UNITED ARAB REPUBLIC

THE INSTITUTE OF NATIONAL PLANNING

Memo No. 840

THE DEMAND - INCOME FUNCTION - A STUDY IN AGGREGATION

BY

Dr. M. M. El-Imam

May, 1968

CONTENTS

I. Models and Structures:

Econometric analysis is concerned with the empirical investigation of quantitative relationships between economic variables as envisaged by economic theory. The analysis is related to the concept of a model, and its outcome defines a certain structure. By a model we mean a given hypothetical set of (one or more) relationships which relate economic variables to each other in a well-defined manner, each of which aiming at explaining a certain economic phenomenon. This might lead to the introduction of non-economic magnitudes which are either variable or relatively constant, hence the term "parameters". In theoretical analysis one can do without an explicit formulation of the exact functional relationships, provided sufficient information is available for defining relevant properties (e.g., first derviatives). For econometric investigation this is not sufficient.

Three conditions have to be met before a theoretical model is to be considered as workable from an econometric point of view:

1. The explicit formulation of relationships; defining their form, and the parameters to be included.

2. The definition of variables in a manner which makes them capable of actual statistical observation.

3. The specification of the distributions of stochastic variables which have to be introduced in order to account for discrepancies between ideal and observed concepts (errors of observation), and between theoretical and actual economic relationships (errors of equations).

By a structure we mean a given point of parameter values which the model assumes in a particular case. Thus a model is the complete set of structures which satisfy its full specifications. This implies that the relevant parameters are not only those of the functional economic relationships; they include also the parameters of the distributions of the stochastic element. If we confine ourselves to the former group of parameters, we would be dealing with a purely economic structure, related to a given economic model. But if we include the complete set of parameters, we would be dealing with the full econometric structure corresponding to a given econometric model.(1)

(1) In many cases in the literature an econometric model is meant to satisfy the first of the two conditions only. This has led to a confusion of the term "econometric" itself.

II. Measurement of Economic Variables:

Whether for purposes of theoretical analysis or of econometric investigation, it is of fundamental importance to define precisely the dimensions of the economic variables to be included. This requires the specification of the exact types of units of each of the following aspects:

1. The economic agents: individuals, groups of individuals and organisms.

2. The economic entities: commodities and groups of commodities, defined according to their nature and uses.

3. The time unit of measurement. In this connection we have to differentiate the timeless stocks and ratios (e.g., prices) from flows. Even when we choose to consider flows as continuous functions of time, their expression should be in terms of given time rates per discrete time units.

It can be easily noticed that in each case one could differenttiate elementary units of measurement, below which any further subdivision is either impossible or operationally meaningless. Any other units should be considered as composite units which can be defined as given functions of elementary units.

Thus the elementary unit of economic agents can be considered as the smallest unit capable of formulating and implementing decisions with respect to a given economic phenomenon. For example the individual consumer is such an elementary unit, though it might be a group of natural persons forming a household. Similarly for an individual firm. If we apply this principle we can say that Government itself is an elementary agent with respect to certain types of action, although it is constituted of a number of separate administrations.

The elementary unit of economic entities is directly related to the concept of commodity. It is known that the homogeneity condition should be strictly satisfied, meaning by that the perfect substitutability with respect to a given type of uses, or operations. However, there is a certain important difference between this elementary unit and the former one with respect to actual measurement. Thus if we consider the purchase of a given commodity by a certain consumer, we have in mind the number of units of that commodity purchased by that single consumer. This means that the relvant unit of measurement relates to the nature of the commodity. But the actual magnitude relates to an individual consumer. Further, the concept of the price of the given commodity is directly related to the commodity itself. If prices are the same for all individual consumers, then there is no need to specify them according to consumer. The concept of price remains the same although its exact magnitude might change if consumers pay different prices for the same commodity (e.g., in different markets).

The role of time is a multiple one. Thus in dynamic analysis, or in actual measurement, the values of variables change over time. This applies to stocks as well as flows. But from a dimensional point of view, one has to determine the elementary time unit with reference to the operations studied. Thus, the elementary unit of analysis should relate to the period of time necessary for the implementation of one single decision. This raises certain complications for the econometrician:

1. The elementary unit of time differs according to the type of decision.

2. It need not concide with the units used in the actual collection of data.

It is clear that actual statistical observations relate to composite time units, which are found to be convenient for collection of data, especially on economic flows. However this raises certain difficulties with respect to the measurement of variables void of the time dimension, namely stocks and ratios. The solution of these difficulties is found through familiar statistical techniques (e.g., averages). This envolves an aggregation procedure which is of a relatively simple nature. Other aggregation problems, relating to the other types of units are of a more complex nature.

III. Construction of Composite Units

Starting from elementary agents, one can construct a large variety of composite units, according to a well-defined principle. For example, we can group together all consumers within a given market, or economy, or belonging to a given region or a given social group. Again firms belonging to a given branch or sector, or falling within a certain class of size, or using a certain technique, can be grouped together.

Further, the aggregation might be carried out at a series of stages. The first stage, which is concerned with the aggregation of elementary agents into certain groups of agents, can be considered as primary aggregation. The following stages can be considered as cases of secondary aggregation. In each of these stages the aggregation is done over aggregates obtained through primary or secondary aggregates to obtain more composite aggregates.

