\Delta}$, and $\beta^{O}_{A} = \beta^{O}$.

If the matrix C of auxiliary regression coefficients is partitioned conformably with $X_A = [\, X^{\Delta}_{A} \,\vdots\, X^{O}_{A} \,]$ and $Z = [\, Z^{\Delta} \,\vdots\, Z^{O} \,]$, then because $X^{O}_{A} = Z^{O}$, C can always be taken to be

$$C = \begin{bmatrix} C^{\Delta\Delta} & \vdots & O \\ \hdashline C^{\Delta O} & \vdots & I^{O} \end{bmatrix} ,$$

where I^{O} is the unit matrix of order p^{O}.

The vectors λ and μ are partitioned conformably with X_A and Z, and the subvectors are denoted accordingly.

The following proposition provides an expectational interpretation. There are however no weight-sum relations for λ, because there is no matrix M such that $X_A M = Z$; cf. 4.4.6.

Proposition 6.1.4

Consider linear transformation of units of analysis. Let d be any vector of GLS(W) macro regression coefficients. Let $[\, C^{\Delta\Delta'} \,\vdots\, C^{\Delta O'} \,]'$ be any matrix satisfying the normal equations of the GLS(W) auxiliary regressions of X_A^{Δ} upon $[\, Z^{\Delta} \,\vdots\, Z^{O} \,]$.

Let μ be a p-vector such that $\mu' d = \mu^{\Delta'} d^{\Delta} + \mu^{O'} d^{O}$ is a unique macro coefficient function. Define the p^{Δ}- and p^{O}-vectors

$$\begin{bmatrix} \lambda^{\Delta} \\ \hline \lambda^{O} \end{bmatrix} = \begin{bmatrix} C^{\Delta\Delta'}\mu^{\Delta} + C^{\Delta O'}\mu^{O} \\ \hline \mu^{O} \end{bmatrix} .$$

Then

$$E_{H\Phi}(\mu' d) = \lambda^{\Delta'} \beta_A^{\Delta} + \lambda^{O'} \beta_A^{\Delta} = \lambda^{\Delta'} \beta^{\Delta} + \lambda^{O'} \beta^{O} . \quad \circ\circ$$

This is a special case of P.4.4.3, except for the second equation which follows because $\beta_A^{\Delta} = \beta^{\Delta}$ and $\beta_A^{O} = \beta^{O}$. □

If there are no designed micro or macro regressors, the interpretation is simple, $\lambda = \mu$.

The results of 6.1.3 and 6.1.4 may be summed up as follows. In linear transformation of units of analysis as delimited in 6.1.1, it is the designed regressors that cause all the difficulties. Normally there is at least one designed micro and macro regressor, the intercept regressor.

6.1.5 An example

There are n observations (y_t, x_t), $t = 1, 2, \ldots, n$, generated by the following linear model.

$$y_t = \alpha + \beta x_t + \varepsilon_t ,$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \omega_t ,$$

$$E(\omega_t) = 0 ,$$

$$E(\omega_t^2) = \tau^2 ,$$

$$E(\omega_s \omega_t) = 0 \text{ unless } s = t ,$$

$$|\rho| < 1 .$$

The disturbances ε_t follow a first order autoregressive scheme, whose parameter ρ is assumed to be known.

The data are transformed into n new observations (u_t, z_t) as follows.

$$u_1 = \sqrt{1 - \rho^2}\, y_1 ,$$

$$z_1 = \sqrt{1 - \rho^2}\, x_1 ,$$

$$\left.\begin{aligned} u_t &= y_t - \rho y_{t-1} , \\ z_t &= x_t - \rho x_{t-1} \end{aligned}\right\} \quad t = 2, 3, \ldots, n .$$

The intercept h and slope d of the common least-squares regression of u upon z are interpreted as estimators as follows.

$$E(h) = (1 - \rho)\alpha ,$$

$$E(d) = \beta .$$

This procedure is recommended e.g. by Wonnacott and Wonnacott [1970], section 16-4.

The procedure is tantamount to a linear transformation of the n micro units of analysis into n macro units of analysis. The aggregating functions for the regressand and for the observed regressor are $u = T'y$ and $z = T'x$, where

$$T' = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 \ldots\ldots 0 & 0 \\ -\rho & 1 & 0 \ldots\ldots 0 & 0 \\ 0 & -\rho & 1 \ldots\ldots 0 & 0 \\ \vdots & \vdots & \vdots \quad\quad \vdots & \vdots \\ 0 & 0 & 0 \ldots\ldots -\rho & 1 \end{bmatrix}$$

There is one designed regressor vector $X^{\Delta} = Z^{\Delta} = j_n$.

The semi-aggregated and macro regressor matrices are

$$X_A = [\, T'j_n \quad T'x \,] = \left[\begin{array}{c|c} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\, x_1 \\ 1-\rho & x_2 - \rho x_1 \\ 1-\rho & x_3 - \rho x_2 \\ \vdots & \vdots \\ 1-\rho & x_n - \rho x_{n-1} \end{array}\right],$$

$$Z = [\, j_n \quad z \,] = \left[\begin{array}{c|c} 1 & \sqrt{1-\rho^2}\, x_1 \\ 1 & x_2 - \rho x_1 \\ 1 & x_3 - \rho x_2 \\ \vdots & \vdots \\ 1 & x_n - \rho x_{n-1} \end{array}\right]$$

Assume that the rank of Z is 2 .

The 2×2 matrix of auxiliary regression coefficients is

$$C = \left[\begin{array}{c|c} c^{\Delta\Delta} & 0 \\ \hline c^{\Delta O} & 1 \end{array}\right] ,$$

where $c^{\Delta\Delta} \cong 1 - \rho$ and $c^{\Delta O} \cong 0$, at least if n is large.

By P.6.1.3, a necessary condition for consistency to be attainable is that for some scalar δ^{Δ}

$$T'j_n \alpha = j_n \delta^{\Delta} \ .$$

Because of the first element of $T'j_n$, consistency is not attainable.

By P.6.1.4, the macro intercept and slope are to be interpreted as follows.

$$E_{H\Phi}(h) = c^{\Delta\Delta} \alpha \ ,$$
$$E_{H\Phi}(d) = c^{\Delta O} \alpha + \beta \ .$$

The recommended interpretations are only approximately correct.

Consistency can be made attainable, and the recommended interpretations exact, by either of two modifications of the procedure. The first alternative is to omit the first macro unit of analysis. This wastes some information.

The second alternative is to exchange the macro intercept regressor j_n for another designed macro regressor $(1 - \rho)^{-1} T'j_n$. This complicates the computation of h and d, but makes these BLUEs.

6.1.6 A special case: Weighted regression

Weighted least-squares regression is a special case of linear transformation of units of analysis. A weight $k_j > 0$ is associated with each micro unit, $j = 1, \ldots, n$. Each micro unit is transformed into a macro unit. For simplicity, let there be a single designed regressor, the intercept regressor. The micro data are $[\, y \mid j_n \mid X^O \,]$.

Let K be the $n \times n$ diagonal matrix, and k the n-vector, whose j'th (diagonal) elements are both k_j, $j = 1, \ldots, n$. The aggregating functions for observed variables are

$$[\, u \mid Z^O \,] = K \, [\, y \mid X^O \,] \ .$$

The designed macro regressor vector is defined to be k. (Note that in general $k \neq j_n$.) The macro data are $[\, u \mid k \mid Z^O \,]$. Since $Kj_n = k$, the semi-aggregated micro regressor matrix $[\, Kj_n \mid KX^O \,]$ agrees with the macro regressor matrix. The matrix of auxiliary regression coefficients can always be taken to be $C = I_p$.

Consider the consistency problem. The micro and macro parameters associated with the designed regressor are β^{Δ} and δ^{Δ}. By P.6.1.3, a necessary and

sufficient condition for consistency to be attainable is that there exists a scalar δ^{Δ} such that

$$K j_n \beta^{\Delta} = k \delta^{\Delta} .$$

This is always the case; choose $\delta^{\Delta} = \beta^{\Delta}$.

Consider the interpretation problem. In spite of the consistency, there is no simple model-free interpretation. Since $C = I_p$, the expectational interpretation has $\lambda = \mu$, i.e. $E_{H\Phi}(\mu' d) = \mu' \beta$.

When the micro relation Φ is a heteroskedastic linear model with disturbance covariance matrix $\sigma^2 K^{-2}$, weighted regression provides BLUEs. One must not forget to replace the intercept regressor by k , however.

6.2 Unweighted partitioned aggregation of non-designed units

6.2.1 Aggregation by summation: Definition and notation

Unweighted partitioned aggregation by summation of non-designed units of analysis will now be defined; cf. 3.1.8. A variant, aggregation by averaging, is introduced later, in 6.2.4 below. What was said in 6.1.1 - 6.1.4 is modified as follows.

The n micro units are partitioned into m disjoint, exhaustive subsets. For $h = 1, \ldots, m$, the h'th macro unit is formed from the $n_h \geq 1$ members of the h'th subset. The regressand and the $p^O \geq 0$ observed regressors are aggregated by summation over the n_h micro units. There are $p^M \geq 0$ repetitive mixed regressors. These are constant across the members of any given subset of micro units. Their values for the h'th macro unit are equal to those for any member of the h'th subset of micro units. Finally, there is at most a single designed regressor, the micro and macro intercept regressor.

The micro and macro regression data matrices are denoted

$$[\, y \mid j_n \mid X^M \mid X^O \,] ,$$

$$[\, u \mid j_m \mid Z^M \mid Z^O \,] .$$

The subvectors of β, λ, δ, μ associated with the three subsets of regressors are denoted β^{Δ}, β^{M}, β^{O}, and so on.

For $h = 1, \ldots, m$, let i_h denote the n_h-vector of unit elements. Define

$$T = \begin{bmatrix} i_1 & 0 & \ldots\ldots & 0 \\ 0 & i_2 & \ldots\ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots\ldots & i_m \end{bmatrix}$$

The $n \times m$ transformation matrix T is block-diagonal.

In aggregation by summation, the aggregating functions for purely observed variables are

$$[u \mid Z^{O}] = T' [y \mid X^{O}] .$$

The repetitive mixed regressors satisfy the "converse" relation

$$X^{M} = TZ^{M} .$$

The basic observed regressor data are Z^{M} and X^{O}; cf. 3.1.8.

Finally, let N be the $m \times m$ diagonal matrix, whose h'th diagonal element is n_h, $h = 1, \ldots, m$. Note that $N = T'T$.

6.2.2 The consistency problem

In aggregation by summation, the semi-aggregated micro and semi-disaggregated macro relations are

$$E_{H\Phi}(u) = T' j_n \beta^{\Delta} + NZ^{M}\beta^{M} + T'X^{O}\beta^{O} ,$$

$$E_{\Psi G}(u) = j_m \delta^{\Delta} + Z^{M}\delta^{M} + T'X^{O}\delta^{O} .$$

Let u_h, $z_h^{M'}$ and $z_h^{O'}$ denote the h'th element and row vectors of u, Z^{M} and $Z^{O} = T'X^{O}$. The two m-vector equations $H\Phi$ and ΨG consist of the following m pairs of scalar equations, $h = 1, \ldots, m$.

$$E_{H\Phi}(u_h) = n_h \beta^{\Delta} + z_h^{M'} (n_h \beta^{M}) + z_h^{O'} \beta^{O} ,$$

$$E_{\Psi G}(u_h) = \delta^{\Delta} + z_h^{M'} \delta^{M} + z_h^{O'} \delta^{O} .$$

The following proposition answers the consistency problem.

Proposition 6.2.2

Consider unweighted partitioned aggregation by summation of non-designed units of analysis.

Consistency is always attainable if either or both of (i) and (ii) is true, and never otherwise.

(i) There is no intercept, and $p^{M} = 0$.

(ii) $n_1 = n_2 = \ldots = n_m$. ∘∘

Analogously with the demonstration of P.4.2.6, the aggregation is consistent if and only if for $h = 1, \ldots, m$

$$\begin{bmatrix} n_h \beta^{\Delta} \\ \hline n_h \beta^{M} \\ \hline \beta^{O} \end{bmatrix} = \begin{bmatrix} \delta^{\Delta} \\ \hline \delta^{M} \\ \hline \delta^{O} \end{bmatrix} .$$

This can be achieved by an appropriate choice of δ if and only if (i) or (ii) is true, or both. □

The consistency conditions of P.6.2.2 are in effect not restrictions on the micro parameters.

6.2.3 An expectational interpretation

In aggregation by summation, the semi-aggregated micro relation is

$$E_{H\Phi}(u) = [\, T' j_n \,\vdots\, NZ^{M} \,\vdots\, T'X^{O} \,]\beta ,$$

i.e. $\beta_A = \beta$, while the macro regressor data are $[\, j_m \,\vdots\, Z^{M} \,\vdots\, Z^{O} \,]$. The matrix of GLS(W) auxiliary regression coefficients is

$$C = \begin{bmatrix} c^{\Delta\Delta} & c^{M\Delta'} & 0 \\ c^{\Delta M} & C^{MM} & 0 \\ c^{\Delta O} & C^{MO} & I^{O} \end{bmatrix} ,$$

where I^O is the unit matrix of order p^O. The following proposition provides an expectational interpretation. As in 6.1.4 there are no weight-sum relations.

Proposition 6.2.3

Consider unweighted partitioned aggregation by summation of non-designed units of analysis.

If $\mu'd$ is a unique macro coefficient function, then

$$E_{H\Phi}(\mu^\Delta d^\Delta + \mu^{M'} d^M + \mu^{O'} d^O) = \lambda^\Delta \beta^\Delta + \lambda^{M'} \beta^M + \lambda^{O'} \beta^O ,$$

where

$$\begin{bmatrix} \lambda^\Delta \\ \lambda^M \\ \lambda^O \end{bmatrix} = \begin{bmatrix} c^{\Delta\Delta}\mu^\Delta + c^{\Delta M'}\mu^M + c^{\Delta O'}\mu^O \\ c^{M\Delta}\mu^\Delta + C^{MM'}\mu^M + C^{MO'}\mu^O \\ \mu^O \end{bmatrix} \quad \circ$$

This is a special case of P.4.4.3. □

This proposition is not particularly illuminating.

Now consider the special case where

$$n_1 = n_2 = \ldots = n_m = \bar{n} = \frac{n}{m} .$$

Then $T'j_n = \bar{n} j_m$, $NZ^M = \bar{n} Z^M$, and $T'X^O = Z^O$. Thus C is block-diagonal, $c^{\Delta\Delta} = \bar{n}$, and $C^{MM} = \bar{n} I^M$, where I^M is the unit matrix of order p^M. Consequently λ in P.6.2.3 is as follows.

$$\begin{bmatrix} \lambda^\Delta \\ \lambda^M \\ \lambda^O \end{bmatrix} = \begin{bmatrix} \bar{n}\mu^\Delta \\ \bar{n}\mu^M \\ \mu^O \end{bmatrix} .$$

In spite of the consistency demonstrated in P.6.2.2, and in spite of the simplicity of the expectational interpretation, there is no simple model-free interpretation.

6.2.4 An alternative: Aggregation by averaging

Unweighted partitioned aggregation by averaging of non-designed units of analysis will now be briefly discussed. The notation is as in 6.2.1. Note that $N^{-1}T'j_n = j_m$ and $N^{-1}T'T = I_m$.

In aggregation by averaging, the regressand aggregating function is $H: u = N^{-1}T'y$. The aggregating function for observed regressors is similar, and the designed and mixed regressors can also be taken to be aggregated in the same way. In summary

$$[\,u \,\vdots\, j_m \,\vdots\, Z^M \,\vdots\, Z^O\,] = N^{-1}T'\,[\,y \,\vdots\, j_n \,\vdots\, X^M \,\vdots\, X^O\,]\,,$$

where the relation $X^M = TZ^M$ is used.

The semi-aggregated micro and semi-disaggregated macro relations are

$$E_{H\Phi}(u) = j_m\beta^{\Delta} + Z^M\beta^M + N^{-1}T'X^O\beta^O\,,$$

$$E_{\Psi G}(u) = j_m\delta^{\Delta} + Z^M\delta^M + N^{-1}T'X^O\delta^O\,.$$

The macro and semi-aggregated micro regressor matrices are identical, $Z = X_A$. Further, $\beta_A = \beta$.

Proposition 6.2.4

Consider unweighted partitioned aggregation by averaging of non-designed units of analysis.

(i) Consistency is always attainable.

(ii) If $\mu'd$ is a unique macro coefficient function, then

$$E_{H\Phi}(\mu'd) = \mu'\beta_A = \mu'\beta\,. \quad \circ\!\circ$$

(i) P.4.2.6 applies. Consistency is attained by choosing $\delta = \beta$.

(ii) P.4.4.3 applies. Since $X_A = Z$, one can put $C = I_p$. Further, $\beta_A = \beta$. □

In spite of the consistency, and in spite of the simplicity of the expectational interpretation, there is no simple model-free interpretation; but cf. 6.4.3.

Less formal statements essentially equivalent to P.6.2.4 are found in Prais and Aitchison [1954].

6.3 Aggregation of sets of units

6.3.1 Definition and notation

There are one regressand and p regressors. There are $k \geq 2$ disjoint micro sets of units of analysis. The data for the h'th set of $n_h \geq 1$ units form the $n_h \times (1+p)$ regression data matrix $[\, y_h \,|\, X_h \,]$, $h = 1, \ldots, k$. There is one macro set of units of analysis. The data for the macro set of n units form the $n \times (1+p)$ regression data matrix $[\, u \,|\, Z \,]$.

The regressand and regressor aggregating functions are as follows.

$$H,\ G:[\, u \,|\, Z \,] = \left[\begin{array}{c|c} y_1 & X_1 \\ \hline \vdots & \vdots \\ \hline y_h & X_h \\ \hline \vdots & \vdots \\ \hline y_k & X_k \end{array}\right]$$

The macro set of units is the union of the micro sets of units, and $n = n_1 + n_2 + \ldots + n_k$; cf. 3.1.6.