The same applies also to economic entities. From the elementary units we can construct groups of commodities such as food. Further composite units can be constructed from these latter, e.g. total consumers' goods. While the aggregation of economic agents usually envolves simple processes of summation, the aggregation of economic entities is generally more complicated. For example, it is not possible to add up quantities of different commodities, even though their physical units are the same (e.g. tons). Index numbers are an attempt to approximate quantity and price composites. The abundant controversy around this subject is a clear indication of its complexity, as well as its ambiguity. We have to observe also that the attempt to solve the index numbers problem without reference to the exact context in which they are to be used, is responsible for many of their shortcomings.

Supposing that a convenient unit of time (usually composite) has been chosen, we still have to determine the level of measurement of the other two aspects of economic variables: Their combination yields four categories

1. Variables relating to elementary agents and elementary entities

2. Variables relating to elementary agents and composite entities.

3. Variables relating to composite agents and elementary entities.

4. Variables relating to composite agents and composite entities.

In other words we have to consider cases of no aggregation; of aggregation over entities, agents or both.

IV. Types of Analysis:

It is customary, in economic analysis, to differentiate the micro and the macro levels. Ideally the micro-analysis will be concerned with the interpretation of variables of type(1) although this might be done with reference to variables of other types. On the other hand macro-economics is essentially concerned with the interpretation of "broad aggregates", i.e., variables of type (4). We have to allocate types (2) and (3) to either branch of analysis. This has to be done according to a precise definition of the two branches, in an exhanstive and a mutually exclusive manner.

According to Bushaw and Clower:[1]

"Broadly speaking, microeconomics is concerned, first, with specifying alternative possible decisions that an individual economic unit might make and, second, with describing the process by which decisions are selected from the admissible alternatives".

This definition emphasizes the decision-making problem at the level of the elementary agent. However, it is not a universal one. Thus, Henderson and Quandt[2] would consider micronmics as: "the study of the economic actions of individuals and well-defined groups of individuals". This latter definition is more in line with the more traditional subdivision of economics into price analysis and income analysis. The term "well-defined groups of individuals" can be seen to mean those groups which are related to an elementary entity. Accordingly this definition implies that microeconomics is essentially the study of elementary entities whether in relation to elementary or composite agents. It follows that macroeconomics would be "the study of broad aggregates such as total employment and national income". Therefore, microeconomics includes categories(1) and (3), while macroeconomics deals with (2) and (4). According to the former definition, microeconomics would deal with elementary agents, categories(1) and (2), while macroeconomics deals with categories (3) and (4). Hence the study of market equilibrium for a single commodity was classified by Bushaw and Clower as macro-economic, while it was included by Henderson and Quandt in their treatment of micro-economics.

---

(1) D. W. Bushaw and R. W. Clower: Introduction to mathematical Economics, p. 102. Richard Irwin, 1957.

(2) J. M. Henderson and R.E. Quandt: Microeconomic theory - A Mathematical Approach, p.2. Mc graw Hill, 1958.

In justifying the need for a separate discipline for the study of macro-economics, Dernburg and McDugall have emphasized the agents rather than the entities aspect. They stress the fact that:(3)

"Aggregate economic behaviour does not correspond to the summation of individual activities".

What could be included in the ceteris paribus clause in micro-analysis need not remain so in macro-economic problems. In other words the lists of exogenous and endogenous variables are different. According to this point of view, the study of market equilibrium belongs to the field of macro-economics. However, the authors confined their monograph to problems arising in connection with category (4) only. The same emphasis on the aspect of studying the behaviour and activities of individuals as the criterion differentiating micro-economics, appears in many parts of the literature.(4) However, a precise definition of the two disciplines should be based on a better understanding of the role of aggregation.

V. The Need for Aggregation:

The economic theory of any model runs in micro-terms, based on decisions taken by elementary agents (5). This means that we basically possess micro-relationships, i.e., economic relationships explaining variables of type (1). In principle, there is no difficulty in formulating such relationships and discussing their implications so long as the analysis is partial. For a general study, i.e., the study of the whole economy, we can always think of a model which encompasses a whole setup of such micro-relationships. Such a model would in fact be quite cumbersome. Further, there might be some need to concentrate on certain aspects at the aggregate level.

Accordingly it is found convenient to develop economic models on the basis of broad aggregates which attempt at explaining variables of category (4). On the other hand, econometric analysis might

(3) T. F. Dernburg and D.M. Mc Dougall: Macro-Economics - The measurement, Analysis, and Control of Aggregate Economic Activity, p.2 McGraw-Hill 1960.

(4) See, e.g. R.G.D. Allen: Mathematical Economics, p. 694 Macmillan - 1956; C. Abraham and A Thomas: Microéconomie; Decisions Optimales dans L'Entreprise et dans la Nation, p.IX Dunod, 1966

(5) Allen, op. cit.

make it necessary to reconstruct the micro-relations into a macro form before subjecting them to empirical verification. For example a micro-demand-function for a commodity such as wheat might be expressed in terms of individual income, wheat prices and the prices of many related commodities. If we do not possess statistical information at this micro level, we have to replace it by a macro-function, in terms of aggregate income, wheat price and a general price index.