The micro and macro relations

$$\Phi: \quad y_h = X_h \beta_h + \varepsilon_h; \quad E(\varepsilon_h) = 0, \quad h = 1, \ldots, k,$$

$$\Psi: \quad u = Z\delta + \eta; \quad E(\eta) = 0$$

are sometimes considered.

The semi-aggregated regressor matrix is defined as follows,

$$X_A = \left[\begin{array}{ccc} X_1 & \ldots\ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots\ldots & X_h \\ \vdots & & \vdots \\ 0 & \ldots\ldots & 0 \end{array}\right] \quad \begin{array}{c} \ldots\ldots 0 \\ \vdots \\ \ldots\ldots 0 \\ \vdots \\ \ldots\ldots X_k \end{array} \quad ,$$

i.e. as the $n \times kp$ block-diagonal arrangement of the micro regressor matrices.

The total micro parameter vector is the kp-vector $\beta = (\beta_1' \ldots \beta_h' \ldots \beta_k')'$ as in 4.1.2. Defining $y = u$, the micro relation can be written

$$\Phi: \quad E(y) = X_A \beta .$$

Sometimes $p^\Delta \geq 0$ designed and $p^O \geq 0$ observed regressors are distinguished, where $p^\Delta + p^O = p$. The micro and macro regressor matrices and parameter vectors are partitioned accordingly, as follows.

$$\Phi: \quad E(y_h) = X_h^\Delta \beta_h^\Delta + X_h^O \beta_h^O , \quad h = 1, \ldots, k ,$$

$$\Psi: \quad E(u) = Z^\Delta \delta^\Delta + Z^O \delta^O .$$

Analogously with X_A and β, define

$$X_A^\Delta = \begin{bmatrix} X_1^\Delta & \ldots\ldots 0 & \ldots\ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \ldots\ldots & X_h^\Delta & \ldots\ldots 0 \\ \vdots & \vdots & & \vdots \\ 0 & \ldots\ldots 0 & \ldots\ldots & X_k^\Delta \end{bmatrix}$$

and $\beta^\Delta = (\beta_1^{\Delta\prime} \ldots\ldots \beta_h^{\Delta\prime} \ldots\ldots \beta_k^{\Delta\prime})'$, and similarly X_A^O and β^O. In terms of these, the micro relation can be written

$$\Phi: \quad E(y) = X_A^\Delta \beta^\Delta + X_A^O \beta^O .$$

The orders of X_A^Δ and X_A^O are $n \times kp^\Delta$ and $n \times kp^O$.

Finally, define the $kp \times p$ matrix.

$$L = [\, I_p \quad I_p \ldots\ldots\ldots I_p \,]'$$

and the analogous $kp^\Delta \times p^\Delta$ and $kp^O \times p^O$ matrices L^Δ and L^O. The aggregating functions can be reformulated

$$G^*: \quad Z = X_A L \ ,$$

$$H^*: \quad u = y \ .$$

Also, $Z^\Delta = X_A^\Delta L^\Delta$, and $Z^O = X_A^O L^O$.

The reformulation $\{G^*, H^*\}$ transforms aggregation of sets of units into, formally, a special kind of aggregation of regressors.

6.3.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are $E_{H\Phi}(u) = \alpha_A + X_A^O \beta_A^O$ and $E_{\Psi G}(u) = \gamma_D + X_A^O \delta_D^O$, where

$$\begin{bmatrix} \alpha_A \\ \hline \beta_A^O \end{bmatrix} = \begin{bmatrix} X_A^\Delta \beta^\Delta \\ \hline \beta^O \end{bmatrix} \ ,$$

$$\begin{bmatrix} \gamma_D \\ \hline \delta_D^O \end{bmatrix} = \begin{bmatrix} X_A^\Delta L^\Delta \delta^\Delta \\ \hline L^O \delta^O \end{bmatrix} \ .$$

The following proposition answers the consistency problem.

Proposition 6.3.2

Consider aggregation of sets of units of analysis.

A necessary and sufficient condition for consistency to be attainable is that statements (i) and (ii) are true.

If the rank of every X_h^Δ is p^Δ, statements (ii*) and (ii) are equivalent.

(i) $\beta_1^O = \beta_2^O = \ldots\ldots = \beta_k^O$.

(ii) The designed regressor data and associated micro parameters are such that

$$X_A^\Delta \beta^\Delta = X_A^\Delta L^\Delta \delta^\Delta$$

for at least one p^Δ-vector δ^Δ .

(ii*) $\beta_1^\Delta = \beta_2^\Delta = \ldots\ldots = \beta_k^\Delta$ ∘∘

This is a special case of P.4.2.6 or of P.5.1.2, whose statement (i): $\beta^O = L^O \delta^O$ has been broken into k parts, while statement (ii) has been retained.

Equations (ii) say that $X_h^\Delta \beta_h^\Delta = X_h^\Delta \delta^\Delta$ for $h = 1, \ldots, k$. If the rank of an X_h^Δ is p^Δ, that X_h^Δ can be "divided out". □

6.3.3 A model-free interpretation

Throughout 6.3.3., the $n \times n$ matrix W is restricted to be block-diagonal as follows. For $h = 1, \ldots, k$, let W_h be a positive-definite $n_h \times n_h$ matrix. Define W as the block-diagonal arrangement

$$W = \begin{bmatrix} W_1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & W_h & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & W_k \end{bmatrix} .$$

The matrix W is positive-definite.

For $h = 1, \ldots, k$, let b_h be any p-vector satisfying the normal equations $X_h' W_h X_h b_h = X_h' W_h y_h$ for the h'th (GLS(W_h)) micro regression. Define the total micro coefficient vector b of order kp as follows,

$$b = (b_1' \ldots b_h' \ldots b_k')' ,$$

i.e. analogously with the total micro parameter vector β of 4.1.2. The k micro normal equation systems can be written jointly

$$X_A' W X_A b = X_A' W y .$$

The block-diagonality of W is essential here. The situation differs from that in Zellner [1962a], where the point is that W is not block-diagonal.

The $p \times kp$ matrix of GLS(W) auxiliary regression coefficients will be partitioned $C = [\, C_1 \ldots C_h \ldots C_k \,]$ conformably with the columns of X_A. The auxiliary normal equation system $Z'WZC = Z'WX_A$ can be decomposed as follows.

$$(\sum_{i=1}^{k} X_i' W_i X_i) C_h = X_h' W_h X_h \quad , \quad h = 1, \ldots, k \quad .$$

The block-diagonality of W is essential here too.

The macro regression data $[\, u \,\vdots\, Z \,]$ and the semi-aggregate regression data $[\, u \,\vdots\, X_A \,]$ are related by the equation $X_A L = Z$.

The semi-aggregate regression data $[\, u \,\vdots\, X_A \,]$ and the total set of micro regression data $[\, y \,\vdots\, X_A \,]$ are identical. Thus any vector b_A of GLS(W) semi-aggregate regression coefficients is also a total micro coefficient vector b, where b_h is GLS(W_h), h = 1, ..., k, and conversely; cf. 4.3.6.

The following proposition provides a model-free interpretation.

Proposition 6.3.3

Consider aggregation of sets of units of analysis. Let the matrix W be block-diagonal conformably with X_A. Let d, b_A and b be any vectors of GLS(W) macro, semi-aggregate, and total micro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

Let μ be a p-vector such that $\mu' d$ is a unique macro coefficient function. Define the kp-vector $\lambda = C'\mu$ with subvectors $\lambda_h = C_h'\mu$ of order p. Then:

(i) $\quad \mu' d = \lambda' b_A = \sum_{h=1}^{k} \lambda_h' b_h \quad .$

(ii) $\quad \sum_{h=1}^{k} \lambda_h = \mu \, .$ $\quad$ ∘∘

This is a special case of P.4.3.4 or of P.5.1.3. The second equation of (i) follows because $b_A = b$, and b has the subvectors b_h. Equation (ii) is a reformulation of $L'\lambda = \mu$. □

In the terms of 4.3.2, this model-free interpretation chooses $Q_h = W_h$, h = 1, ..., k. The block-diagonality of W is essential. For another model-free interpretation see 6.4.2.

If the rank of Z is p, all macro and auxiliary coefficients d and C are unique. Then by P.6.3.3, $d = \Sigma C_h b_h$, where $\Sigma C_h = I_p$. Any macro coefficient is a weighted sum of all micro coefficients. Group the kp micro coefficients into p families, each associated with a separate regressor. The sum of the weights for any one family is zero, with one exception. The sum of the weights of the corresponding family of k micro coefficients is unity.

A similar model-free interpretation is given in Klevmarken [1972] section 5.1.3.3, for aggregation of sets with partly different regressors, and under linear constraints.

That a unique macro coefficient need not fall between the smallest and largest corresponding micro coefficients is shown by an example in Larson and Barr [1972].

6.3.4 An expectational interpretation

In 6.3.4, the matrix W need not be block-diagonal.

The micro models Φ imply the semi-aggregated micro relation $E_{H\Phi}(u) = X_A \beta_A$, where $\beta_A = \beta$. As before, $X_A L = Z$.

The following proposition provides an expectational interpretation.

Proposition 6.3.4

Consider aggregation of sets of units of analysis. Let d be any vector of GLS(W) macro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

Let μ be a p-vector such that $\mu'd$ is a unique macro coefficient function. Define the kp-vector $\lambda = C'\mu$ with subvectors $\lambda_h = C_h'\mu$ of order p. Then:

(i) $E_{H\Phi}(\mu'd) = \lambda'\beta_A = \sum_{h=1}^{k} \lambda_h'\beta_h$.

(ii) $\sum_{h=1}^{k} \lambda_h = \mu$. ○

This is a special case of P.4.4.3 or of P.5.1.4. The second equation of (i) follows because $\beta_A = \beta$, and β has the subvectors β_h. Equation (ii) is a reformulation of $L'\lambda = \mu$. □

If W is block-diagonal as in 6.3.3, then the implied micro parameter functions $\lambda_h'\beta_h$ in P.6.3.4 have the same weight vectors λ_h as the implied micro coefficient functions $\lambda_h'b_h$ in P.6.3.3; cf. 4.4.5.

6.3.5 A special case: Individual micro models

A special case of aggregation of sets of units occurs when each unit of analysis has its own incomplete linear model with parameter vector β_h, i.e.

$n_1 = \ldots = n_k = 1$. For simplicity, assume that the only designed regressor is the intercept regressor. Then by P.6.3.2 consistency is attainable if and only if all β_h are equal, i.e. there is a single incomplete linear model common to all units of analysis.

Model-free interpretation by P.6.3.3 is hardly of interest here, for when $n_h = 1$ and $p \geq 2$ every b_h is underdetermined. An expectational interpretation is always provided by P.6.3.4. Such an interpretation is derived for the case where d is unique and $W = I_n$ by Zellner [1962b].

6.4 A model-free interpretation in deviation form regression

6.4.1 Two fundamental relations

Section 6.4 is based on 2.5.4 above. Please read 2.5.4 again! The block-diagonality of the matrix Q is essential.

The following abbreviated notation is introduced for the $p \times p$ deviation form moment matrices.

$$M_T = (X - j_n \bar{\bar{x}}')' Q (X - j_n \bar{\bar{x}}') ,$$

$$M_{Wh} = (X_h - i_h \bar{x}_h')' Q_h (X_h - i_h \bar{x}_h') , \quad h = 1, \ldots, m,$$

$$M_B = (\bar{X} - j_m \bar{\bar{x}}')' \bar{Q} (\bar{X} - j_m \bar{\bar{x}}') .$$

By P.2.5.4, $M_T = M_{W1} + \ldots + M_{Wm} + M_B$.

The p variables are partitioned into $q = p - 1$ regressors and one regressand, indicated by the symbols x and y. The moment matrices are partitioned accordingly,

$$M_T = \left[\begin{array}{c|c} M_T^{xx} & M_T^{xy} \\ \hline M_T^{yx} & M_T^{yy} \end{array}\right] ,$$

and similarly for M_{Wh} and M_B.

Consider the following $m + 2$ regressions, all in deviation form in the sense of 2.5.2.

The normal equations of the GLS(Q) total regression are

$$M_T^{xx} b_T = M_T^{xy} \quad ,$$

where b_T is the q-vector of total regression (slope) coefficients.

For $h = 1, \ldots, m$, the normal equations of the GLS(Q_h) regression within the h'th subset are

$$M_{Wh}^{xx} b_h = M_{Wh}^{xy} \quad ,$$

where b_h is the q-vector of regression (slope) coefficients within the h'th subset.

The normal equations of the GLS($\overline{Q}$) regression between subsets are

$$M_B^{xx} b_B = M_B^{xy} \quad ,$$

where b_B is the q-vector of regression (slope) coefficients between subsets. The regression between subsets treats the m subsets as units of analysis. The data for each subset are formed from those of the n_h member units by averaging in the sense of 6.2.4.

By P.2.5.4,

$$M_T^{xy} = \sum_{h=1}^{m} M_{Wh}^{xy} + M_B^{xy} \quad .$$

By substitution of the m + 2 systems of normal equations,

$$(R) \qquad M_T^{xx} b_T = \sum_{h=1}^{m} M_{Wh}^{xx} b_h + M_B^{xx} b_B \quad .$$

Further, by P.2.5.4,

$$(S) \qquad M_T^{xx} = \sum_{h=1}^{m} M_{Wh}^{xx} + M_B^{xx} \quad .$$

The fundamental relation (R) holds for any q-vectors b_T, b_h, b_B satisfying the respective normal equations systems. The fundamental relation (S) may be called a set of weight-sum relations.

Two special cases are of interest. In the first special case, $Q = I_n$, so that all Q_h are also unit matrices. The total and within-subsets regressions are by common least-squares. The between-subsets regression is by weighted least-squares. The weight in M_B of the h'th subset is n_h.

In the second special case, $\overline{Q} = I_m$, and all Q_h are scalar matrices. The

between-subsets and within-subsets regressions are by common least-squares. The total regression is by weighted least-squares. The weight in M_T of each member of the h'th subset is n_h^{-1}.

6.4.2 Aggregation of sets

In aggregation of sets of units, the macro coefficients are the deviation form total regression coefficients b_T. If these are unique, the following proposition provides a kind of a model-free interpretation. Not only micro terms are involved.

Proposition 6.4.2

Consider aggregation of sets of units. Let Q, Q_h and $\overline{Q}$ be as in 2.5.4. Assume that the rank of the total regressor moment matrix M_T^{xx} is q.

Let b_T be the q-vector of GLS(Q) total regression coefficients. For $h = 1, \ldots, m$, let b_h be any q-vector of coefficients in the GLS(Q_h) regression within the h'th subset. Let b_B be any q-vector of coefficients in the GLS($\overline{Q}$) regression between subsets.

Let λ_T be any q-vector. Define the q-vectors

$$\lambda_h = M_{Wh}^{xx}(M_T^{xx})^{-1}\lambda_T \quad , \quad h = 1, \ldots, m \quad ,$$

$$\lambda_B = M_B^{xx}(M_T^{xx})^{-1}\lambda_T \quad .$$

Then:

(i) $\lambda_h' b_h$, $h = 1, \ldots, m$, and $\lambda_B' b_B$ are unique coefficient functions in the respective regressions.

(ii) $\lambda_T' b_T = \sum_{h=1}^{m} \lambda_h' b_h + \lambda_B' b_B$.

(iii) $\sum_{h=1}^{m} \lambda_h + \lambda_B = \lambda_T$. ○○

The vectors λ_h and λ_B are unique, and are in the row spaces of M_{Wh}^{xx} and M_B^{xx}, respectively. By P.2.1.2.A and P.2.2.4.A this implies (i).

If (R) and (S) of 6.4.1 are premultiplied by $\lambda_T'(M_T^{xx})^{-1}$, the outcome is equations (ii) and (iii). □

6.4.3 Partitioned aggregation by averaging of units

In partitioned aggregation by averaging of units of analysis, the macro coefficients are the coefficients b_B in the deviation form regression between subsets. If these are unique, the following proposition provides a kind of a model-free interpretation. Not only micro terms are involved.

Proposition 6.4.3

Consider partitioned aggregation by averaging of units of analysis. Let Q, Q_h, and $\overline{Q}$ be as in 2.5.4. Assume that the rank of the between subsets regressor moment matrix M_B^{xx} is q.

Let b_B be the q-vector of coefficients in the GLS($\overline{Q}$) regression between subsets. For $h = 1, \ldots, m$, let b_h be any q-vector of coefficients in the GLS(Q_h) regression within the h'th subset. Let b_T be any q-vector of GLS(Q) total regression coefficients.

Let λ_B be any q-vector. Define the q-vectors

$$\lambda_h = - M_{Wh}^{xx} (M_B^{xx})^{-1} \lambda_B \quad , \quad h = 1, \ldots, m \quad ,$$

$$\lambda_T = M_T^{xx} (M_B^{xx})^{-1} \lambda_B \quad .$$

Then:

(i) $\lambda_h' b_h$, $h = 1, \ldots, m$, and $\lambda_T' b_T$ are unique coefficient functions in the respective regressions.

(ii) $\lambda_B' b_B = \sum_{h=1}^{m} \lambda_h' b_h + \lambda_T' b_T$.

(iii) $\sum_{h=1}^{m} \lambda_h + \lambda_T = \lambda_B$. ○○

The vectors λ_h and λ_T are unique, and are in the row spaces of M_{Wh}^{xx} and M_T^{xx}, respectively. By P.2.1.2.A and P.2.2.4.A this implies (i).

If (R) and (S) of 6.4.1 are premultiplied by $\lambda_B' (M_B^{xx})^{-1}$, the outcome is equations (ii) and (iii). □

6.4.4 On analogous expectational interpretations

Consider an attempt to establish an expectational interpretation for aggregation of sets of units, analogous to the model-free interpretation of 6.4.2. The following $m + 1$ incomplete linear micro models must then be assumed. A model Φ_h with parameter vector β_h is valid for the n_h members of the h'th subset of units. This is so for $h = 1, \ldots, m$. A model Φ_B with parameter vector β_B is valid for the m subsets, when these are treated as units of analysis in the regression between subsets.

As a rule, these assumptions are not mutually consistent. In the regression between subsets, the h'th subset is represented by data obtained from its n_h members by averaging. Therefore by the reasoning in 6.2.4, the model Φ_h is valid for the h'th subset average. On the other hand, the model Φ_B is also assumed to be valid. By the reasoning in 6.3.2, inconsistency is avoided only if $\beta_1 = \ldots = \beta_m = \beta_B$.

Consider also an attempt to establish an expectational interpretation for partitioned unweighted aggregation by averaging of non-designed units of analysis, analogous to the model-free interpretation of 6.4.3. The following $m + 1$ incomplete linear micro models must then be assumed. A model Φ_h with parameter vector β_h is valid for the n_h members of the h'th subset of units. This is so for $h = 1, \ldots, m$. A model Φ_T with parameters β_T is valid for all the n units of analysis.

As a rule, these assumptions are not mutually consistent. By the reasoning in 6.3.2, inconsistency is avoided only if $\beta_1 = \ldots = \beta_m = \beta_T$.

7 A FORMAL ANALYSIS: AGGREGATION OF ASPECTS

7.1 General aggregation of regressands and regressors

7.1.1 Definition and notation

There are n units of analysis. The micro data form the $k \geq 2$ regression data matrices $[\, y_i \,\vdots\, X_i \,]$ of orders $n \times (1 + p_i)$, $i = 1, \ldots, k$. The union of the k sets of micro regressors forms the $n \times p$ total micro regressor matrix X, and the k micro regressands form the $n \times k$ micro regressand matrix Y; cf. 5.3.1. The macro data form the $n \times (1 + q)$ regression data matrix $[\, u \,\vdots\, Z \,]$.

The regressor and regressand aggregating functions are

$$G: \quad Z = XG \ ,$$

$$H: \quad u = Yh \ ,$$

where G is a given $p \times q$ matrix of constants, and h is a given k-vector of constants; cf. 5.1.1 and 5.3.1.

General aggregation of regressands and regressors includes cases that it may not be natural to call aggregations; cf. 3.2.4.

The micro and macro relations

$$\Phi: \quad y_i = X_i \beta_i + \epsilon_i \ ; \quad E(\epsilon_i) = 0 \ ; \quad i = 1, \ldots, k \ ,$$

$$\Psi: \quad u = Z\delta + \eta \ ; \quad E(\eta) = 0$$

are sometimes considered. Alternatively, the k micro models are written jointly

$$\Phi: \quad E(Y) = XB^* \ ,$$

where the augmented micro parameter $p \times k$ matrix B^* and its column vectors β_i^* are defined as in 5.3.1.

Designed and observed micro and macro regressors are sometimes distinguished. The regressor aggregating function becomes

$$G: \quad [\, Z^{\Delta} \,\vdots\, Z^{O} \,] = [\, X^{\Delta} \,\vdots\, X^{O} \,] \begin{bmatrix} G^{\Delta} & \vdots & G^{T} \\ \cdots & \vdots & \cdots \\ 0 & \vdots & G^{O} \end{bmatrix} \ .$$

The joint micro relations and the macro relation become

$$\Phi: \quad E(Y) = X^{\Delta}B^{*\Delta} + X^{O}B^{*O}$$

$$\Psi: \quad E(u) = Z^{\Delta}\delta^{\Delta} + Z^{O}\delta^{O} \quad .$$

Cf. 5.1.1 and 5.3.1, respectively.

The semi-aggregated regressor matrix and its designed and observed submatrices are defined to be identical to the total micro regressor matrix and its submatrices, $X_A = [\, X_A^{\Delta} \,\vdots\, X_A^{O} \,] = [\, X^{\Delta} \,\vdots\, X^{O} \,] = X$.

7.1.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are $E_{H\Phi}(u) = \alpha_A + X_A^{O}\beta_A^{O}$ and $E_{\Psi G}(u) = \gamma_D + X_A^{O}\delta_D^{O}$, where

$$\begin{bmatrix} \alpha_A \\ \hline \beta_A^{O} \end{bmatrix} = \begin{bmatrix} X^{\Delta}B^{*\Delta}h \\ \hline B^{*O}h \end{bmatrix} \quad ,$$

$$\begin{bmatrix} \gamma_D \\ \hline \delta_D^{O} \end{bmatrix} = \begin{bmatrix} X^{\Delta}(G^{\Delta}\delta^{\Delta} + G^{T}\delta^{O}) \\ \hline G^{O}\delta^{O} \end{bmatrix} \quad ;$$

cf. 5.3.2 and 5.1.2 respectively. The following proposition answers the consistency problem.

Propositon 7.1.2

Consider general aggregation of regressands and regressors.

A necessary and sufficient condition for consistency to be attainable is that the following two statements are both true.

(i) The total micro parameter submatrix B^{*O} associated with observed regressors is such that

$$B^{*O}h = G^{O}\delta^{O}$$

for at least one q^{O}-vector δ^{O}.

(ii) The designed total micro regressor submatrix X^{Δ} and the associated total micro parameter submatrix $B^{*\Delta}$ are such that, for some δ^{O} satisfying (i),

$$X^{\Delta}B^{*\Delta}h = X^{\Delta}(G^{\Delta}\delta^{\Delta} + G^{T}\delta^{O})$$

for at least one q^{Δ}-vector δ^{Δ}. $\circ\!\!\circ$

This is a special case of P.4.2.6. □

Statement (i) taken alone is a necessary condition for consistency to be attainable. In particular, if $h = j_k$, then $B^{*O}j_k = \Sigma\beta_i^{*O}$ must be in the column space of G^{O}.

7.1.3 An expectational interpretation

The micro models $E(Y) = XB^*$ imply the semi-aggregated micro model $E_{H\Phi}(u) = X_A\beta_A$, where $X_A = X$ and $\beta_A = B^*h$; cf. 5.3.3. Further, $X_AG = Z$; cf. 5.1.4.

For any p-vector θ, let θ_i be the subvector corresponding to the micro regressor subset included in X_i; cf. 5.3.3. Let h_i be the i'th element of h. The following proposition provides an expectational interpretation.

Proposition 7.1.3

Consider general aggregation of regressands and regressors. Let d be any vector of GLS(W) macro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

Let μ be a q-vector such that $\mu'd$ is a unique macro coefficient function. Define the p-vector $\theta = C'\mu$. For $i = 1, \ldots, k$, define the p_i-vector $\lambda_i = h_i\theta_i$. Then:

(i) $E_{H\Phi}(\mu'd) = \theta'\beta_A = \sum_{i=1}^{k}\lambda_i'\beta_i$.

(ii) $G'\theta = \mu$. $\circ\!\!\circ$

This is a special case of P.4.4.3, except for the second equation of (i), which follows since $\theta'\beta_A = \theta'B^*h = \Sigma h_i\theta'\beta_i^* = \Sigma h_i\theta_i'\beta_i$. □

In particular, if $h = j_k$, then $\lambda_i = \theta_i$, $i = 1, \ldots, k$.

Except for special cases such as when $X_i = X$ for every i, there is no simple model-free interpretation.

7.2 Unweighted simple aggregation of aspects

7.2.1 Definition and notation

Unweighted simple aggregation of aspects will now be defined. As a preliminary, please read 3.2.1 - 3.2.2 again! What was said in 7.1. applies, and is specialized as follows.

There are n units of analysis and $k \geq 2$ aspects. The micro and macro data form $k + 1$ similarly partitioned $n \times (1 + p + q)$ regression data matrices

$$[\, y \,\vdots\, X^{\Delta} \,\vdots\, X^{C} \,\vdots\, X^{S}_{h} \,], \quad h = 1, \ldots, k \ ,$$

$$[\, u \,\vdots\, Z^{\Delta} \,\vdots\, Z^{C} \,\vdots\, Z^{S} \,],$$

where $[\, X^{\Delta} \,\vdots\, X^{C} \,]$ and $[\, Z^{\Delta} \,\vdots\, Z^{C} \,]$ correspond to X^{C} and Z^{C} of 3.2.1.

There are $p^{\Delta} \geq 0$ designed and $p^{C} \geq 0$ observed micro regressors X^{Δ} and X^{C} common to the aspects, and $p^{\Delta} + p^{C} = p$. There are $q \geq 0$ observed micro regressors X^{S}_{h} specific to each aspect $h = 1, \ldots, k$, and $p + q \geq 1$. For simplicity, there are no designed specific regressors. Cf. Figure 3.2.1.

The micro regressand vectors y_{h}, ordered $h = 1, \ldots, k$, form the micro regressand $n \times k$ matrix Y. For each $i = 1, \ldots, q$ proceed as follows. Select the $(p + i)$'th regressor vector from each micro regression data matrix. Order these $h = 1, \ldots, k$ to form the i'th specific micro regressor $n \times k$ submatrix X^{*}_{p+i}. The total micro regressor matrix is

$$X^{*} = [\, X^{\Delta} \,\vdots\, X^{C} \,\vdots\, X^{*}_{p+1} \,\vdots\, \ldots\ldots \,\vdots\, X^{*}_{p+q} \,]$$

of order $n \times (p + kq)$. Cf. Figure 3.2.2.

The column vectors of the $n \times q$ macro regressor submatrix Z^{S} will be denoted z^{S}_{p+i}, $i = 1, \ldots q$. Let I^{Δ} and I^{C} denote the unit matrices of orders p^{Δ} and p^{C}. As usual, let j_{k} denote the k-vector of unit elements. The regressor and regressand aggregating functions are as follows.

$$G: \quad [\, Z^{\Delta} \mid Z^{C} \mid z^{S}_{p+1} \mid \cdots\cdots \mid z^{S}_{p+q} \,] =$$

$$= [\, X^{\Delta} \mid X^{C} \mid X^{*}_{p+1} \mid \cdots\cdots \mid X^{*}_{p+q} \,] \begin{bmatrix} I^{\Delta} & 0 & 0 & \cdots\cdots & 0 \\ 0 & I^{C} & 0 & \cdots\cdots & 0 \\ 0 & 0 & j_k & \cdots\cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots\cdots & j_k \end{bmatrix},$$

$$H: \quad u = Yj_k \,.$$

The transformation $(p + kq) \times (p + q)$ matrix of G is block-diagonal. Each of the last q diagonal blocks is j_k. Cf. 5.2.1 and 5.3.1.

For brevity, the regressor aggregating function can be written

$$G: \quad Z = X^{*}G \,,$$

where Z is the macro regressor matrix and G the transformation matrix.

The micro and macro relations

$$\Phi: \quad y_h = X^{\Delta}\beta^{\Delta}_h + X^{C}\beta^{C}_h + X^{S}_h\beta^{S}_h + \varepsilon_h \,, \quad E(\varepsilon_h) = 0 \,, \quad h = 1, \ldots, k \,,$$

$$\Psi: \quad u = Z^{\Delta}\delta^{\Delta} + Z^{C}\delta^{C} + Z^{S}\delta^{S} + \eta \,, \quad E(\eta) = 0$$

are sometimes considered. Alternatively, the k micro models are expressed jointly in terms of the total micro regressor matrix as follows.

$$\Phi: \quad E(Y) = [\, X^{\Delta} \mid X^{C} \mid X^{*}_{p+1} \mid \cdots\cdots \mid X^{*}_{p+q} \,] \begin{bmatrix} B^{\Delta} \\ \hline B^{C} \\ \hline B^{*}_{p+1} \\ \hline \vdots \\ \hline B^{*}_{p+q} \end{bmatrix} .$$

There are q + 2 micro parameter submatrices. The $p^{\Delta} \times k$ matrix B^{Δ} consists of the p^{Δ}-vectors β^{Δ}_h ordered $h = 1, \ldots, k$. The $p^{C} \times k$ matrix B^{C} consists of the p^{C}-vectors β^{C}_h ordered $h = 1, \ldots, k$. Each of the $k \times k$ matrices B^{*}_{p+i}, $i = 1, \ldots, q$, is diagonal. For $h = 1, \ldots, k$, the h'th diagonal element of B^{*}_{p+i}

is the i'th element of β_h^S. The (p+kq) × k parameter matrix in Φ is an augmented micro parameter matrix in the sense of 5.3.1.

For brevity, the micro relations can be written

$$\Phi: \quad E(Y) = X^* B \ ,$$

where B is the total (augmented) micro parameter matrix.

For later purposes define

$$\beta^*_{p+i} = B^*_{p+i} j_k \ , \quad i = 1, \ldots, q \ .$$

Since B^*_{p+i} is diagonal, β^*_{p+i} is its main diagonal k-vector.

The semi-aggregated micro regressor matrix is defined to be identical to the total micro regressor matrix, $X_A = X^*$. Partitioning,

$$X_A^\Delta = X^\Delta \ ,$$

$$X_A^O = [\, X^C \,\vdots\, X^*_{p+1} \,\vdots\, \ldots\ldots \,\vdots\, X^*_{p+q} \,] \ .$$

The matrix of auxiliary regression coefficients can always be taken to be as follows, where I_p is the unit matrix of order $p = p^\Delta + p^C$.

$$C = \begin{bmatrix} I_p & C_1 & \ldots\ldots & C_q \\ 0 & c'_{11} & \ldots\ldots & c'_{q1} \\ \vdots & \vdots & & \vdots \\ 0 & c'_{1q} & \ldots\ldots & c'_{qq} \end{bmatrix}$$

The submatrices c'_{sr} are row k-vectors. Cf. 5.2.1.

7.2.2 The consistency problem

The semi-aggregated micro and semi-disaggregated macro relations are $E_{H\Phi}(u) = \alpha_A + X_A^O \beta_A^O$ and $E_{\Psi G}(u) = \gamma_D + X_A^O \delta_D^O$. For i = 1, ..., q, let δ^S_{p+i} denote the i'th element of δ^S. Then

$$\begin{bmatrix} \alpha_A \\ \hline \\ \beta^O_A \\ \\ \end{bmatrix} = \begin{bmatrix} X^\Delta B^\Delta j_k \\ \hline B^C j_k \\ \beta^*_{p+1} \\ \vdots \\ \beta^*_{p+q} \end{bmatrix}$$

$$\begin{bmatrix} \gamma_D \\ \hline \\ \delta^O_D \\ \\ \end{bmatrix} = \begin{bmatrix} X^\Delta \delta^\Delta \\ \hline \delta^C \\ j_k \delta^S_{p+1} \\ \vdots \\ j_k \delta^S_{p+q} \end{bmatrix}$$

Cf. 5.3.2 and 5.2.2. The following proposition answers the consistency problem.

Proposition 7.2.2

Consider unweighted simple aggregation of aspects.

A necessary and sufficient condition for consistency to be attainable is that

$$\beta^S_1 = \ldots\ldots = \beta^S_k \quad . \qquad \circ\!\circ$$

This is a special case of P.7.1.2. Some parts of the condition stated there are automatically fulfilled, and have been omitted. The critical part is the requirement that for $i = 1, \ldots, q$ the total micro parameter subvector β^*_{p+i} is such that

$$\beta^*_{p+i} = j_k \delta^S_{p+i}$$

for some scalar δ^S_{p+i}. This is so if and only if the i'th elements of $\beta^S_1, \ldots, \beta^S_k$ are all equal. The argument is repeated for $i = 1, \ldots, q$. □

The consistency condition can be formulated in words. For any given $i = 1, \ldots, q$, the following must be true. The micro parameters associated with the $(p + i)$'th micro regressors specific to the k aspects are all equal. The micro parameters associated with regressors common to the aspects, e.g. the inter-

cepts, are not restricted. The k micro models are parallel hyperplanes in the space of the regressand and the q specific regressors.

7.2.3 An expectational interpretation

The micro models $E(Y) = X^*B$ imply the semi-aggregated micro model $E_{H\Phi}(u) = X_A\beta_A$, where $X_A = X^*$ and

$$\beta_A = \begin{bmatrix} B^{\Delta}{}_{j_k} \\ B^{C}{}_{j_k} \\ \beta^*_{p+1} \\ \vdots \\ \beta^*_{p+q} \end{bmatrix}$$

For $r = 1, \ldots, p$, let the r'th element of β_A be denoted $\Sigma\beta_{hr}$, where the sum is over $h = 1, \ldots, k$. Further, $X_A G = Z$.

Expectational interpretations will be given for two kinds of macro regression coefficients; cf. 5.2.3. The first kind, d^S_{p+r}, $r = 1, \ldots, q$, is associated with a macro regressor that is the sum of k micro regressors specific to the aspects.

Proposition 7.2.3.A

Consider unweighted simple aggregation of aspects. Let d be any vector of GLS(W) macro regression coefficients. Let C be any matrix of GLS(W) auxiliary regression coefficients.

If d^S_{p+r} is a unique coefficient function in the macro regression, then

(i) $E_{H\Phi}(d^S_{p+r}) = \sum_{i=1}^{q} c'_{ir}\beta^*_{p+i}$,

(ii) $j'_k c_{ir} = \begin{cases} 1 & \text{if } i = r \\ 0 & \text{if } i \neq r. \end{cases}$ ∘∘

As in the demonstration of P.5.2.3.A, $d^S_{p+r} = \mu'_{p+r} d$, and $\lambda_{p+r} = C'\mu_{p+r}$ is the $(p + r)$'th row vector of C.

The proposition is a special case of P.7.1.3, except that $\theta'\beta_A$ is not explicitly decomposed into k implied micro parameter functions. □

The second kind of macro regression coefficient, d_r^C, $r = 1, \ldots, p$, is associated with a designed or observed regressor common to the aspects. For $i = 1, \ldots, q$, let c'_{ir} denote the r'th row vector of the submatrix C_i of C.

Proposition 7.2.3.B

The same assumptions as in P.7.2.3.A.

If d_r^C is a unique coefficient function in the macro regression, the

(i) $E_{H\Phi}(d_r^C) = \sum_{h=1}^{k} \beta_{hr} + \sum_{i=1}^{q} c'_{ir} \beta^*_{p+i}$,

(ii) $j'_k c_{ir} = 0$, $i = 1, \ldots, q$. ○○

As in the demonstration of P.5.2.3.B, $d_r^C = \mu'_r d$, and $\lambda_r = C'\mu_r$ is the r'th row vector of C.