Both in economic analysis and in econometric investigation the question must be raised: What is the exact relationship between the micro- and the macro-relationships? The present conditions are such that the two disciplines co-exist with little to say about their correspondence. Generally speaking, macro-analysis is made with no explicit reference to the underlying micro-theory. Two approaches are open:

1. The economist might refer to certain "first principles" - sometimes to direct observations - in order to build up a certain hypothesis concerning a given macro-relationship. For example the market law of demand has been formulated by Cournot long ago before Marshall introduced his micro-analysis, based on individual decisions.

2. Alternatively, he might go through the existing stock of micro-theory, indicate the relevant factors then elaborates his hypothesis through an implicit (mental) process of aggregation. Thus from an investigation of the factors that affect individual decisions on consumption expenditure, Keynes could derive his macro propensity to consume. He also took into consideration factors that might be at work at the macro-level, such as the effect of income redistribution, to allow for differences between the micro-parameter values.

If the two hypotheses - the micro and the macro - are to be consistent, we have to make an explicit study of the degree by which the one implies the other. Theoretical analysis could do without this imperative condition, but not always with success. The dangers of this crude approach is best illustrated by the fallacy of copying a micro-relationship between the rate of wages and the size of employment into an exactly similar macro-relationship.

The econometrician who attempts to estimate economic models on the basis of micro-theory but using macro-variables faces the same dangers. As we mentioned before, in an attempt to estimate the demand function for wheat, he might build the observable macro-function in exactly the same form as the basic micro-function, substituting macro- for micro-variables. Before testing the empirical validity of the hypothesis, he should test the new function for consistency with the underlying economic theory. Otherwise he would be investigating a completely new hypothesis without sufficient theoretical investigation. In fact it was the careful attempt to build aggregative econometric models on the basis of economic theory, that led klein to formulate the problem in its proper shape.(6) A number of authors have since joined the controversy, and various criteria and approaches had been suggested.

VI. Approaches to Aggregation:

In general, two main approaches can be distinguished(7):

1. To build up a macro-model which satisfies, the conditions of micro-theory, and makes use of available statistical aggregates.

2. To derive an aggregated model from the micro-theory, then try to construct the statistical aggregates consistent with the derived theoretical ones.

To illustrate the two approaches, let us assume that a micro-function relates the dependent variable y to a number of explanatory variables z_j, such that we have for the i-th individual

$$y_i = a_i + \sum_j b_{ji} z_{ji} \qquad (1)$$

Further, available statistical data define certain aggregates, e.g. of the form:

$$Y = \sum_i s_{oi} y_i , \qquad Z_j = \sum_i s_{ji} z_{ji} \qquad (2)$$

Where the s_{ji} are given weights; if they are equal to unity, we obtain natural sums. An aggregate equation similar to (1) in shape and using variables (2) would be:

(6) L.R. Klein: "Macroeconomics and the Theory of Rational Behavior" - Econometrica, vol. 14, pp. 93-108
"Remarks on the Theory of Aggregation" - Econometrica vol. 14, pp. 303-12.

(7) L.R. Klein: Economic Fluctuations in the United States, 1921-1941 pp. 13-14 - John Wiley, 1950.

$$Y = a + \sum_j b_j Z_j \qquad (3)$$

By aggregation we mean the analysis of the conditions under which (3) can be considered as an acceptable representation of the group of equations (1). But according to the second approach, we apply the aggregation of Y to equations (1) to obtain a macro-equation consistent with them:

$$Y = \sum_i s_{oi} a_i + \sum_i \sum_j s_{oi} b_{ji} z_{ji} \qquad (4)$$

The last term need not be directly observable, since it is a function of the (unknown) micro-parameters. Even if the micro-parameters are known a priori we might still be unable to estimate the aggregates thus defined since we do not possess observations on the micro-variables. Another difficulty arises from the fact that when we attempt at building a multi-equational model, the aggregates appropriate for each equation would be different from those relating to another, since the micro-parameters are usually different. It should be also noticed that any deviation from the linear form indicated would complicate the problem.

Using malinvaud's notation(8), let:

Y_o = the set of elements y_i containing the values of the dependent micro-variable.

Z_o = the set of elements $\underline{z_i} = (z_{1i}, \ldots\ldots, z_{mi})$ containing the values of explanatory micro-variables.

Y = the set of elements Y containing the values of the dependent macro-variable

Z = the set of elements $Z = (Z_1, \ldots, Z_m)$, containing the values of macro explanatory variables.

Let f_o be the transform (function) of Z_o into Y_o, i.e., the micro-function; and f the transform of Z into Y. Further let g be the transform of the micro-explanatory variables Z_o into Z, and h the transform of Y_o into Y. Then the theory of aggregation means that we have to state the conditions under which the following scheme is consistent:

(8) E. Malinvaud: "L'Aggrégation dans les Modèles Economiques" - pp. 69-146, Cahiers du Séminaire d'Econométrie, Paris, 1956.

$$\begin{array}{ccc} Z_o & \xrightarrow{\quad f_o \quad} & Y_o \\ g \big\downarrow & & \big\downarrow h \\ Z & \xrightarrow[\quad f \quad]{} & Y \end{array} \qquad (5)$$

The first approach assumes that f_o as well as g and h are independently given. It investigates the conditions under which a given choice of f would close the scheme. The second approach assumes that bath f_o and h are given, and obtains f as a result of applying h to f_o. The functions g thus derived are then reconciliated to those which normally define Z as functions of Z_o.

It is clear from the above discussion that any solution of the problem is by necessity of an approximative nature. For practical reasons the first approach is preferred. But this need not lead to a clear-cut solution unless we apply the principles of the second approach in order to judge the consistency of the results.

VII. The Treatment of Disturbances:

It has been shown that whenever we have econometric investigation in mind, the first approach would be more convenient, since it recognizes the full implications of the given functions g and h. However, in the econometric formulation of economic relationships due allowance should be given to the specification of the distributions of random disturbances involved. We want to indicate that a complete solution of the aggregation problem should be based on a consideration of the econometric rather than the economic structures of the models envolved.(9)

Suppose that the equation for the i-th individual in period t is as follows:

$$y_{it} = a + b\, z_{it} + u_{it} \qquad (6)$$

where $i = 1, \ldots, N_t$, hence the number of elementary agents is variable over t. Equation (6) assumes similar economic structures for simplification. To complete the specification of the econometric structure, let us assume that for all i and all t:

$$E(u_{it}) = 0, \qquad V(u_{it}) = \sigma^2 \qquad (7)$$

(9) See section I above.

which means that we have similar economic structures and further, if the micro-variables are observable, then (6) will be suitable for purposes of direct statistical estimation.

Let us define the (observable) macro-variables according to the following(functions g an h):

$$Y_t = \sum_i y_{it}, \qquad Z_t = \sum_i z_{it},$$
$$\text{or,} \quad y_t = Y_t/N_t \,, \qquad z_t = Z_t/N_t \qquad (8)$$

which are the aggregates and the corresponding per-capita values. The problem of aggregation is relatively simple: We have simply to add over i and obtain the macro-function f which would envolve the given macro-variables Y and Z:

$$Y_t = a\,N_t + b\,Z_t + U_t$$
$$\text{or,} \quad y_t = a + b\,z_t + \bar{u}_t \qquad (9)$$

where,

$$U_t = \sum_i u_{it} \qquad \bar{u}_t = U_t/N_t \qquad (10)$$

The economic aggregation is rather trivial: function f is exactly similar to f_o and all macro-variables are directly observable.

But from as econometric point of view we still have to dervie the distribution of the aggregate disturbances (10). Assuming that all u_{it} are independent (with respect to i), we have:

$$E(U_t) = 0, \qquad V(U_t) = N_t\,\sigma^2$$
$$E(\bar{u}_t) = 0, \qquad V(\bar{u}_t) = \frac{1}{N_t}\,\sigma^2 \qquad (11)$$

If N_t is variable over time, the aggregate disturbances will not satisfy the condition of homoscedasticity; i.e., the constance of variance for all points of observation. In order to be able to estimate the macro-function we have to transform it into a form satisfying that condition.

Let:

$$k_t = \sqrt{N_t} \qquad (12)$$

Define the new variables:

$$\left.\begin{aligned} \eta_t &= y_t\, k_t = Y_t/k_t \\ \zeta_t &= z_t\, k_t = Z_t/k_t \\ u_t &= \bar{u}_t\, k_t = U_t/k_t \end{aligned}\right\} \quad (13)$$

Then

$$E(u_t) = 0 \qquad V(u_t) = \sigma^2 = \text{constant}$$

Using these transformations we can replace (9) by the following equation:

$$\eta_t = a\, k_t + b\, \zeta_t + u_t \qquad (14)$$

which can be subjected to familiar methods of statistical estimation.

Thus if (6) is an equation expressing the individual propensity to consume, then an aggregate equation of the same form:

$$Y_t = a' + b\, Z_t + U_t$$

would be the exact correspondant. But when compared with (9), it would be found to violate the condition $E(U_t) = 0$. It is often claimed that the use of this latter equation would overestimate the parameter b, owing to the spurious correlation between Y and Z resulting from population changes. The use of per capita values is prefered since it eleminates the effect of N_t on the economic structure. However, it does not follow that the per-capita form is suitable for statistical estimation, since the disturbance is heteroscedastic. If we insist on using per capita values for estimation, assuming that its disturbance has a constant variance, $V(\bar{u}_t) = \sigma^2$, then this implies that the micro-disturbances have a variable variance: $V(u_{it}) = \sigma^2 N_t$. On the other hand, if we use the aggregated version of (9), assuming $V(U_t) = \sigma^2$, then $V(u_{it}) = \sigma^2/N_t$.

If we decide to use per capita values - as is usually done - we have to justify the increasing variability of the individual variances It might be argued, for example, that the wider the market is, the greater will be the factors leading to deviations from the micro-function relating consumption to income. Unless this is proved to be true, equation (14) has to be used for estimation purposes.

VIII. A Specific Econometric Problem:

The present paper has been motivated by a specific problem raised in connection with the research project undertaken by the I.N.P.C. on perspective planning. In projecting the demand for wheat and other agricultural products, it was felt that it is important to account for differences in the consumption patterns existing between rural and urban areas. On the other hand, there is no direct statistical information which enable the estimation of a separate function for each region. The problem is to formulate a national demand function which takes these differences into consideration, and at the same time makes use of published statistical aggregates. We have to allow also for the fact that the analysis is based on time series, rather than cross-sectional data.