The same argument as for P.7.2.3.A. □

The expectational interpretation can be formulated in words. The micro expectation of a unique macro coefficient d_r, $r = 1, \ldots, p + q$, is the sum of $p + q$ terms. Each of the first p terms is a weight times the sum of the k micro parameters associated with a regressor common to the aspects. If $i = r$, the i'th weight is unity, otherwise zero. Each of the other q terms is a weighted sum of the k micro parameters associated with a set of corresponding micro regressors specific to the aspects. If $p + i = r$, the i'th weight sum is unity, otherwise zero. Cf. 5.2.3.

As a rule, there are no analogous model-free interpretations.

A simple and instructive demonstration of a somewhat simplified version of P.7.2.3.A and P.7.2.3B is given by Kloek [1961].

7.2.4 A model-free interpretation in deviation form

Consider unweighted simple aggregation of aspects where $p = 1$, and the single common regressor is the intercept regressor. Consider a unique intercept-free coefficient function $\mu'd$ in the GLS(I_n) macro regression. Then $\mu'd$ can be computed from the deviation form macro regression, and interpreted in two steps as follows. Not only micro terms are involved.

Recall the analysis of 3.2.3, especially Figure 3.2.3. In the terminology of

6.4.1, the macro regression is the regression between units in the figure. First, if the macro regression is of full rank, then P.6.4.3 interprets it in terms of the n regressions within units and the total regression. Second, if the total regression is of full rank, then P.6.4.2 interprets it in terms of the k regressions within aspects (i.e. the micro regressions) and the regression between aspects. The distinction between partitioned aggregation of units by summation and by averaging causes no complications.

The suggested interpretation is model-free but hardly useful.

7.2.5 Analysis in terms of regressands and regressors

Unweighted simple aggregation of aspects can be decomposed into two consecutive steps as follows; cf. 3.2.2.

The first step starts from the micro data $[\, Y \,\vdots\, X^* \,]$ and performs an unweighted aggregation of regressands $u = Yj_k$ producing the intermediate data $[\, u \,\vdots\, X^* \,]$.

The second step starts from the intermediate data $[\, u \,\vdots\, X^* \,]$ and performs an unweighted partitioned aggregation of regressors $Z = X^*G$ producing the macro data $[\, u \,\vdots\, Z \,]$.

In the first step, consistency is easily achieved; cf. P.5.3.2. The joint micro models $E(Y) = X^*B$ imply the intermediate model $E(u) = X^*\beta_A$, where the intermediate parameter vector $\beta_A = Bj_k$ is as in 7.2.3.

In the second step, consistency is not automatic. Since the second step is a partitioned aggregation of regressors, P.7.2.2 is a special case of P.5.2.2.

The expectational interpretation in 7.2.3 is primarily in terms of the intermediate parameter vector β_A. Since the second step is a partitioned aggregation of regressors, P.7.2.3.A and P.7.2.3.B are special cases of P.5.2.3.A and P.5.2.3.B.

An analogous model-free interpretation would be primarily in terms of an intermediate coefficient function in the regression of u upon X^*. Since the first step is an aggregation of regressands, such an interpretation could not be brought further; cf. 5.3.3 and 4.3.6.

7.2.6 The connection with Theil's analysis

Unweighted simple aggregation of aspects is more general than "Aggregation over one set of individuals", but less general than "Aggregation over several sets of individuals or commodities", as defined by Theil [1954]. Let the latter be simplified as follows, in Theil's notation. For some λ, $H_\lambda = 1$. For the other λ, $H_\lambda = I$, and it is known that $\beta_{\lambda h,i} = 0$ except when $i = h$. The outcome of the simplification is unweighted simple aggregation of aspects.

If Theil [1954] Theorem 9(ii) is simplified as just indicated, the outcome is P.7.2.2. If Theil [1954] Theorem 2 is similarly simplified, the outcome is P.7.2.3.A and P.7.2.3.B, except that Theil distinguishes explicitly only one micro regressor common to the aspects, the intercept regressor; cf. 5.2.4.

In Theil's terminology, the intermediate parameters $\beta_A = Bj_k$ are "derived microparameters".

8 A FORMAL ANALYSIS: DISCUSSION AND SOME CONCLUSIONS

8.1 The nature of the macro relation derived by Theil

8.1.1 A summary of the formal analysis

The formal analysis is now concluded. It has been concerned with two problems and five types of linear aggregation.

The consistency problem was formulated in 4.2, especially 4.2.7. An answer has been provided for each type and variant of aggregation considered.

The interpretation problem has occurred in two variants. The model-free interpretation was introduced in 4.3., especially 4.3.6. Simple model-free interpretations have been provided for two sub-types of aggregation only. One is partitioned aggregation of regressors, P.5.2.3.A and P.5.2.3.B. The other one is aggregation of sets of units when the macro GLS matrix W is block-diagonal, P.6.3.3. The model-free interpretations suggested in 6.4 and 7.2.4 are not exclusively in micro terms.

The expectational interpretation was introduced in 4.4, especially 4.4.4. An expectational interpretation has been provided for each type and variant of aggregation considered. The weights of the implied micro parameter function(s) were often found to be subject to simple weight-sum relations.

The following summary considers six sub-types of linear aggregation. First, unweighted partitioned aggregation of regressors, 5.2. Second, aggregation of regressands, 5.3. Third, unweighted partitioned aggregation by summation of units of analysis, 6.2.1 - 6.2.3. Fourth, ditto by averaging, 6.2.4. Fifth, aggregation of sets of units, 6.3. Sixth, unweighted simple aggregation of aspects, 7.2.

The restriction in several cases to "unweighted" aggregation could easily be removed. Generalization to the more general variants of aggregation of regressors, 5.1, of units of analysis, 6.1, and of aspects, 7.1, would meet with no problems of principle.

The results of the formal analysis are collected in Table 8.1.1. The model-free interpretation is however not covered.

Consistency will be said to be automatic when, whatever the total micro parameter vector β, a macro parameter vector δ can always be found such that consistency is attained. Otherwise, consistency will be said to be exceptional.

The expectational interpretation will be called data-free when the weights λ (or λ_h) of the implied micro parameter function(s) are fully determined by the weights μ of the interpreted macro coefficient function. Otherwise the expectational interpretation will be called dependent on the micro regressor data and on the macro GLS matrix.

Table 8.1.1 The results of the formal analysis

Sub-type of aggregation	Consistency	Expectational interpretation	Weight-sum relations
unw. partit. of regressors	exceptional P.5.2.2	dependent P.5.2.3.A/B	yes
of regressands	automatic P.5.3.2	data-free P.5.3.3	-
unw. part. by sum. of units	impossible (or automatic) P.6.2.2	dependent (or data-free) P.6.2.3	no
unw. part. by ave. of units	automatic P.6.2.4	data-free P.6.2.4	-
of sets	exceptional P.6.3.2	dependent P.6.3.4	yes
unw. simple of aspects	exceptional P.7.2.2	dependent P.7.2.3.A/B	yes

In those sub-types of aggregation where consistency is exceptional, consistency makes the otherwise dependent expectational interpretation data-free.

The further discussion in 8.1 and 8.2 is primarily about those types of aggregation where consistency is exceptional, the expectational interpretation is de-

pendent on the micro data, and there are simple weight-sum relations for the implied micro parameter function(s). This is the case in (unweighted) simple aggregation of aspects, in (unweighted) partitioned aggregation of regressors, and in aggregation of sets of units. The discussion applies also to inconsistent unweighted partitioned aggregation by summation of units of analysis.

8.1.2 A fundamental question

Consider simple aggregation of aspects, or partitioned aggregation of regressors, or aggregation of sets of units. Consistency is exceptional, the expectational interpretation is dependent on the micro regressor data, and there are weight-sum relations.

Assume given linear regressor and regressand aggregating functions G and H. Consider a given p-columned macro regressor matrix Z such that each macro regression coefficient d_i, $i = 1, \ldots, p$, is a unique macro coefficient function.

The expectational interpretation assumes a micro relation Φ that consists of $k \geq 1$ incomplete linear models with parameter vectors β_h, $h = 1, \ldots, k$. It interprets the macro regression coefficients as follows.

$$E_{H\Phi}(d_i) = \sum_{h=1}^{k} \lambda'_{ih} \beta_h \quad , \quad i = 1, \ldots, p \quad .$$

As a rule, the weight vectors λ_{ih} of the implied micro parameter functions are dependent on the (macro and) micro regressor data, and on the macro GLS matrix W.

As stated in 1.3.3, Theil [1954] postulates a linear macro relation in the macro variables, say Ψ_T, whose parameters are

$$\delta_{Ti} = E_{H\Phi}(d_i) \quad , \quad i = 1, \ldots, p \quad .$$

As a rule, Theil's macro parameters are thus functions of the micro regressor data. Theil discusses this property of δ_{Ti} in Theil [1954], section 2.5.

Theil's use of the expectational interpretation should be confronted with the consistency problem.

The consistency problem considers a micro relation Φ as above, and a macro

relation Ψ that consists of an incomplete linear model with parameter vector δ. It establishes conditions that $\{\beta_1, \ldots, \beta_k\}$ must satisfy if there is to exist a δ consistent with them. These conditions are very restrictive.

The following conclusion can be drawn. As a rule, Theil's macro relation Ψ_T cannot be an incomplete linear model. In particular, its disturbance terms, say η_{Tj}, cannot be those of an incomplete linear model. Theil discusses their properties in Theil [1954], section 6.3. See also Wu [1973].

What kind of a relation is the macro relation Ψ_T derived by Theil?

8.1.3 An algebraical example

The fundamental question just asked will be discussed in the remainder of 8.1, and in 8.2. The discussion will be in terms of simple aggregation of aspects; this is rather customary. Analogous arguments apply to partitioned aggregation of regressors, and to aggregation of sets of units; cf. the end of 8.1.1.

Consider unweighted simple aggregation of $k \geq 2$ aspects indexed $h = 1, \ldots, k$. There are the intercept regressor and a single observed regressor specific to the aspects. There are $n \geq 3$ units of analysis indexed $j = 1, \ldots, n$. The rank of the $n \times 2$ macro regressor matrix is assumed to be 2.

The regressor aggregating function G, the regressand aggregating function H, and the micro relation Φ are as follows.

$$G: \quad z_j = \sum_{h=1}^{k} x_{hj} \quad , \quad j = 1, \ldots, n \ .$$

$$H: \quad u_j = \sum_{h=1}^{k} y_{hj} \quad , \quad j = 1, \ldots, n \ .$$

$$\Phi: \quad y_{hj} = \beta_h^{\Delta} + \beta_h^{O} x_{hj} + \epsilon_{hj} \ ; \ E(\epsilon_{hj}) = 0 \ ; \ h = 1, \ldots, k; \ j = 1, \ldots, n.$$

By P.7.2.2, a necessary condition for consistency to be attainable is that $\beta_1^O = \ldots = \beta_k^O$. This condition is assumed not to be satisfied.

Only common least-squares will be considered. The macro regression line fitted by common least-squares regression is denoted

$$u = d^{\Delta} + d^{O} z \ .$$

The common least-squares auxiliary regressions, say A, and their residuals, are denoted

$$A: \quad x_{hj} = c_h^{\Delta} + c_h^{O} z_j + v_{hj} \quad , \quad h = 1, \ldots, k \quad .$$

The auxiliary regressions are model-free. The macro regression coefficients can be interpreted by means of the micro relation Φ.

Theil's macro slope and intercept parameters are defined by the expectational interpretations of the macro slope and intercept coefficients. These interpretations, and the associated weight-sum relations, are given by P.7.2.3.A and P.7.2.3.B. The macro parameters derived by Theil are thus as follows.

$$(T) \quad \begin{cases} \delta_T^{O} = \sum_{h=1}^{k} c_h^{O} \beta_h^{O} \quad , \\ \text{where} \quad \sum_{h=1}^{k} c_h^{O} = 1 \; ; \\ \delta_T^{\Delta} = \sum_{h=1}^{k} \beta_h^{\Delta} + \sum_{h=1}^{k} c_h^{\Delta} \beta_h^{O} \quad , \\ \text{where} \quad \sum_{h=1}^{k} c_h^{\Delta} = 0 \quad . \end{cases}$$

Via the auxiliary regression coefficients, δ_T^{O} and δ_T^{Δ} are dependent on the micro regressor data.

If the auxiliary regressions A are substituted into the semi-aggregated micro relation $H\Phi$, the outcome is the following linear relation in the macro variables.

$$H\Phi A: \; u_j = \sum_{h=1}^{k} \beta_h^{\Delta} + \sum_{h=1}^{k} c_h^{\Delta} \beta_h^{O} + \left(\sum_{h=1}^{k} c_h^{O} \beta_h^{O} \right) z_j +$$

$$+ \sum_{h=1}^{k} \beta_h^{O} v_{hj} + \sum_{h=1}^{k} \epsilon_{hj} \quad , \quad j = 1, \ldots, n \quad .$$

The intercept and slope of $H\Phi A$ are Theil's δ_T^{Δ} and δ_T^{O}.

The relation $H\Phi A$ derived from Φ is the macro relation Ψ_T derived by Theil. In symbols,

$$\Psi_T = H\Phi A \quad .$$

This analysis of Ψ_T is implicit in Theil [1954]. Perhaps the first explicit statement of it is Allen [1956], section 20.2, where however the micro disturbances ϵ_{hj} are omitted. The complete version is found i.a. in Kloek [1961].

Theil's macro relation can be written as follows.

$$\Psi_T: \left\{ \begin{array}{l} u_j = \delta_T^{\Delta} - \delta_T^{O} z_j - \eta_{Tj} \quad , \\ \\ E_{\Phi}(\eta_{Tj}) = \sum_{h=1}^{k} \beta_h^{O} v_{hj} \quad , \end{array} \right\} \qquad j = 1, \ldots, n \quad .$$

It may be represented in figures by the line

$$u = \delta_T^{\Delta} + \delta_T^{O} z \quad .$$

The macro relation Ψ_T is derived from the micro relation Φ. It remains to investigate what properties of Ψ_T follow from those of Φ.

8.1.4 Three numerical examples

The algebraical example of 8.1.3 will now be specialized into three different numerical examples. Throughout, "lf(ε)" denotes any homogeneous linear function of the micro disturbances ε_{hj}.

Examples 8.1.4.1 and 8.1.4.2. $k = 2$ and $n = 4$. The micro parameters are as follows.

$$E_{\Phi} y_{1j} = \beta_1^{\Delta} + \beta_1^{O} x_{1j} \equiv 15 + x_{1j} \; ;$$

$$E_{\Phi} y_{2j} = \beta_2^{\Delta} + \beta_2^{O} x_{2j} \equiv 5 + 2x_{2j} \quad .$$

Example 8.1.4.1. The micro and macro data are as in Table 8.1.4.1.A. The following results are easily verified.

$$d^{O} = -\frac{1}{2} + \mathrm{lf}(\varepsilon), \qquad d^{\Delta} = 58 + \mathrm{lf}(\varepsilon) \; ;$$

$$\delta_T^{O} = -\frac{1}{2}, \qquad \delta_T^{\Delta} = 58 \; ;$$

$$c_1^{O} = \frac{5}{2}, \qquad c_1^{\Delta} = -38 \; ;$$

$$c_2^{O} = -\frac{3}{2}, \qquad c_2^{\Delta} = 38 \quad .$$

The relations (T) of 8.1.3 are satisfied. The disturbances of the macro relation Ψ_T are as in Table 8.1.4.1.B.

Table 8.1.4.1.A Micro and macro data in example 8.1.4.1

j	x_{1j}	y_{1j}	x_{2j}	y_{2j}	z_j	u_j
1	8	$23 + \varepsilon_{11}$	11	$27 + \varepsilon_{21}$	19	$50 + \varepsilon_{11} + \varepsilon_{21}$
2	11	$26 + \varepsilon_{12}$	8	$21 + \varepsilon_{22}$	19	$47 + \varepsilon_{12} + \varepsilon_{22}$
3	13	$28 + \varepsilon_{13}$	8	$21 + \varepsilon_{23}$	21	$49 + \varepsilon_{13} + \varepsilon_{23}$
4	16	$31 + \varepsilon_{14}$	5	$15 + \varepsilon_{24}$	21	$46 + \varepsilon_{14} + \varepsilon_{24}$

Table 8.1.4.1.B Macro disturbances in example 8.1.4.1

j	v_{1j}	v_{2j}	η_{Tj}
1	$-\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2} + \varepsilon_{11} + \varepsilon_{21}$
2	$\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{3}{2} + \varepsilon_{12} + \varepsilon_{22}$
3	$-\frac{3}{2}$	$\frac{3}{2}$	$\frac{3}{2} + \varepsilon_{13} + \varepsilon_{23}$
4	$\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{3}{2} + \varepsilon_{14} + \varepsilon_{24}$

Example 8.1.4.2. The micro and macro data are as in Table 8.1.4.2.A. The following results are easily verified.

$$d^O = 1 + \mathrm{lf}(\varepsilon), \qquad d^\Delta = 32 + \mathrm{lf}(\varepsilon);$$

$$\delta_T^O = 1, \qquad \delta_T^\Delta = 32;$$

$$c_1^O = 1, \qquad c_1^\Delta = -12;$$

$$c_2^O = 0, \qquad c_2^\Delta = 12.$$

The relations (T) of 8.1.3 are satisfied. The disturbances of the macro relation Ψ_T are as in Table 8.1.4.2.B.

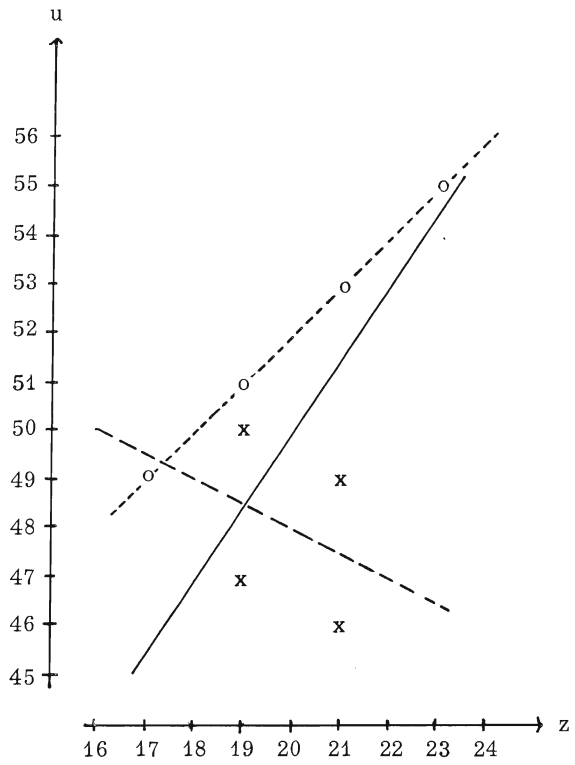
Table 8.1.4.2.A Micro and macro data in example 8.1.4.2

j	x_{1j}	y_{1j}	x_{2j}	y_{2j}	z_j	u_j
1	5	$20 + \varepsilon_{11}$	12	$29 + \varepsilon_{21}$	17	$49 + \varepsilon_{11} + \varepsilon_{21}$
2	7	$22 + \varepsilon_{12}$	12	$29 + \varepsilon_{22}$	19	$51 + \varepsilon_{12} + \varepsilon_{22}$
3	9	$24 + \varepsilon_{13}$	12	$29 + \varepsilon_{23}$	21	$53 + \varepsilon_{13} + \varepsilon_{23}$
4	11	$26 + \varepsilon_{14}$	12	$29 + \varepsilon_{24}$	23	$55 + \varepsilon_{14} + \varepsilon_{24}$

Table 8.1.4.2.B Macro disturbances in example 8.1.4.2

j	v_{1j}	v_{2j}	η_{Tj}
1	0	0	$\varepsilon_{11} + \varepsilon_{21}$
2	0	0	$\varepsilon_{12} + \varepsilon_{22}$
3	0	0	$\varepsilon_{13} + \varepsilon_{23}$
4	0	0	$\varepsilon_{14} + \varepsilon_{24}$

Examples 8.1.4.1 and 8.1.4.2 simplified. The analysis above remains valid when $\varepsilon_{hj} = 0$ for $h = 1,2$ and $j = 1,2,3,4$. Figure 8.1.4.A illustrates this simplified (or degenerate) situation.



x --------- data and relation in example 8.1.4.1

o ---------- data and relation in example 8.1.4.2

——————— plausible relation, see 8.2.1

Figure 8.1.4.A Macro data and relations in examples 8.1.4.1 and 8.1.4.2 when all micro disturbances vanish

Example 8.1.4.3. $k = 4$ and $n = 5$. The micro parameters are as follows.

$$E_{\Phi} y_{1j} = \beta_1^{\Delta} + \beta_1^{O} x_{1j} \equiv \quad 15 + x_{1j} ;$$

$$E_{\Phi} y_{2j} = \beta_2^{\Delta} + \beta_2^{O} x_{2j} \equiv \quad 5 + 2x_{2j} ;$$

$$E_{\Phi} y_{3j} = \beta_3^{\Delta} + \beta_3^{O} x_{3j} \equiv - \; 5 + 3x_{3j} ;$$

$$E_{\Phi} y_{4j} = \beta_4^{\Delta} + \beta_4^{O} x_{4j} \equiv - \; 15 + 4x_{4j} .$$

The micro and macro data as as in Table 8.1.4.3.A. The following results are easily verified.

$$d^{O} = \frac{5}{2} + \mathrm{lf}(\varepsilon), \qquad d^{\Delta} = \mathrm{lf}(\varepsilon) ;$$

$$\delta_T^{O} = \frac{5}{2} , \qquad \delta_T^{\Delta} = 0 ;$$

$$c_1^{O} = \frac{9}{10} , \qquad c_1^{\Delta} = - \; 25 ;$$

$$c_2^{O} = - \frac{6}{10} , \qquad c_2^{\Delta} = \quad 31 ;$$

$$c_3^{O} = 0 , \qquad c_3^{\Delta} = \quad 13 ;$$

$$c_4^{O} = \frac{7}{10} , \qquad c_4^{\Delta} = - \; 19 .$$

The relations (T) of 8.1.3 are satisfied. The disturbances of the macro relation Ψ_T are as in Table 8.1.4.3.B.

Table 8.1.4.3.A Micro and macro data in example 8.1.4.3

j	x_{1j}	x_{2j}	x_{3j}	x_{4j}	Ey_{1j}	Ey_{2j}	Ey_{3j}	Ey_{4j}	z_j	Eu_j
1	6	11	13	6	21	27	34	9	36	91
2	9	8	13	8	24	21	34	17	38	96
3	15	3	13	9	30	11	34	21	40	96
4	11	8	13	10	26	21	34	25	42	106
5	14	5	13	12	29	15	34	33	44	111

Table 8.1.4.3.B Macro disturbances in example 8.1.4.3

j	v_{1j}	v_{2j}	v_{3j}	v_{4j}	η_{Tj}
1	$-\frac{14}{10}$	$\frac{16}{10}$	0	$-\frac{2}{10}$	$1 + \Sigma \epsilon_{h1}$
2	$-\frac{2}{10}$	$-\frac{2}{10}$	0	$\frac{4}{10}$	$1 + \Sigma \epsilon_{h2}$
3	$\frac{40}{10}$	$-\frac{40}{10}$	0	0	$-4 + \Sigma \epsilon_{h3}$
4	$-\frac{18}{10}$	$\frac{22}{10}$	0	$-\frac{4}{10}$	$1 + \Sigma \epsilon_{h4}$
5	$-\frac{6}{10}$	$\frac{4}{10}$	0	$\frac{2}{10}$	$1 + \Sigma \epsilon_{h5}$

The analysis above remains valid when $\epsilon_{hj} = 0$ for $h = 1,2,3,4$ and $j = 1, \ldots, 5$. Figure 8.1.4.B illustrates this simplified (or degenerate) situation.

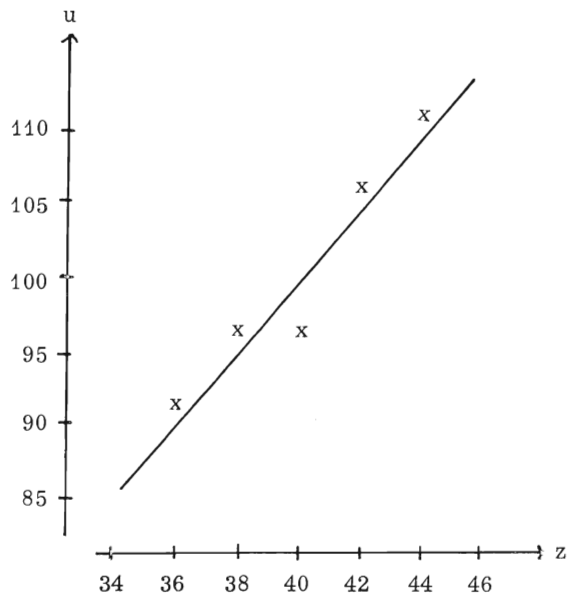


Figure 8.1.4.B Macro data and relation in example 8.1.4.3 when all micro disturbances vanish

8.1.5 Systematic disturbances and variable parameters

The macro relation Ψ_T derived by Theil will now be compared with the incomplete linear model of 2.3.1, and with certain generalizations of the latter.

The incomplete linear model is concerned with a given n-rowed regressor matrix Z . With each row vector z_j' of regressor data occurring in Z there is associated a stochastic regressand element u_j . The incomplete linear model partly specifies the distribution of u_j as follows.

$$u_j = z_j'\delta + \eta_j , \quad E(\eta_j) = 0 .$$

The same parameters δ apply to each row $j = 1, \ldots, n$ of Z . The disturbances η_j have zero expectation.

In the macro relation Ψ_T of 8.1.3, the micro relation Φ implies that as a rule

$$E_\Phi(\eta_{Tj}) \neq 0 .$$

Examples 8.1.4.1 and 8.1.4.3 illustrate this property of Ψ_T , see Figures 8.1.4.A and 8.1.4.B. The macro disturbances are in this sense systematic.

If the specification $E(\eta_j) = 0$ is simply dropped, the incomplete linear model of 2.3.1 becomes vacuous. The following generalization permits $E(\eta_j) \neq 0$ but avoids vacuity. There exists a single-valued function $f(\cdot)$ such that

$$u_j = z_j'\delta + \eta_j , \quad E(\eta_j) = f(z_j') .$$

The expected value of u_j is then still uniquely determined by the vector z_j' .

In the macro relation Ψ_T of 8.1.3, there need not exist such a function $f(\cdot)$. Example 8.1.4.1 illustrates this, see Figure 8.1.4.A. The expected macro disturbance $E_\Phi(\eta_{Tj})$ is a function of the micro regressor data aggregated into the corresponding macro data z_j' .

When the linear model is applied in practice, and the regressors are not all designed (2.3.3), the following generalization is usually taken for granted. Many row vectors z_r' of regressor data that did not occur in Z could have done so; but it is hardly ever stated precisely what vectors z_r' are eligible. Associated with any such vector z_r' there would be a stochastic regressand element u_r . The linear relation $u = z'\delta$ that is valid for the given n units of analysis would

be valid for any new units of analysis too. This generlization is probably inspired by the regression model (2.3.4).

In the macro relation Ψ_T of 8.1.3, the macro parameters δ_T^O and δ_T^Δ derived by means of the micro relation Φ are not independent of the regressor data. Examples 8.1.4.1 and 8.1.4.2 jointly illustrate this property of Ψ_T, see Figure 8.1.4.A. The total micro parameter vector β is identical in the two examples, but the macro parameters differ.

The macro parameters δ_T are functions, via the auxiliary regression coefficients, of the total set of micro regressor data. If in example 8.1.4.2, units $j = 1$ and $j = 4$ are replaced by replications of units $j = 2$ and $j = 3$, then the two examples have identical macro regressor matrices. Yet the macro parameters differ. The macro parameters are in this sense variable.

In summary, the micro models Φ imply the following two important properties of the macro relation Ψ_T derived by Theil.

The first property is for a given set of micro regressor data. The macro disturbances are systematic. The usually non-zero expected value of the macro disturbance of a given unit of analysis is affected by the micro regressor data of the unit. Thus the macro relation Ψ_T in the macro variables does not provide unbiased forecasts of the macro regressand.

The second property is concerned with changing the micro regressor data. The macro parameters of Ψ_T are not determined by the micro parameters of Φ alone. They are affected also by the total set of micro regressor data given. As a rule, two different sets of micro regressor data give different macro parameter vectors δ_T, and in fact different macro disturbance vectors η_T too. Thus when the set of micro regressor data is changed, nothing in the macro relation Ψ_T need remain permanent.

The macro relation Ψ_T is presumably intended for use also when the micro regressor data underlying the actual macro regressor data are not known. Its parameters δ_T are however not invariant w.r.t. the unknown micro regressor data. To call δ_T macro parameters may therefore by rather misleading.

8.1.6 On zero means and correlations

The macro relation Ψ_T derived by Theil is not a generalized linear model in the sense discussed in 8.1.5. It is perhaps tempting to argue that it is nevertheless essentially equivalent to a linear model in the following sense. "The macro disturbance η_T has zero mean, and zero correlation with any non-intercept macro regressor." Is this argument correct?

Two kinds of means and correlations must be distinguished. First, if x and y are two random variables, their joint distribution determines two means μ_x and μ_y, and a correlation ρ_{xy}. Such means and correlations will be called distributional. Second, if x_j and y_j, $j = 1, \ldots, n$, are two series of given numbers, one can compute from them two means $\bar{x}$ and $\bar{y}$, and a correlation r_{xy}. Such means and correlations will be called model-free (or descriptive).

Again, the explicit attention will be limited to the class of aggregations considered in 8.1.3. Three different models of the form

$$u_j = \delta^{\Delta} + \delta^{O} z_j + \eta_j$$

will be compared. One of them is Theil's model Ψ_T. There, the macro disturbance is

$$\eta_{Tj} = \sum_{h=1}^{k} \beta_h^{O} v_{hj} + \sum_{h=1}^{k} \epsilon_{hj} ,$$

where v_{hj} are residuals from the auxiliary regressions.

First, consider a regression model in the sense of 2.3.4. The standard assumptions are as follows. For any j, the distributional mean of η_j is zero. For any j and k, inclusive of k = j, the distributional correlation between η_j and z_k is zero. On the other hand, consider a given realization (η_j, z_j), $j = 1, \ldots, n$ of the regression model. As a rule, the model-free mean $\bar{\eta}$ and the model-free correlation $r_{\eta z}$, computed from the realization, are not exactly zero.

Second, consider a linear model in the sense of 2.3.1. The standard assumptions are as follows. For any j, the distributional mean of η_j is zero. Since z_k is a non-stochastic constant, the distributional correlation between η_j and z_k is zero by definition. On the other hand, consider a given realization η_j, $j = 1, \ldots, n$, of the linear model for given z_j, $j = 1, \ldots, n$. As a rule, the model-free mean $\bar{\eta}$ and the model-free correlation $r_{\eta z}$, computed from the

realization, are not exactly zero.

Third, consider Theil's model Ψ_T and the model-free properties of a given realization η_{Tj}, $j = 1, \ldots, n$, for given z_j, $j = 1, \ldots, n$. By P.2.2.3(ii), for $h = 1, \ldots, k$ the following is true. The model-free mean of the h'th auxiliary residual v_h is zero, and so is its model-free correlation with z. But for $h = 1, \ldots, k$ the following is also true. The model-free mean of the h'th micro disturbance ε_h is not exactly zero, nor is its model-free correlation with z. Thus the model-free mean of η_T, and its model-free correlation with z, are both as a rule not zero.

Fourth, consider Ψ_T and the distributional properties of its disturbances η_{Tj}, $j = 1, \ldots, n$. Since z_k is a non-stochastic constant, the distributional correlation between η_{Tj} and z_k is zero by definition. But the distributional mean of η_{Tj}, as determined by Φ, is as a rule not zero.

Is Ψ_T essentially equivalent to a linear model in the sense indicated at the beginning of 8.1.6? The model-free mean of η_T and its model-free correlation with z are not relevant to this question. Moreover, they are not zero. It is the distributional means of η_{Tj} and their distributional correlations with z_k that matter. The distributional correlations are indeed zero by definition. But the distributional means of η_{Tj} are not zero. Therefore Ψ_T is not essentially equivalent to a linear model.

8.1.7 A simple reinterpretation

Once more, the explicit argument will be in terms of the class of simple aggregations of aspects of 8.1.3. Analogous arguments apply to other classes and types of aggregation.

The parameters of the macro relation Ψ_T according to Theil are derived by means of two consecutive operations. The first operation computes the macro regression coefficients d^O and d^Δ. The second operation takes the micro expectations $E_{H\Phi}(d^O)$ and $E_{H\Phi}(d^\Delta)$ of the macro regression coefficients. Both operations are linear in the macro regressand data u_j, $j = 1, \ldots, n$. If the order of the two operations is reversed, the result remains unaffected.

Let the order be reversed. The first operation takes the micro expectations $E_{H\Phi}(u_j)$ of the macro regressand data,

$$E_{H\Phi}(u_j) = \sum_{h=1}^{k} \beta_h^{\Delta} + \sum_{h=1}^{k} \beta_h^{O} x_{hj} \quad , \quad j = 1, \ldots, n \ .$$

By substitution of the auxiliary regressions,

$$E_{H\Phi}(u_j) = \sum_{h=1}^{k} \beta_h^{\Delta} + \sum_{h=1}^{k} c_h^{\Delta} \beta_h^{O} + \left(\sum_{h=1}^{k} c_h^{O} \beta_h^{O} \right) z_j + \sum_{h=1}^{k} \beta_h^{O} v_{hj} \ , \ j = 1, \ldots, n.$$

Compare with HΦA of 8.1.3.

The second operation computes the slope δ_T^O and intercept δ_T^{Δ} in the regression of $E_{H\Phi}(u)$ upon z. Since for $h = 1, \ldots, k$ the h'th auxiliary residual v_h has zero model-free mean and correlation with z, the outcome is as follows.

$$\delta_T^O = \sum_{h=1}^{k} c_h^O \beta_h^O \ ,$$

$$\delta_T^{\Delta} = \sum_{h=1}^{k} \beta_h^{\Delta} + \sum_{h=1}^{k} c_h^{\Delta} \beta_h^O \ .$$

The second operation is a model-free regression computed from the data

$$\{z_j \ , \ E_{H\Phi}(u_j)\} \ , \quad j = 1, \ldots, n \ .$$

The regressand has been derived by means of the micro models Φ, but the regression itself refers to no model.

Three examples of the model-free regression of $E_{H\Phi}(u)$ upon z are shown in Figures 8.1.4.A and 8.1.4. B.

A simple reinterpretation of Theil's macro parameters δ_T can now be formulated in general terms. The macro parameters derived by Theil are the regression coefficients in the model-free regression of $E_{H\Phi}(u)$ upon Z, where $E_{H\Phi}(u)$ is the micro expectation of the macro regressand, and Z is the set of macro regressors.

8.2 The concept of aggregation bias

8.2.1 Some terminology

The macro parameters derived by Theil are linear functions of the micro parameters. If the aggregation is sufficiently simple, it is tempting to guess the weights of those functions. Thus in the class of simple aggregations of aspects

introduced in 8.1.3, the following linear functions have som intuitive appeal.

$$\delta_P^O = \frac{1}{k} \sum_{h=1}^{k} \beta_h^O \quad ,$$

$$\delta_P^\Delta = \sum_{h=1}^{k} \beta_h^\Delta \quad .$$

Macro parameters defined by such a priori plausible weight systems will be called plausible macro parameters.

No attempt will be made to formulate a general definition of plausible macro parameters. It is doubtful whether such a rule could be agreed upon even for simple aggregation of aspects, unless the aggregation is unweighted. The discussion below in 8.2 is limited to the class of aggregations of 8.1.3.