If we start from regional demand-functions then we have to deal with a problem of secondary aggregation (10). It is preferable to state the problem in a more general form starting from the micro-functions themselves. We assume that the total of N_t consumers is composed of g groups (in the present case g = 2).

Each group contains N_{it} individuals, so that:

$$N_t = \sum_{i=1}^{g} N_{it}, \qquad N_{it} = n_{it} N_t, \qquad \sum_{i}^{g} n_{it} = 1 \qquad (15)$$

The j-th individual in the i-th group acts according to a given micro-demand function:

$$y_{ijt} = a_{ij} + b_{ij} z_{ijt} + u_{ijt} \qquad (16)$$

The micro-variables denote the economic variables or any function of them; e.g., their logarithms. In the former case, we have a linear demand function, where the variables are:

y_{ijt} = the individual's consumption

z_{ijt} = the individual's income

u_{ijt} = a random residual.

(10) See section III above.

All variables relate to the j-th individual, in the i-th group, at year t. The overall function is obtained through summation over j (= 1,2,....,N_{it}) and i (= 1,.....,g). The first summation gives the group functions, and it can be considered as the process of primary aggregation. It will be shown in the following sections that the knowledge of the micro-function is important for the determination of group functions. Thus we have to investigate this process before deciding on the form of the group functions themselves. The further summation over i (= 1, 2,.....g) gives the required overall function. Since the size of population is usually variable, as in (15), it would be preferable to express all equations on a per capita basis.

Starting from the micro-variables, we can define the group per capita variables as follows:

$\bar{y}_{it} = \sum_j y_{ijt}/N_{it}$ = group per capita consumption

$z_{it} = \sum_j z_{ijt}/N_{it}$ = group per capita income

$u_{it} = \sum_j u_{ijt}/N_{it}$ = disturbances of per capita group function

all variables relating to group i at time t. For the whole economy, the overall per capita variables, at year t, are

$y_t = \sum_i \sum_j y_{ijt}/N_t = \sum_i n_{it}\, y_{it}$ = consumption

$z_t = \sum_i \sum_j z_{ijt}/N_t = \sum_i n_{it}\, z_{it}$ = income

$\bar{u}_t = \sum_i \sum_j u_{ijt}/N_t = \sum_i n_{it}\, u_{it}$ = disturbance

In some cases functions of the macro-variables are used. The application of the summation rule does not yield the above macro-variables directly. Since we are going to preserve the linear form in those functions of the variables, we have to stick to the summation rule. Therefore we have to relate the aggregates obtained through this process to those given above, since it is these latter that can be measured empirically.

If the functions of variables take the form of logarithms, their averaging will give geometric means (or rather their logarithms). It is known that the geometric mean is somewhat smaller than the corresponding arithmetic mean. We introduce the ratios between the two means to relate our aggregates to those given above. We shall denote the geometric means of z by g, and of y by h.

$$\left.\begin{aligned} \log g_{it} &= \frac{1}{N_{it}} \sum_j \log z_{ijt} \\ \log g_t &= \frac{1}{N_t} \sum_i \sum_j \log z_{ijt} = \sum_i n_{it} \log g_{it} \\ \log G_t &= \sum_i n_{it} \log z_{it} \end{aligned}\right\} \quad (17)$$

and,

$$\left.\begin{aligned} \log h_{it} &= \frac{1}{N_{it}} \sum_j \log h_{ijt} \\ \log h_t &= \frac{1}{N_t} \sum_i \sum_j \log h_{ijt} = \sum_i n_{it} \log h_{it} \\ \log H_t &= \sum_i n_{it} \log y_{it} \end{aligned}\right\} \quad (18)$$

Two types of relationships between these variables will be considered:

1. Relationships between group and overall variables:
Let:

$$z_{it} = d_{it} z_t, \qquad y_{it} = e_{it} y_t \qquad (19)$$

The averages of these ratios can be obtained by weighting them by n_{it}. Thus their arithmetic means are:

$$\sum_i n_{it} d_{it} = 1, \qquad \sum_i n_{it} e_{it} = 1 \qquad (20)$$

Their geometric means are denoted by δ_t , ϵ_t :

$$\text{Log } \delta_t = \sum_i n_{it} \log d_{it}, \qquad \log \epsilon_t = \sum_i n_{it} \log e_{it} \qquad (21)$$

The restrictions (20) on the values of these ratios show that they are functions of n_{it}. This can be shown by first relating the group per capita incomes to some constant value. For example, if z_o is the overall per capita income in the base year ($t = o$), we can define the ratio D such that:

$$z_{it} = D_{it} z_o, \qquad z_t = D_t z_o \qquad (22)$$

then,

$$d_{it} = D_{it}/D_t, \qquad D_t = \sum_i n_{it} D_{it} \qquad (23)$$

Similar relations can be obtained for the e's.