The difference between a macro parameter as derived by Theil, and the corresponding plausible macro parameter, is called by Theil [1954], moment 6.4.5, an aggregation bias. In 8.1.3 there are the following two aggregation biases.

$$\theta^O = \delta_T^O - \delta_P^O = \sum_{h=1}^{k} (c_h^O - \frac{1}{k}) \beta_h^O \quad ,$$

$$\theta^\Delta = \delta_T^\Delta - \delta_P^\Delta = \sum_{h=1}^{k} c_h^\Delta \beta_h^O \quad .$$

Because of the weight-sum relations for the auxiliary regression coefficients, these aggregation biases can be expressed in terms of certain model-free covariances across the aspects, between auxiliary regression coefficients and micro parameters. See Theil [1954], moment 2.3.3.

The aggregation biases in the three examples of 8.1.4 are listed in Table 8.2.1.

Table 8.2.1 Aggregation biases in examples 8.1.4.1 - 8.1.4.3

example	δ_T^O	δ_P^O	θ^O	δ_T^Δ	δ_P^Δ	θ^Δ
8.1.4.1	$-\frac{1}{2}$	$\frac{3}{2}$	- 2	58	20	38
8.1.4.2	1	$\frac{3}{2}$	$-\frac{1}{2}$	32	20	12
8.1.4.3	$\frac{5}{2}$	$\frac{5}{2}$	0	0	0	0

The plausible macro relation

$$\Psi_P : u = \delta_P^{\Delta} + \delta_P^{O} z$$

common to examples 8.1.4.1 and 8.1.4.2 has been drawn into Figure 8.1.4.A. In example 8.1.4.3, Figure 8.1.4.B, Ψ_P coincides with Ψ_T.

8.2.2 Two conflicting points of view

In standard statistical terminology, a bias is the expected difference between an estimator p and the parameter π estimated, $E(p) - \pi$. It is natural to try to interpret the term "aggregation bias" accordingly. The estimator is a macro regression coefficient. For example, $p = d^O$ and $E(p) = \delta_T^O$. As to what is the parameter estimated, two different points of view are possible.

The first point of view is that the parameter estimated is the corresponding plausible macro parameter, $\pi = \delta_P^O$. The aggregation bias θ^O is then a bias in the ordinary sense of the word. If this point of view is adopted, the macro parameter δ_T^O derived by Theil loses its status as the real macro parameter.

The second point of view is that the parameter estimated is the macro parameter δ_T^O derived by Theil. If this point of view is adopted, the estimator d^O is unbiased, and the aggregation bias is not a bias in the ordinary sense of the word. The plausible macro parameter δ_P^O and the aggregation bias θ^O lose most of their interest.

A somewhat sharp formulation of the two conflicting points of view is as follows. Either the macro parameters derived by Theil are not parameters, or else the aggregation biases are not biases.

8.2.3 Aggregation bias and consistency

The interrelations of three possible properties of an aggregation will now be discussed. Again, the discussion is limited to the class of aggregations of 8.1.3.

The first property is that consistency in the sense of 4.2.4 is attainable. By P.7.2.2, a necessary and sufficient condition is as follows.

$$\beta_1^O = \ldots\ldots = \beta_h^O = \ldots\ldots = \beta_k^O \ .$$

This property will be abbreviated CA (consistency attainable).

The second property is that the expected disturbances in the macro relation Ψ_T vanish. From Ψ_T of 8.1.3,

$$E_{\Phi}(\eta_{Tj}) = \sum_{h=1}^{k} \beta_h^O v_{hj} = 0 , \quad j = 1, \ldots, n .$$

This property will be abbreviated VMD (vanishing macro disturbances).

The third property is that the aggregation biases vanish. From 8.2.1,

$$\theta^O = \sum_{h=1}^{k} (c_h^O - \frac{1}{k}) \beta_h^O = 0 ,$$

$$\theta^{\Delta} = \sum_{h=1}^{k} c_h^{\Delta} \beta_h^O = 0 .$$

This property will be abbreviated VAB (vanishing aggregation biases).

If the auxiliary regressions A of 8.1.3 are summed over the aspects, the result is

$$z_j = \sum_{h=1}^{k} c_h^{\Delta} + \left(\sum_{h=1}^{k} c_h^O \right) z_j + \sum_{h=1}^{k} v_{hj} .$$

By the weight-sum relations in (T) of 8.1.3, this implies

$$\sum_{h=1}^{k} v_{hj} = 0 , \quad j = 1, \ldots, n .$$

This is a restriction on the auxiliary residuals.

By the restriction on the auxiliary residuals, CA implies VMD. By the weight-sum relations (T) for auxiliary regression coefficients, CA implies VAB.

Example 8.1.4.2 shows that VMD implies neither VAB nor CA. Cf. Figure 8.1.4.A and Table 8.2.1.

Example 8.1.4.3 shows that VAB implies neither VMD nor CA. Cf. Figure 8.1.4.B and Table 8.2.1.

Thus either of VMD and VAB is a necessary but not sufficient condition for CA. In particular, vanishing aggregation biases do not imply consistency.

8.2.4 Some references to literature

Theil [1954] assumes a micro relation Φ that consists of linear models, and derives a macro relation $\Psi_T = H\Phi A$. That Ψ_T is not an incomplete linear model

is made clear in Theil [1954], moment 6.3.4.

Both Φ and Ψ_T consist of linear equations plus disturbances. To economists and other non-statisticians the formal differences between Φ and Ψ_T may not appear important. Such neglect of statistical technicalities as in fact encouraged by Theil [1954] himself in section 6.1 on "aggregation in a changing economy". As stated in 3.2.5 above, this is a sequence of two aggregations. The first stage produces intermediate relations of the type Ψ_T. These are then aggregated in the second stage just as if they were linear models. Further, in the second stage Theil uses the model-free approach of 6.4.2 to establish an expectational interpretation, which is also, by 6.4.4, questionable.

Boot and de Wit [1960] study empirically a simple aggregation of aspects. The authors replace the unknown micro parameters by their least-squares estimates. They then decompose each estimated macro parameter $\hat{\delta}_{Ti}$ or disturbance $\hat{\eta}_{Tj}$ into three parts, on a "true" parameter or disturbance, the second one an aggregation bias, and the third one an implied sampling error. In the notation of 8.1.3, the "true" parts correspond to

$$\frac{1}{k} \sum_{h=1}^{k} \beta_h^O \quad \text{and} \quad \sum_{h=1}^{k} \epsilon_{hj} \quad ,$$

while the aggregation biases correspond to

$$\sum_{h=1}^{k} \left(c_h^O - \frac{1}{k} \right) \beta_h^O \quad \text{and} \quad \sum_{h=1}^{k} \beta_h^O v_{hj} \quad .$$

Since 8.1.3 is in model terms, the implied sampling errors have no counterpart there.

The numerical decompositions of the estimated macro parameters and disturbances are given in Boot and de Wit [1960], Tables 3 and 6. The authors then proceed to compute the model-free variances and covariances across the n units of analysis of the three components of the estimated macro disturbance, Table 7. This procedure sweeps under the carpet the difference between the "true" macro disturbance $\Sigma_h \epsilon_{hj}$, which is stochastic, and the "aggregation bias" $\Sigma_h \beta_h^O v_{hj}$, which is not; cf. 8.1.6. Generalization to other data would therefore be questionable.

An analogous empirical study is Gupta [1971]. That $\Sigma_h \epsilon_{hj}$ is stochastic while $\Sigma_h \beta_h^O v_{hj}$ is not is indicated in Theil [1954], moment 6.3.5.

The term "aggregation bias" is used differently by certain other authors. A test for the absence of aggregation bias in Zellner [1962a], section 4, is a test for consistency in simple aggregation of aspects. The test is unnecessarily restrictive, for the hypothesis tested includes that the micro intercept parameters agree; cf. the end of 7.2.2. The aggregation bias referred to by Ijiri [1971], section 4.2, is the bias in the prediction of the macro regressand caused by the aggregation.

8.3 On the coefficient of determination in aggregation

8.3.1 On the coefficient of determination in the linear model

The coefficient of determination R^2 was defined in 2.5.3. Its definition makes no reference to a linear model.

Consider a given regressor matrix X and a complete linear model as defined in 2.3.1. The parameters of the model are the parameter vector β, the covariance matrix $\sigma^2 W$, and (cf. 2.3.4) the regressor matrix X.

Corresponding to (almost) any realization of the regressand vector y, there is a coefficient of determination R^2. Such an R^2 characterizes primarily the set $[\, y \mid X \,]$ of regression data from which it is computed.

Attempts have been made to define an additional parameter P^2 for the complete linear model (not the regression model!), such that P^2 is somehow a theoretical counterpart of R^2. One alternative is suggested by Barten [1962], another one by Koerts and Abrahamse [1969].

The latter authors show that the distribution of R^2 implied by the normal linear model depends on X. They warn against the use of R^2 to compare different linear models, Koerts and Abrahamse [1969], section 8.7.

The coefficient of determination R^2 is a model-free concept. Its conceptual relation to a postulated linear model is not simple.

8.3.2 On aggregation of aspects

Grunfeld and Griliches [1960] refer to an empirical study of simple aggregation of eight aspects. They consider three kinds of fitted relations.

The first set of relations consists of a fitted micro relation $\hat{\varphi}_h$ for each aspect $h = 1, \ldots, k$. The coefficients of determination will be denoted r_h^2, $h = 1, \ldots, k$.

The second relation is a fitted macro relation $\hat{\Psi}$. The authors call this the aggregate relation. Its coefficient of determination will be denoted R_a^2.

The third relation is the indirectly fitted semi-aggregated micro relation $H\hat{\Phi}$ defined by the regressand aggregating function and the fitted micro relations. The authors call this the composite relation. Its coefficient of determination will be denoted R_c^2.

First, the authors find that, somewhat unexpectedly,

$$R_a^2 > r_h^2 \quad , \quad h = 1, \ldots, k \quad .$$

They manage to find an explanation in terms of the joint behaviour of the micro regressors. The explanation is model-free; see Grunfeld and Griliches [1960], appendices A to E.

Second, the authors find that, very unexpectedly,

$$R_a^2 > R_c^2 \quad .$$

They fail to find an illuminating model-free explanation of this phenomenon. Instead, they argue that inclusion of the macro regressors in the set of micro regressors for every aspect would be likely to change the relation between the aggregate and composite coefficients of determination. Their tentative conclusion is that the micro models are probably mis-specified and should include the macro regressors. See Grunfeld and Griliches [1960], section IV.

The second finding has been further considered by Green [1964], section 12.4, and by Gupta [1969].

The question why in simple aggregation of aspects sometimes $R_a^2 > R_c^2$ is quite intriguing. At least three different kinds of answers may be sought. The first kind is entirely model-free. The second kind involves the k micro models but keeps the micro regressor data fixed. The third kind of answers would associate a probability distribution with the micro regressor data too. It is important to be clear as to what kind of answer is sought. The analysis of Grunfeld and Grilicher [1960] is not explicit on this point.

8.3.3 On aggregation of units

The common correlation squared is a special case of the coefficient of determination.

Sociologists sometimes study the following kind of partitioned aggregation by averaging of units of analysis. The micro units are individuals. The macro units are disjoint groups of the individuals. The variables may, but need not, be dummy variables that allot the value 1 to an individual who has a certain property, and 0 to one who has not. If so, the macro data are percentages.

A correlation between two variables computed from the micro data is called an individual correlation. A similar correlation computed from the macro data is called an ecological correlation.

There seems to have been some tendency earlier to use the ecological correlation as an approximation to the individual correlation. On the other hand, the ecological correlation has often been found to increase steadily as the individuals are grouped into fewer and larger groups. A striking non-sociological example of this tendency is given in Yule and Kendall [1950], sections 13.2-13.7, where the "modifiable units" of analysis are geographical regions. A sociological example is Slatin [1969].

Is there a simple relation between ecological and individual correlations? For simple correlations, the answer is given by Robinson [1950]. The answer uses the same algebraical relations as 6.4 above, and the concept of correlation ratios. Robinson's answer is completely model-free.

Ecological correlation and regression is further discussed i.a. in Blalock [1964], Alker [1969], and Hannan [1971].

9 A WIDER PERSPECTIVE

9.1 Miscellaneous reflections on aggregation and models

9.1.1 On aggregation versus disaggregation

Statistical data give a numerical picture of some segment of the real world. Data that give pictures of different degrees of detail are said to refer to different levels of aggregation.

A theory about the real world is represented in statistics by a model. The theory, and thus the model, refers to some more or less clearly specified level of aggregation. Data referring to the proper level of aggregation are required, if the model is to be estimated or tested. Forecasts by the model are statements about data referring to the proper level of aggregation. Sometimes, however, data referring to the proper level are not available, except perhaps at prohibitive cost. Sometimes, it is not even known for certain what is the proper level of aggregation.

Suppose that a certain model is thought to apply to data referring to a specified level L_M of aggregation. Suppose that the only data available refer to a different level L_D of aggregation. Two questions arise. First, what can the data tell about the model? How, if at all, can data on level L_D be used to estimate or test the model on level L_M ? Second, what can the model tell about the data? How, if at all, can the model on level L_M be used to forecast data on level L_D ? These general and vague questions can be made specific and precise in several ways.

If the data are too highly aggregated for the model, there are what may be called problems of aggregation. In the opposite case there are problems of disaggregation. The consistency and interpretation problems of the formal analysis of chapters 4 to 8 are problems of aggregation. On the whole, they presume that the micro level is the proper level of aggregation, while the available data are on the macro level. The consistency problem (4.2) formally treats the micro and macro levels symmetrically, but the question asked is prejudiced in favour of the micro level. The model-free (4.3) and expectational (4.4) interpretations both interpret macro statistics in micro terms, and the latter is explicitly based on the micro model or models.

In an aggregation, the micro data determine the macro data uniquely via the aggregating functions. As a rule the macro data do not, conversely, determine the micro data uniquely; there are no (exact) disaggregating functions. For this simple reason, aggregation is, loosely speaking, more amenable than disaggregation to formal analysis. Not that problems of disaggregation are always more difficult to solve than problems of aggregation. The significant difference is that it is less easy even to formulate a solvable problem of disaggregation.

To study aggregation is easier, but to study disaggregation is equally legitimate. A similar opinion in pronounced by Ijiri [1971], section 6.3.

9.1.2 On aggregation and disaggregation in regression

In many applications of regression analysis, the following applies. The data could conceivably, if perhaps not in practice, be aggregated or disaggregated in some way, or in several ways. An analogous regression analysis could be performed using the aggregated or disaggregated data. The regression on the new level of aggregation would then be interpreted much like the actual regression. There are no convincing arguments, from theory or otherwise, to decide whether the actual level L_D of aggregation is the proper one. The model implied by the theory studied may in fact refer to a different level L_M of aggregation.

Depending on the type of aggregation and on the parameters of the model, the suspected aggregation or disaggregation may or may not admit consistency. If it is not known to be actually consistent, the standard conclusions from the actual regression analysis are not well-founded. They may well be misleading.

In particular, many econometric (and other) data refer to a period of time. There are e.g. monthly, quarterly, and annual data. At least if the theory studied contains time lags, this property of the data is definitely disquieting. Aggregation and disaggregation in the dimension of time are likely to be fertile sources of inconsistency. See for example Nerlove [1959], Mundlak [1961], and Moriguchi [1970]. Cf. also Brewer [1973]. An empirical study of aggregation over time without lags is found in Dean [1941].

There is a temptation to become over-apprehensive, and to think that "perhaps most" applications of regression analysis involve hidden aggregations or disaggregations. The border-line between wise caution and absurd obsession appears to be remarkably elusive here.

9.1.3 On models versus the real world

Different scientific theories introduce different ways of looking at the real world. They formulate different concepts, to be used as tools to grasp the real world. A good theory does not only answer many questions that were not answered before. Its real merit may well be that the questions were not even asked before. Cf. Toulmin [1953], chapter 2, or Hempel [1966], section 2.3.

A simple example from economics is the dynamic theory of agricultural supply formulated by Nerlove [1958a,c]. Random disturbances apart, this theory implies a model equation

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} \gamma^i x_{t-i} \quad ,$$

where it is assumed that $|\gamma| < 1$. This model gives a definite meaning to the terms "short run" and "long run". The short run influence of x on y is measured by the parameter β, the long run influence by $(1 - \gamma)^{-1}\beta$. The model makes it meaningful, even if trivial, to ask about the path followed by y on its way to the long run (equilibrium) level.

A generalization is the final form of interdependent econometric equations formulated by Theil and Boot [1962], and anticipated by Goldberger [1959]. Random disturbances apart, the final form is the vector equation

$$y_t = \alpha + \sum_{i=0}^{\infty} \Gamma^i B x_{t-i}$$

where Γ is a square matrix whose eigenvalues γ_h are all assumed to satisfy $|\gamma_h| < 1$. The short run influence of the j'th element of x on the i'th element of y is measured by element (i,j) of the matrix B of impact multipliers. The long run influence is measured by the corresponding element of the matrix $(I - \Gamma)^{-1} B$ of total multipliers. The path followed by y_i on its way to the long run level is of great interest i.a. for economic policy.

Two theories in the same area may well bring into focus different levels of aggregation. The theory concerned with the higher level of aggregation is not necessarily less correct simply because it is less detailed. Theories, and hence models, referring to different levels of aggregation are as legitimate as maps giving different degrees of detail. Cf. Toulmin [1953], chapter 4.

One reason why economists study aggregation is that they would like to establich a bridge connecting macroeconomics to microeconomics. In the opinion of Peston [1959], the economic profession had better concentrate on the development of macroeconomics as such. Whether a macroeconomic theory is acceptable and fruitful should be judged, theoretically and empirically, within a macroeconomic frame of reference. Its possible relation to microeconomic theories is a separate question.

The concepts and models used in science are not necessarily intuitively appealing. The inflow of water into a dam, and the number of bacterio-phage attached to a bacterium, are phenomena naturally represented, the former by a continuous variable, the latter by a discrete variable. Yet Gani [1971] discusses two successful statistical models representing, respectively, the former phenomenon by a discrete variable and the latter one by a continuous variable.

The concept of a stochastic variable is of central importance in the linear model and in other statistical models. It is based on the concept of a random event. When a statistical model is applied in practice, some observed data are regarded as realizations of stochastic variables. When is this permitted, i.e. what events can be regarded as random? Statistical text-books are often remarkably reluctant to discuss this fundamental question. An exception is H. Cramér [1945], chapter 13. Cramér writes i.a. (section 13.2): "It does not seem possible to give a precise definition of what is meant by the word 'random'. The sense of the word is best conveyed by some examples."

Many text-book examples of stochastic variables refer to gambling, heredity, or radio-active decay, and identify an observed datum directly with a stochastic variable. Similar examples referring to economics are rare, but a few are given by J.S. Cramer [1969], chapter 2. However, the typical econometric stochastic variable is a disturbance term covering the lack of exact fit of a linear model equation. Such random disturbances are sometimes said to represent the joint influence of all factors not explicitly included in the model.

Models and theoretical concepts are tools. There is more than one legitimate way to use them.

9.1.4 On parameters and stochastic model components

A model expresses the form of a regularity. Suppose that the quantity x of a fodder additive has been found to increase linearly, on the average, the quantity y of wool produced by sheep. The deterministic equation $y = \alpha + \beta x$ is a simple model of this regularity. The simple model implies i.a. that if two sheep are given identical quantities of the fodder additive, both will produce identical quantities of wool. Suppose experience shows this not to be the case. The deterministic model is then refuted by the data. Still, it may remain valid as an approximation.

In order to avoid vague claims of "approximate validity", the model can be stochasticized so as to become the incomplete linear model $y = \alpha + \beta x + \varepsilon$, $E(\varepsilon) = 0$. This model is not immediately refuted by almost any data. Still, it expresses a regularity, for data sufficiently at variance with it can be taken to refute it. In order to formalize the confrontation of the model with data, additional assumptions on the disturbance term ε are required.

The parameters of a statistical model express the interesting regularity studied. The parameters are usually assumed to remain valid also outside any given set of data. How far their validy extends is rarely stated with precision.

The stochastic disturbance terms cover the — presumably uninteresting — deviations from the interesting regularity. The stochastic disturbances are usually assumed to be literally unpredictable. Whether they are unpredictable in principle or only in practice is rarely discussed. The formal model as such does not distinguish between these two cases.

Consider again the class of simple aggregations of aspects $\{G, H, \Phi\}$ introduced in 8.1.3 above. Two unsuccessful attempts to derive a linear model Ψ in the macro variables z and u will be mentioned.

The first attempt is suggested by Klein [1953], section 5.3. Summation of the micro models for the k aspects establishes the relation

$$\Psi_S : u_j = \delta_S^{\Delta} + \delta_S^{O} z_j + \eta_{Sj} \quad ,$$

where the parameters and disturbances are

$$\delta_S^{\Delta} = \sum_{h=1}^{k} \beta_h^{\Delta} ,$$

$$\delta_S^{O} = \sum_{h=1}^{k} x_{hj} \beta_h^{O} \Big/ \sum_{h=1}^{k} x_{hj} ,$$

$$\eta_{Sj} = \sum_{h=1}^{k} \epsilon_{hj} .$$

The macro disturbance η_S is as unpredictable as the micro disturbances. The macro intercept δ_S^{Δ} is as constant as the micro intercepts. But the macro slope parameter δ_S^{O} takes a different value for each unit of analysis, and should be written δ_{Sj}^{O}.

The second attempt is that due to Theil [1954], which establishes the relation Ψ_T described in 8.1.3 and discussed i.a. in 8.1.5. The macro disturbance η_T is predictable in terms of the micro regressor data and parameters, in the sense that $E_{\Phi}(\eta_T)$ is a function of those quantities. The macro parameters δ_T^{Δ} and δ_T^{O} are valid for each unit of analysis within a given set of units. But as shown by examples 8.1.4.1 and 8.1.4.2, they take on different values for different sets of units of analysis.

In a sense, δ_S^{O} and δ_T^{O} both fail to be acceptable macro slope parameters for the same reason. There are too many micro quantities that affect them in a deterministic fashion. A simplifying regularity is called for.

9.1.5 On stochasticizing micro quantities

Under the assumptions maintained in chapters 4 to 8, an aggregation involves two sets of non-stochastic micro quantities. The first set consists of the elements of the total micro parameter vector β (4.1.2). The second set consists of the elements of the independent micro data vector x (4.2.4). The assumptions can be modified, and either β or x, or both, made stochastic.

Unweighted simple aggregation of aspects with stochastic micro parameters β is discussed by Zellner [1969] and by Theil [1971], section 11.5. Both authors assume that the distribution of β is such that $E(\beta)$ satisfies the consistency condition of P.7.2.2. Thus in the notation of 7.2.2, for $i = 1, \ldots, q$, the micro parameter subvectors β_{p+i}^{*} satisfy

$$E_{\Phi}(\beta^{*}_{p+i}) = j_k \beta^{S}_{p+i}$$

for some scalars β^{S}_{p+i}. Consequently, the expectational consistency

$$E_{H\Phi}(u) = E_{\Psi G}(u)$$

is easily attained by choosing $\delta^{S}_{p+i} = \beta^{S}_{p+i}$ for $i = 1, \ldots, q$.

If the stochastic character of β is motivated by random sampling of aspects from some population of aspects, then not only β but also x becomes stochastic. As pointed out by Theil [1971], section 11.5, this complicates the analysis.

The two authors cited assume that β is stochastic while x is not. The opposite assumptions are at least as natural. Like other parameters, the micro parameters β serve to express an interesting regularity, which is thought of as more permanent than the actual data. The observed independent micro data x record observations of the real world, and could have taken different values; cf. 2.3.3. The habit to regard x as non-stochastic is motivated by the mathematical theory of the linear model rather than by the applications of regression analysis in for example econometrics; cf. 2.3.4.

To assume that the micro parameters are non-stochastic, while the observed micro regressor data are stochastic, appears quite justifiable.

9.2 A sometimes possible way out

9.2.1 A set of normal variables

The following two properties of the multivariate normal distribution will be referred to without explicit reference.

(1) Let t be a stochastic n-vector, whose distribution is multivariate normal with mean vector $E(t) = \mu$ and non-singular covariance matrix $V(t) = S$. Let H be a non-stochastic $m \times n$ matrix of rank m, and k a non-stochastic m-vector. Then $u = k + Ht$ is a stochastic m-vector, whose distribution is multivariate normal with mean vector $E(u) = k + H\mu$ and non-singular covariance matrix $V(u) = HSH'$.

(2) Let $t = (t_1' \mid t_2')'$ be a partitioned stochastic vector, whose distribution is multivariate normal with mean vector and non-singular covariance matrix as follows.

$$E \begin{bmatrix} t_1 \\ \hline t_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix} ,$$

$$V \begin{bmatrix} t_1 \\ \hline t_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ \hline S_{21} & S_{22} \end{bmatrix} .$$

Then the conditional expectation of t_2 for a given t_1 is the following linear function of t_1; for later purposes the equation is transposed.

$$E(t_2' \mid t_1') = a_{21}' + t_1' B_{21} , \quad \text{where} \quad B_{21} = S_{11}^{-1} S_{12} \quad \text{and}$$

$$a_{21}' = \mu_2' - \mu_1' B_{21} .$$

Properties (1) and (2) are demonstrated in Graybill [1961], chapter 3, and in many other text-books.

For $h = 1, \ldots, k \geq 2$ let x_h be a stochastic vector of order $q \geq 1$, and ε_h a stochastic scalar. Further, let α_h be a non-stochastic scalar, and β_h a non-stochastic q-vector. Define

$$y_h = \alpha_h + x_h' \beta_h + \varepsilon_h .$$

Then y_h is a stochastic scalar.

Define the kq-vector x as follows.

$$x = (x_1' \ \ldots \ x_h' \ \ldots \ x_k')' .$$

Define the k-vectors y, ε and α analogously. Define the kq × k block-diagonal matrix B as follows. (B can be read "capital beta".)

$$B = \begin{bmatrix} \beta_1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \beta_h & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & \beta_k \end{bmatrix}$$

In this more compact notation,

$$y' = \alpha' + x'B + \varepsilon' \ .$$

This is a row vector of k equations, one for each y_h.

Let G be the kq × q matrix defined by

$$G = [\, I_q \quad I_q \quad \cdots \quad I_q \,]' \ .$$

Let i be the k-vector of unit elements. Define

$$z' = x'G \ ,$$
$$u = y'i \ ,$$
$$\beta = Bi \ .$$

Then z is a stochastic q-vector, u a stochastic scalar, and β a non-stochastic kq-vector with subvectors β_h, h = 1, ..., k.

Assume that the mean vector and covariance matrix of the (kq + k)-vector $(x' \mid \varepsilon')'$ are as follows.

$$E \begin{bmatrix} x \\ \hline \varepsilon \end{bmatrix} = \begin{bmatrix} \mu \\ \hline 0 \end{bmatrix}$$

$$V \begin{bmatrix} x \\ \hline \varepsilon \end{bmatrix} = \begin{bmatrix} S & 0 \\ \hline 0 & Q \end{bmatrix} \ .$$

Assume that S and Q are non-singular matrices.

It follows that the (kq+k+q+1)-vector $(x' \mid y' \mid z' \mid u)'$ has the following mean vector and covariance matrix.

$$E\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha + B'\mu \\ G'\mu \\ i'\alpha + \beta'\mu \end{bmatrix}$$

$$V\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix} = \begin{bmatrix} S & SB & SG & S\beta \\ B'S & B'SB + Q & B'SG & B'S\beta + Qi \\ G'S & G'SB & G'SG & G'S\beta \\ \beta'S & \beta'SB + i'Q & \beta'SG & \beta'S\beta + i'Qi \end{bmatrix}$$

Because of $z = G'x$ and $u = i'y$, the covariance matrix is singular.

Assume that the distribution of $(x' \mid \varepsilon')'$ is multivariate normal. Let v be any subvector of $t = (x' \mid y' \mid z' \mid u)'$ such that no element of v is defined to be an exact linear function of the other elements of v. Then the distribution of v is multivariate normal, and its mean vector and covariance matrix are the corresponding subvector of $E(t)$ and submatrix of $V(t)$.

It foolows that several conditional expectations are linear functions of the subvector held constant. The following expressions will be useful later.

First, from the distribution of $(x' \mid y')$,

$$(\varphi) \quad E(y' \mid x') = \alpha' + x'B \ .$$

This is a row vector of k equations, one for each $E(y_h \mid x')$. Because for every h

$$\left.\begin{array}{l} E(y_h \mid x_h') = \\ \\ E(y_h \mid x') = \end{array}\right\} = \alpha_h + x_h'\beta_h \ ,$$

the k elements of (φ) can also be taken to be the equations for $E(y_h \mid x_h')$.

Second, from the distribution of $(z' \mid u)$,

$$(\psi) \quad E(u \mid z') = \gamma + z'\delta \ ,$$
$$\text{where } \delta = (G'SG)^{-1}G'S\beta \text{ and}$$
$$\gamma = i'\alpha + \mu'\left[I - G(G'SG)^{-1}G'S\right]\beta \ .$$

The q-vector δ and the scalar γ are non-stochastic.

Third, from the distributions of $(z' \mid x_h')$, $h = 1, \ldots, k$,

$$(g^{-1})\quad E(x' \mid z') = \theta' + z'\Pi \ ,$$

where $\Pi = (G'SG)^{-1}G'S$ and

$$\theta' = \mu'[I - G(G'SG)^{-1}G'S] \ .$$

Equation (g^{-1}) is a row vector of k row subvectors of equations, one for each $E(x_h' \mid z')$. The $q \times kq$ matrix Π and the kq-vector θ are non-stochastic.

9.2.2 A counter-example

Consider an unweighted simple aggregation of k aspects indexed $h = 1, \ldots, k$. The intercept regressor is the only regressor common to the aspects. There are q observed micro regressors specific to each aspect. There are $n \geq q + 1$ units of analysis.

The micro data form the n-rowed matrix

$$[\,Y \mid j \mid X\,] = [\,Y \mid j \mid X_1^S \mid \ \ldots\ \mid X_h^S \mid \ \ldots\ \mid X_k^S\,] \ .$$

The micro regressand k-column matrix is as in figure 3.2.2. The designed micro regressor vector j consists of n unit elements. Each observed micro regressor q-column submatrix X_h^S specific to an aspect, $h = 1, \ldots, k$, is as in Figure 3.2.1. Note that the order in X of the kq observed micro regressors specific to the aspects, is different from that in the total micro regressor matrix X^* of 3.2.2 and 7.2.1.

The macro data form the n-rowed matrix

$$[\,u \mid j \mid Z\,] \ .$$

There are q observed macro regressors Z. The regressor and regressand aggregating functions for observed data are as follows

$$G: \quad Z = XG \ ,$$

$$H: \quad u = Yi \ ,$$

where G and i are as defined in 9.2.1.

Assume that the observed micro data are generated as follows. Each row of $[\,X \mid Y\,]$ is a randomly drawn vector from the multivariate normal distribution of $(x' \mid y')$ described in 9.2.1. The n rows are drawn independently of each other.

It follows that the aggregation has all the stochastic properties specified in the formal analysis of chapters 4 to 8, with two modifications. First, the micro and macro relations consist of regression models (2.3.4) rather than linear models. Second, the auxiliary regressions (4.3.3) are not model-free, but correspond to auxiliary regression models.

The micro relation consists of k incomplete regression models, one for each aspect. They can be written jointly

$$\Phi: \quad E(Y \mid X) = j\alpha' + XB \ ,$$

where α' and B are as in (φ) of 9.2.1.

The macro relation consists of the incomplete regression model

$$\Psi: \quad E(u \mid Z) = j\gamma + Z\delta \ ,$$

where γ and δ are as in (ψ) of 9.2.1.

The kq incomplete auxiliary regression models can be written jointly

$$G^{-1}: \quad E(X \mid Z) = j\theta' + Z\Pi \ ,$$

where θ and Π are as in (g^{-1}) of 9.2.1.

Expectation according to the micro relation is expectation conditional upon given micro regressor data. Expectation according to the macro relation is expectation conditional upon given macro regressor data. In symbols,

$$E_{\Phi}(\cdot) = E(\cdot \mid X) \ ,$$

$$E_{\Psi}(\cdot) = E(\cdot \mid Z) \ .$$

The columns of the macro regressor matrix $[\, j \,\vdots\, Z \,]$ are almost certainly linearly independent. The complementary event occurs with probability zero.

9.2.3 Analysis of the consistency problem

Substitution of Φ into H and of G into Ψ produces the semi-aggregated micro and semi-disaggregated macro relations

$$H\Phi: \quad E(Yi \mid X) = j\alpha' i + X\beta \ ,$$

$$\Psi G: \quad E(u \mid XG) = j\gamma + XG\delta \ ,$$

where β is as defined in 9.2.1.

Both $H\Phi$ and ΨG can be used to forecast the macro regressand vector u from the observed micro regressor matrix X. As in P.4.2.6 the two forecasts are identical linear functions of X if and only if

$$\begin{bmatrix} j\alpha' i \\ \hline \beta \end{bmatrix} = \begin{bmatrix} j\gamma \\ \hline G\delta \end{bmatrix} .$$

This vector equation holds if and only if, as in P.7.2.2, $\beta = G\xi$ for some q-vector ξ. For if $\beta = G\delta$, then $\beta = G\xi$ for $\xi = \delta$. And conversely, the definitions of δ and γ imply that if $\beta = G\xi$, then $G\delta = \beta$ and $\gamma = \alpha' i$.

Mechanical application of the consistency problem of 4.2 would lead to the following conclusion: "Unless $\beta = G\xi$ for some ξ, the micro and macro relations Φ and Ψ are incompatible and cannot both be simultaneously valid." But this conclusion is wrong. In the counter-example, Φ and Ψ are both valid irrespective of β.

Thus when the observed micro regressor data X are stochasticized, the consistency problem and its solution as formulated in 4.2 somehow become inapplicable. Why this happens will now be explained.

When X is non-stochastic as in 4.2, $H\Phi$ and ΨG make statements of the following form.

$$H\Phi: \quad E(u) = f_1(X) ,$$
$$\Psi G: \quad E(u) = f_2(X) .$$

If $f_1(X) \neq f_2(X)$ for some admissible X, $H\Phi$ and ΨG contradict each other. Since G and H are not questioned, it can be concluded that Φ and Ψ are incompatible.

When X is stochastic as in 9.2, $H\Phi$ and ΨG make statements of the following form.

$$H\Phi: \quad E(u \mid X) = f_1(X) ,$$
$$\Psi G: \quad E(u \mid XG) = f_2(X) .$$

Because the events "$X = X_o$" and "$XG = X_o G$" are defined by different regions in the sample space of X, the two statements are not comparable. That $f_1(X) \neq f_2(X)$ does not mean that $H\Phi$ and ΨG contradict each other. No conclusion can be drawn as to the compatibility of Φ and Ψ.

9.2.4 Analysis of the expectational interpretation

Let X_o and Z_o be such that $Z_o = X_o G$. Because the events "$X = X_o$" and "$X = X_o$ and $Z = Z_o$" are by definition identical,

$$E(\cdot \mid X = X_o \text{ and } Z = Z_o) = E(\cdot \mid X = X_o) \ .$$

Consequently, for any X and Z such that $Z = XG$,

$$E(\cdot \mid Z) = E[E(\cdot \mid X,Z) \mid Z] = E[E(\cdot \mid X) \mid Z] \ .$$

This rule will be used below.

Consider the model-free GLS(W) macro and auxiliary regressions, where W is an arbitrary positive-defnite $n \times n$ matrix. Let the fitted macro and auxiliary regressions be denoted

$$u = j d^{\Delta} + Z d^{O} \ ,$$

$$[j \mid X] = [j \mid Z] \begin{bmatrix} 1 & c^{\Delta'} \\ 0 & C^{O} \end{bmatrix} \ .$$

By the auxiliary regression models,

$$E\left\{ \begin{bmatrix} c^{\Delta'} \\ C^{O} \end{bmatrix} \middle| Z \right\} = \begin{bmatrix} \theta' \\ \Pi \end{bmatrix} \ ,$$

for $c^{\Delta'}$ and C^{O} are GLS estimates of θ' and Π.