2. Relationships between arithmetic and geometric means. Using the relations given in (17) and (18), we can summarize the relations between the two sets of means as follows:

Table (1)- Ratios of Geometric to Arithmetic means

Averaging variables	Averaging over	to obtain means of	gives Arithmetic mean	gives Geometric mean	Ratio of Arithmetic to geometric
z_{ijt}	j	groups	z_{it}	g_{it}	r_{it}
z_{it}	i	(group averages)	z_t	G_t	δ_t
z_{ijt}	i,j	(whole economy)	z_t	g_t	p_t
y_{ijt}	j	groups	y_{it}	h_{it}	s_{it}
y_{it}	i	group averages	y_t	H_t	ε_t
y_{ijt}	i,j	whole economy	y_t	h_t	q_t

The ratios δ and ε are the same as those defined in (21). This can be shown as follows:

$$\log G_t = \sum_i n_{it} \log z_{it} = \sum_i n_{it} \log (d_{it} z_t)$$

$$= \log z_t + \sum_i n_{it} \log d_{it} = \log z_t + \log \delta_t$$

Similarly for H_t. Further, let us introduce:

$$\log R_t = \sum_i n_{it} \log r_{it}; \quad \log S_t = \sum_i n_{it} \log s_{it} \qquad (24)$$

Thus R and S are the geometric means of the corresponding ratios. Then,

$$p_t = \delta_t R_t, \qquad q_t = \varepsilon_t S_t \qquad (25)$$

For example:

$$\log g_t = \sum_i n_{it} \log g_{it} = \sum_i n_{it} (\log z_{it} + \log r_{it})$$
$$= \log G_t + \log R_t = \log z_t + \log \delta_t + \log R_t$$
$$= \log z_t + \log p_t$$

hence, $\log p_t = \log \delta_t + \log R_t$

We shall deal also with power functions of the variables. For example, if the micro-function envolves the squares of individual incomes, the application of the summation process leads to sums of squares which are not usually observable. Since the definition of the variance envolves such sums, we can overcome this difficulty by introducing the variances, V^2 for z and W^2 for y. It follows that:

$$\left.\begin{aligned} \frac{1}{N_{it}} \sum_j z_{ijt}^2 &= V_{it}^2 + z_{it}^2 \\ \frac{1}{N_{it}} \sum_j y_{ijt}^2 &= W_{it}^2 + y_{it}^2 \end{aligned}\right\} \qquad (26)$$

Averaging for the whole economy we obtain:

$$\frac{1}{N_t} \sum_i \sum_j z_{ijt}^2 = \sum_i n_{it} V_{it}^2 + \sum_i n_{it} z_{it}^2$$
$$= V_t^2 + \left(\sum_i n_{it} \cdot d_{it}^2\right) z_t^2 = V_t^2 + \varphi_t z_t^2 \qquad (27)$$

A similar relation can be established for y.

The above approach does not solve completely all our problems. In fact these problems are thrown into the ratios and paramters introduced. We have to set rules for estimating them, or at least to formulate working hypotheses which would enable us to approximate them. The following section gives the set of alternative assumptions to which we shall refer later.

IX. Alternative Assumptions:

The general function (16) considered before, and the process of aggregation of variables raise the need to state some assumptions with regards to:

1. The types of economic structures of the micro-functions.
2. The distribution of population among groups.
3. The distribution of income
4. The distribution of residual elements.

Let us consider each of these aspects.

By the economic structure we mean the nature of variables envolved, and the specification of their parameters. We shall deal with the various types of functions of the variables in the following sections. As regards the parameters we consider the following alternative assumptions:

Assumptions A: relating to the individual structures

(A/1): All individual structures are different, which means that all micro-parameters are different, as in (16). But the members of any group follow the same average behaviour at all points of time. Thus if a_{io} and b_{io} are group parameters in the base year, we have, for all t:

$$\frac{1}{N_{it}}\sum_j a_{ij} = a_{io}, \quad \frac{1}{N_{it}}\sum_j b_{ij} = b_{io} \tag{28}$$

(A/2): All individuals within group follow the same economic structure:

$$a_{ij} = a_i, \qquad b_{ij} = b_i \tag{29}$$

(A/3): All individuals in the whole economy follow the same economic structure:

$$a_{ij} = a \qquad b_{ij} = b \tag{30}$$

The second group of assumptions relates to the distribution of population over groups, which is, usually, variable over time. In general there will be enough information to observe it. If this is not available, we might depend on some points of observation to estimate it over time by means of certain relationships assumed:

<u>Assumptions B:</u> relating to the distribution of population among groups:

<u>(B/1):</u> Population distribution varies over time as a linear function of a variable x, which itself is a given function of time:

$$n_{it} = m_{io} + m_{i1} x_t \tag{31}$$

The parameters m are assumed to be given, and satisfy the conditions:

$$\sum_i m_{io} = 1, \qquad \sum_i m_{i1} = o \tag{32}$$

<u>(B/2):</u> Population distribution remains constant over time:

$$n_{it} = m_i, \qquad \sum_i m_i = 1 \tag{33}$$

In the case g = 2, it will be shown that these two alternatives are sufficient to cover all possibilities.

The third group of assumptions relates to the distribution of income both among groups and among individuals. The following characteristics of this distribution are to be considered:

a- The group per capita incomes, already defined by (19)-(23).
b- The variances of relative per capita incomes within group, or the coefficients of variation of absolute incomes.
c- The variances of alsolute per capita incomes within groups.
d- The nature of the above two parameters in the special case of a log-normal distribution.