Below, X and Z denote the actual realized observed micro and macro regressor matrices.

Mechanical application of the expectational interpretation of 4.4 leads to the interpretation

$$E\left\{ \begin{bmatrix} d^{\Delta} \\ d^{O} \end{bmatrix} \middle| X \right\} = \begin{bmatrix} 1 & c^{\Delta'} \\ 0 & C^{O} \end{bmatrix} \begin{bmatrix} \alpha' i \\ \beta \end{bmatrix} \ .$$

in accordance with P.4.4.3. If these implied micro parameter functions are regarded as macro parameters, δ_T^{Δ} and δ_T^{O}, the comments of 8.1 apply.

When the observed micro regressor data X are stochasticized, the following modified expectational interpretation becomes possible, and more natural than the original one.

$$E\left\{\begin{bmatrix} d^{\Delta} \\ \hline d^{O} \end{bmatrix} \,\middle|\, Z\right\} = E\left[E\left\{\begin{bmatrix} d^{\Delta} \\ \hline d^{O} \end{bmatrix} \,\middle|\, X\right\} \,\middle|\, Z\right] =$$

$$= \left[\begin{array}{c|c} 1 & \theta' \\ \hline 0 & \Pi \end{array}\right] \begin{bmatrix} \alpha' i \\ \hline \beta \end{bmatrix} .$$

These implied micro parameter functions can be seen to be data-free in the sense of 8.1.1. They can therefore be regarded as parameters.

By the definitions of θ, Π, γ and δ,

$$\left[\begin{array}{c|c} 1 & \theta' \\ \hline 0 & \Pi \end{array}\right] \begin{bmatrix} \alpha' i \\ \hline \beta \end{bmatrix} = \begin{bmatrix} \gamma \\ \hline \delta \end{bmatrix} .$$

Thus, the modified expectational interpretation derives the (macro) parameters γ and δ from the semi-aggregated micro model $H\Phi$ and the auxiliary regression models G^{-1}.

9.2.5 A derived macro relation

Let t be a linear function of X not involving Z. Then

$$E(t \mid Z) = E[t \mid E(X \mid Z)] .$$

This rule will be used below.

Consider the following chain of five equations, where X is understood to mean "X such that XG = Z".

$$\begin{aligned} E(u \mid Z) = \quad & E(Yi \mid Z) = \\ = \quad & E[E(Yi \mid X) \mid Z] = \\ = \quad & E(j\alpha' i + X\beta \mid Z) = \\ = \quad & j\alpha' i + (j\theta' + Z\Pi)\beta = \\ = \quad & j\gamma + Z\delta . \end{aligned}$$

Consider also the following chain of five equations.

$$
\begin{aligned}
E(u \mid Z) = \quad & E(Yi \mid Z) = \\
= \quad & E[\, Yi \mid E(X \mid Z)\,] = \\
= \quad & j\alpha' i + [\, E(X \mid Z)\,]\beta = \\
= \quad & j\alpha' i + (j\theta' + Z\Pi)\beta = \\
= \quad & j\gamma + Z\delta \ .
\end{aligned}
$$

The two chains finally agree.

In either chain, link no. 1 follows from H. Links no. 2 follow from the rules for expectations stated in 9.2.4 and 9.2.5, respectively. In either chain, link no. 3 follows from Φ, link no. 4 from G^{-1}, and link no. 5 from the definitions of θ, Π, γ and δ.

The equation twice established can be labelled

$$H\Phi G^{-1}: \qquad E(u \mid Z) = j\gamma + Z\delta .$$

Thus under the assumptions of 9.2, the macro relation Ψ can be derived from the regressand aggregating function H, the micro relation Φ, and the stochastic regressor disaggregating functions G^{-1}. In symbols,

$$\Psi = H\Phi G^{-1} .$$

Cf. $\Psi_T = H\Phi A$ of 8.1.3.

The counter-example cannot be generalized to micro regressor data with an arbitrary probability distribution, nor to any type of aggregation of regression data. Nevertheless, it sheds new light on the fundamental questions indicated in 1.3.3 and 8.1.2 above.

REFERENCES

Ackley, G. (1961). Macroeconomic Theory. Macmillan, New York.

Alker, H. (1969). A Typology of Ecological Fallacies. In Dogan and Rokkan (eds.): Quantitative Ecological Analysis in the Social Sciences. M.I.T. Press, Cambridge Mass.

Allen, R.G.D. (1956). Mathematical Economics. Macmillan, London.

Alstadheim, H. (1968). Konsistent aggregering i økonomisk teori. En oversikt over problemstillinger. Research report from Sosialøkonomisk institutt, Universitetet i Oslo.

Barten, A.P. (1962). Note on Unbiased Estimation of the Squared Multiple Correlation Coefficient. Statistica Neerlandica, 16, 151-163.

Bentzel, R. (1956). Om aggregation av produktionsfunktioner. In 25 Economic Essays in Honour of Erik Lindahl. Ekonomisk Tidskrift, Stockholm.

Blalock, H. (1964). Causal Inferences in Nonexperimental Research. University of North Carolina Press, Chapel Hill N.C.

Boot, J.C.G., and de Wit, G.M. (1960). Investment Demand: An Empirical Contribution to the Aggregation Problem. International Economic Review, 1, 3-30.

Box, G.E.P. (1966). Use and Abuse of Regression. Technometrics, 8, 625-629.

Brewer, K.R.W. (1973). Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX Models. Journal of Econometrics, 1, 133-154.

Buse, A. (1973). Goodness of Fit in Generalized Least Squares Estimation. The American Statistician, 27, 106-108.

Christ, C. (1966). Econometric Models and Methods. Wiley, New York.

Cormack, R.M. (1971). A Review of Classification. Journal of the Royal Statistical Society, Series A, 134, 321-353.

Cramér, H. (1945). Mathematical Methods of Statistics. Almqvist och Wiksell, Uppsala.

Cramer, J.S. (1964). Efficient Grouping, Regression and Correlation in Engel Curve Analysis. Journal of the American Statistical Association, 59, 233-250.

Cramer, J.S. (1969). Empirical Econometrics. North-Holland, Amsterdam.

Dean, J. (1941). Statistical Cost Functions of a Hosiery Mill. The Journal of Business of the University of Chicago. Supplement: Studies in Business Administration, 11, No. 4.

Fisher, F. (1969). The Existence of Aggregate Production Functions. Econometrica, 37, 553-577.

Fisher, I. (1922). The Making of Index Numbers. Houghton Mifflin, Boston Mass.

Fisher, W. (1969). Clustering and Aggregation in Economics. Johns Hopkins Press, Baltimore Md.

Gani, J. (1971). Models in Applied Probability. Journal of the Royal Statistical Society, Series A, 134, 26-38

Goldberger, A. (1959). Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model. North-Holland, Amsterdam

Goldberger, A. (1961). Stepwise Least Squares: Residual Analysis and Specification Error. Journal of the American Statistical Association, 56, 998-1000.

Goldberger, A. (1964). Econometric Theory. Wiley, New York.

Goldberger, A. (1968). Topics in Regression Analysis. Macmillan, New York.

Goldman, S.M., and Uzawa, H. (1964). A Note on Separability in Demand Analysis. Econometrica, 32, 387-398.

Graybill, F. (1961). An Introduction to Linear Statistical Models, Volume I. McGraw-Hill, New York.

Graybill, F. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth, Belmont Calif.

Green, J. (1964). Aggregation in Economic Analysis. An Introductory Survey. Princeton University Press, Princeton N.J.

Griliches, Z. (1957). Specification Bias in Estimates of Production Functions. Journal of Farm Economics, 39, 8-20.

Grunfeld, Y., and Griliches, Z. (1960). Is Aggregation Necessarily Bad? Review of Economics and Statistics, 42, 1-13.

Gupta, K.L. (1969). Aggregation in Economics. A Theoretical and Empirical Study. Rotterdam University Press, Rotterdam.

Gupta, K.L. (1971). Aggregation Bias in Linear Economic Models. International Economic Review, 12, 293-305.

Haitovsky, Y. (1966). Unbiased Multiple Regression Coefficients Estimated from One-Way-Classification Tables when the Cross Classifications Are Unknown. Journal of the American Statistical Association, 61, 720-728.

Hannan, M. (1971). Aggregation and Disaggregation in Sociology. D.C. Heath, Lexington Mass.

Hannan, M. (1972). Approaches to the Aggregation Problem. Research report from the Laboratory for social research, Stanford University.

Hatanaka, M. (1952). Note on Consolidation Within a Leontief System. Econometrica, 20, 301-303.

Hempel, C. (1966). Philosophy of Natural Science. Prentice-Hall, Englewood Cliffs N.J.

Ijiri, Y. (1968). The Linear Aggregation Coefficient as the Dual of the Linear Correlation Coefficient, Econometrica, 36, 252-259.

Ijiri, Y. (1971). Fundamental Queries in Aggregation Theory. Journal of the American Statistical Association, 66, 766-782.

Johnston, J. (1963). Econometric Methods. McGraw-Hill, New York.

Kendall, M. (1948). The Advanced Theory of Statistics, Volume II. Griffin, London.

Kimbell, L. (1970). The Relations Among Estimates Using Time Series, Cross Sections, Aggregate and Disaggregate Data. Abstract of paper read at the American Econometric Meeting, New York 1969. Econometrica, 38, No. 4, 153-154.

Klein, L. (1946a). Macroeconomics and the Theory of Rational Behavior. Econometrica, 14, 93-108.

Klein, L. (1946b). Remarks on the Theory of Aggregation. Econometrica, 14, 303-312.

Klein, L. (1953). A Textbook of Econometrics. Row, Peterson, Evanston Ill.

Klevmarken, A. (1972). Statistical Methods for the Analysis of Earnings Data. Almqvist och Wiksell, Stockholm.

Kloek, T. (1961). Note on Convenient Matrix Notations in Multivariate Statistical Analysis and in the Theory of Linear Aggregation. International Economic Review, 2, 351-360.

Koerts, J., and Abrahamse, A.P.J. (1969). On the Theory and Application of the General Linear Model. Rotterdam University Press, Rotterdam.

Konyus, A. (1939). The Problem of the True Index of the Cost of Living. Econometrica, 7, 10-29. Original in Russian in 1924.

Koopmans, T. (1953). Identification Problems in Economic Model Construction. In Hood and Koopmans (eds.): Studies in Econometric Method. Wiley, New York.

Kuh, E. (1963). Capital Stock Growth: A Micro-Econometric Approach. North-Holland, Amsterdam.

Lancaster, K. (1966). Economic Aggregation and Additivity. In Krupp (ed.): The Structure of Economic Science. Prentice-Hall, Englewood Cliffs N.J.

Larson, H.J., and Barr, D.R. (1972). A Note on Pooling Regression Estimates. The American Statistician, 26, No. 5, 35.

Leander, E. (1963). Några synpunkter på statistisk deskription med minstakvadrat-metod. Unpublished paper from the Institute of statistics, University of Stockholm.

Leontief, W. (1947). Introduction to a Theory of the Internal Structure of Functional Relationships. Econometrica, 15, 361-373.

Lütjohann, H. (1970a). Linear aggregation in regression and in econometrics. Chapters 1, 2 and 3. Research report from the Institute of statistics, University of Stockholm.

Lütjohann, H. (1970b). The Stepwise Regression Algorithm Seen from the Statistician's Point of View. Metrika, 15, 110-125.

Lütjohann, H. (1971). Linear aggregation in regression and in econometrics. Chapters 4, 5 and 6. Research report from the Institute of statistics, University of Stockholm.

Lütjohann, H. (1972). Another Note on Linear Aggregation of Economic Relations. International Economic Review, 13, 426-428.

Malinvaud, E. (1956). L'agrégation dans les modèles économiques. Cahiers du séminaire d'économétrie, No. 4, 69-146. Paris.

Malinvaud, E. (1964). Méthodes statistiques de l'économétrie. Dunod, Paris.

May, K. (1946). The Aggregation Problem for a One-Industry Model. Econometrica, 14, 285-298.

May, K. (1947). Technological Change and Aggregation. Econometrica, 15, 51-63.

Misra, P.N. (1967). Matrix Formulation of the Theory of Linear Aggregation. Sankhyā, series B, 29, 41-46.

Misra, P.N. (1972). Pooling of Time Series and Cross Section Data. Sankhyā, series B, 34, 385-394.

Moriguchi, C. (1970). Aggregation over Time in Macroeconomic Relations. International Economic Review, 11, 427-440.

Morishima, M. (1961). A Historical Note on Professor Sono's Theory of Separability. International Economic Review, 2, 272-275.

Mundlak, Y. (1961). Aggregation over Time in Distributed Lag Models. International Economic Review, 2, 154-163.

Nataf, A. (1962). Résultats et directions de recherche dans la théorie de l'agrégation. In Nagel, Suppes and Tarski (eds.): Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress. Stanford University Press, Stanford Calif.

Nataf, A. (1964). Théorie des choix et fonctions de demande. Centre National de la Recherche Scientifique, Paris.

Nerlove, M. (1958a). Distributed Lags and Estimation of Long-Run Supply and Demand Elastisticites: Theoretical Considerations. Journal of Farm Economics, 40, 301-311.

Nerlove, M. (1958b). On the Nerlove Estimate of Supply Elasticity: A Reply. Journal of Farm Economics, 40, 723-728.

Nerlove, M. (1958c). Distributed Lags and Demand Analysis for Agricultural and Other Commodities. United States Dept. of Agriculture, Washington D.C.

Nerlove, M. (1959). On the Estimation of Long-Run Elasticities: A Reply. Journal of Farm Economics, 41, 632-640.

Peston, M.H. (1959). A View of the Aggregation Problem. Review of Economic Studies, 27, No. 1, 58-64.

Pokropp, F. (1972). A Note on the Problem of Aggregation. Review of Economic Studies, 39, 221-230.

Prais, S.J., and Aitchison, J. (1954). The Grouping of Observations in Regression Analysis. Revue de l'Institut international de statistique, 22, 1-22.

Pringle, R.M., and Rayner, A.A. (1971). Generalized Inverse Matrices with Applications to Statistics. Griffin, London.

Rajaoja, V. (1958). A Study in the Theory of Demand Functions and Price Indexes. Academic Bookstore, Helsinki.

Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.

Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Ringstad, V. (1971). Om estimering av økonomiske relasjoner fra tverrsnitts-, tidsrekke- og kombinert tverrsnitts-tidsrekke data. Statsøkonomisk tidsskrift, 85, 195-213.

Robinson, W. (1950). Ecological Correlations and the Behavior of Individuals. American Sociological Review, 15, 351-357.

Scheffé, H. (1959). The Analysis of Variance. Wiley, New York.

Schneeweiss, H. (1965). Das Aggregationsproblem. Statistische Hefte, 6, 1-26.

Searle, S.R. (1971). Linear Models. Wiley, New York.

Seber, G.A.F. (1966). The Linear Hypothesis: A General Theory. Griffin, London.

Shou Shan Pu (1946). A Note on Macroeconomics. Econometrica, 14, 299-302.

Slatin, G. (1969). Ecological Analysis of Delinquency: Aggregation Effects. American Sociological Review, 34, 894-907.

Solow, R. (1956). The Production Function and the Theory of Capital. Review of Economic Studies, 23, 101-108.

Sono, M. (1961). The Effect of Price Changes on the Demand and Supply of Separable Goods. International Economic Review, 2, 239-271. Original in Japanese in 1945.

Theil, H. (1954). Linear Aggregation of Economic Relations. North-Holland, Amsterdam.

Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. Revue de l'Institut international de statistique, 25, 41-51.

Theil, H. (1960). Best Linear Index Numbers of Prices and Quantities. Econometrica, 28, 464-480.

Theil, H. (1962). Alternative Approaches to the Aggregation Problem. In Nagel, Suppes and Tarski (eds.): Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress. Stanford University Press, Stanford Calif.

Theil, H. (1971). Principles of Econometrics. North-Holland, Amsterdam.

Theil, H., and Boot, J.C.G. (1962). The Final Form of Econometric Equation Systems. Review of the International Statistical Institute, 30, 136-152.

Thionet, P. (1960). Ébauche d'une théorie statistique des agrégats. Revue de Statistique Appliquée, 8, No. 3, 107-111.

Thionet, P. (1967). Point de vue d'un statisticien sur le problème économétrique de l'agrégation. Paper read at the European Econometric Meeting, Bonn 1967. Abstract in Econometrica, 36, No. 5, 28.

Toulmin, S. (1953). The Philosophy of Science. Hutchinson, London.

Wilks, S.S. (1962). Mathematical Statistics. Wiley, New York.

Wold, H. (1940). Efterfrågan på jordbruksprodukter och dess känslighet för pris- och inkomstförändringar. Statens offentliga utredningar 1940:16, Stockholm.

Wold, H. (1952). Demand Analysis. A Study in Econometrics. Almqvist och Wiksell, Stockholm.

de Wolff, P. (1941). Income Elasticity of Demand, a Micro-Economic and a Macro-Economic Interpretation. Economic Journal, 51, 140-145.

Wonnacott, R., and Wonnacott, T. (1970). Econometrics. Wiley, New York.

Wu, R. (1973). On Some Aspects of Linearly Aggregated Macro Models. International Economic Review, 14, 785-788.

Yule, G., and Kendall, M. (1950). An Introduction to the Theory of Statistics, 14th edition. Griffin, London.

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

Zellner, A. (1962b). Estimation of Cross-Section Relations: Analysis of a Common Specification Error. Metroeconomica, 14, 111-117.

Zellner, A. (1969). On the Aggregation Problem: A New Approach to a Troublesome Problem. In Fox et alii (eds.): Economic Models, Estimation and Risk Programming; Essays in Honor of Gerhard Tintner. Springer, Berlin.

Zellner, A., and Montmarquette, C. (1971). A Study of Some Aspects of Temporal Aggregation Problems in Econometric Analyses. Review of Economics and Statistics, 53, 335-342.

INDEX

See also the table of contents.

University of Stockholm
Institute of Statistics
S–113 85 Stockholm, Sweden

ISBN 91–7222–049–X
Sundt Offset Stockholm 74