<u>Assumptions (C):</u> Relating to group per capita incomes

<u>(C/1):</u> Per capita group incomes hold given, but variable, ratios to each other. We shall assume that the D_{it} defined in (22) are linear functions of some given time function, f_t (e.g., $f_t = t$).

$$D_{it} = d_{io} + d_{i1} f_t \tag{34}$$

The coefficients d can, for example, be estimated from an analysis of results of (at least two) sample surveys. Given the values of n_{it} by assumptions (B), we obtain, by (23)

$$d_{it} = D_{it}/D_t, \qquad D_t = \sum_i d_{io}\, n_{it} + \sum_i d_{i1}\, f_t\, n_{it}$$

which can be substituted in (19) to define z_{it} once z_t is known.

(C/2): Per capita group incomes hold constant relations to each other. In this case:

$$D_{it} = d_i, \qquad d_{it} = d_i/D_t \tag{35}$$

which is the same as (34) with $d_{io} = d_i$, and $d_{i1} = 0$.

Assumptions D: relating to the variance v^2 of the relative income distribution, namely, z_{ijt}/z_{it}.

(D/1): All group variances are equal, but change over time in a given way:

$$\text{Var}\,(z_{ijt}/z_{it}) = v_{it}^2 = v_t^2 \tag{36}$$

(D/2): All group variances differ from each other but remain constant over time:

$$\text{Var}\,(z_{ijt}/z_{it}) = v_{it}^2 = v_i^2 \tag{37}$$

(D/3): All group variances are both constant and equal

$$\text{Var}\,(z_{ijt}/z_{it}) = v_{it}^2 = v^2 \tag{38}$$

Assumptions E: Relating to the variance V^2 of the absolute income distribution. Given V, then v will be in fact the coefficient of variation of obsolute income. Thus:

$$\text{Var}\,(z_{ijt}) = V_{it}^2 = v_{it}^2 \cdot z_{it}^2 \tag{39}$$

It is clear that even when the coefficient v is constant the variance V^2 will be still variable.

(E/1): Group variances are equal but variable over time:

$$\text{Var}\ (z_{ijt}) = V_{it}^2 = V_t^2 \qquad (40)$$

(E/2): Group variances differ but remain constant:

$$\text{Var}\ (z_{ijt}) = V_{it}^2 = V_i^2 \qquad (41)$$

(E/3): Group variances are equal and constant:

$$\text{Var}\ (z_{ijt}) = V_{it}^2 = V^2 \qquad (42)$$

These three alternatives are formally similar to those of (D). However relation (39) gives variable values of V for all variants of (D). Conversely assumptions (E) give variable values, $\bar{v}_{it}$, for the coefficient of variation. In a sense the two assumptions are complementary.

Assumption F: Group incomes are lognormally distributed. In other words:

$$\log z_{ijt} = N(\mu_{it}, \theta_{it}^2) \qquad (43)$$

This assumption is substantiated by the findings of empirical studies(11). Moreover, its application to the distribution of incomes within groups is further justified by the following remark by Aitchison and Brown:(12)

> "The more homogeneous the group of income participants is, the more likely is the lognormal curve to yield a good description of income distribution:

(11) See, e.g., L. Klein: A Textbook of Econometrics, p. 46, Row, Peterson, 1953; also, J.Aitchison and J.A.C. Brown: The Log-Normal Distribution, Ch. 11, Cambridge, 1957.

(12) Op. cit., p. 118

According to this assumption, we have by (43) and (17):

$$\mu_{it} = \log\ g_{it} \tag{44}$$

Further, it is known[13] that:

$$\log z_{it} = \mu_{it} + \frac{1}{2}\ \theta^2_{it} \tag{45}$$

It follows that the ratio r_{it} defined in Table (1) satisfies the relationship:

$$\log r_{it} = -\ \tfrac{1}{2}\theta^2_{it} \tag{46}$$

Also, it is known[13] that the variance of z is:

$$V^2_{it} = v^2_{it}\ z^2_{it}, \qquad v^2_{it} = \exp\ (\ \theta^2_{it}) - 1 \tag{47}$$

Thus if the coefficient of variation is given, all relevant parameters can be determined, provided that the ratios d_{it} are known. The following relationships can be easily derived:

$$\left.\begin{aligned} \theta_{it} &= \sqrt{\log\ (v^2_{it} + 1)} \\ r_{it} &= 1/\sqrt{v^2_{it} + 1} \\ \mu_{it} &= \log z_t + \log d_{it} - \log \sqrt{v^2_{it} + 1} \\ \log g_t &= \sum_i n_{it}\ \mu_{it} = \mu_t \\ \log p_t &= \mu_t - \log\ z_t \\ &= \delta_t - \sum_i n_{it}\ \ \log \sqrt{v^2_{it} + 1} \end{aligned}\right\} \tag{48}$$

It is clear that the alternative assumptions (D) have their corresponding implications for θ and r.

$$v_{it} = v_t \qquad \theta_{it} = \theta_t = \sqrt{\log\ \ (v^2_t + 1)}$$

$$v_{it} = v_i \qquad \theta_{it} = \theta_i = \sqrt{\log\ (v^2_i + 1\ )}$$

$$v_{it} = v \qquad \theta_{it} = \theta = \sqrt{\log\ \ (v^2 + 1)}$$

(13) Ibid, p. 8

Example: suppose that in a case g = 2, it was found for a given point t that $D_{1t} = 2$, $D_{2t} = 1$; $n_{1t} = 1/3$, $n_{2t} = 2/3$ and $v_{1t} = 0.75$ $v_{2t} = 0.60$. Using the above formulae we can derive the following values, knowing that $z_t = 100$:

Variables	Values for i = 1	2	Variables for the whole economy
n	1/3	2/3	Data
D	2	1	
d	1.5	0.75	
v	0.75	0.60	
θ	0.668	0.555	
r	0.798	0.8575	
log r	-0.0969	-0.0668	log R = -0.0768
log d	0.4054	-0.2877	log δ = -0.0567
log z	5.0106	4.3175	log z = 4.6052
μ	4.9137	4.2507	μ = 4.4717

It follows that:

$$\log p_t = \log z_t - \mu_t = \log \delta_t + \log R_t = -0.1335$$

Hence, $p_t = 0.875$.

This means that given z_t we can estimate $\log g_t = \mu_t$, which figures in the aggregate function, by subtracting $\log p_t$ from $\log z_t$. The value of $\log p_t$ is obtained by adding $\log \delta_t$ and $\log R_t$. Notice that the values of $\log z_{it}$ are calculated on the basis of the assumption that d_{it} and z_t are given:

$$\log z_{it} = \log d_{it} + \log z_t$$

X. Effects of Aggregation on Disturbances

The last set of assumptions mentioned in the previous section relates to the distribution of disturbances, U_{ijt}. If the process of aggregation is restricted to simple summation, the aggregate disturbances will be.

group disturbances, $u_{it} = \sum_j u_{ijt}/N_{it}$

overall disturbance, $\bar{u}_t = \sum_i \sum_j u_{ijt}/N_t = \sum_i n_{it} \; u_{it}$

This means that if we specify the characteristics of the distribution of the micro-disturbances, those of the macro ones will follow. The following assumptions will be made:

$$\left.\begin{array}{lll} E\,(u_{ijt}) = 0 & & (\text{all } i,\ j,\ t) \\ Cov.\,(u_{ijt},\ u_{ikt}) = 0 & & (j \neq k) \\ Cov.\,(u_{ijt},\ u_{hkt}) = 0 & & (h \neq i) \end{array}\right\} \qquad (49)$$

A minimum specification requires the determination of variances. The most general assumption would allow all individual variances to be both different and variable. Let λ_{ijt} be given positive factors. It is assumed that:

$$\text{Var}\,(u_{ijt}) = \lambda_{ijt}\,\sigma^2 \qquad (50)$$

Defining the weighted averages:

$$\lambda_{it} = \frac{1}{N_{it}} \sum_j \lambda_{ijt}, \qquad \lambda_t = \frac{1}{N_t} \sum_i \sum_j \lambda_{ijt} = \sum_i n_{it} \lambda_{it} \qquad (51)$$

we can express the aggregate variances as follows:

$$\text{Var}\,(u_{it}) = (1/N_{it}^2) \sum_j \lambda_{ijt}\sigma^2 = (1/N_{it})\,\lambda_{it}\sigma^2 \qquad (52)$$

$$\text{Var}\,(\bar{u}_t) = \sum_i n_{it}^2\, \text{Var}\,(u_{it}) = (1/N_t)\lambda_t\,\sigma^2 \qquad (53)$$

This means that the aggregate functions obtained by means of simple summation will not have constant variances. They have to be adequately transformed before being put to statistical estimation. This is done through multiplication by a factor similar to that defined by (12), namely:

$$k_t = N_t \,/\sqrt{N_t\,\lambda_t} = N_t \,/\sqrt{\sum_i N_{it}\,\lambda_{it}} = N_t \,/\sqrt{\sum_i \sum_j \lambda_{ijt}} \qquad (54)$$

This multiplication transforms the variables into a form similar to that given in (13).

To maintain linarity of estimates, we have to determine λ_{ijt} in some given way. This leads to the introduction of some simplifying assumptions. In the present case, of consumption behaviour, we can make use of an observation frequently noticed in consumers' surveys, pointing to the fact that the variability of behaviour is an increasing function of the level of income.(14)

Assumption (1): Individual disturbances have variable and different variances, proportionate to individual incomes:

$$\lambda_{ijt} = z_{ijt} \tag{55}$$

To estimate the micro-functions, we have to divide throughout by the square-roat of income, Further,

$$\lambda_{it} = z_{it}, \qquad \lambda_t = z_t$$

The factor k_t defined above becomes,

$$k_t = \sqrt{N_t/z_t} \tag{56}$$

Assumption (2): Individual disturbances have variable and different variances, proportionate to the squares of individual incomes:

$$\lambda_{ijt} = z^2_{ijt} \tag{57}$$

Using (26) and (39), we can substitute in (54) to obtain:

$$k_t = \sqrt{N_t / \sum_i n_{it} \, z^2_{it} \, (v^2_{it} + 1)} \tag{58}$$

This can be calculated once we possess information on the v's and the d's. The relevant variants of assumptions (C) and (D) or (E) have to be used. Only when $n_{it} = m_i$, $d_{it} = d_i$, and $v_{it} = v$, can a factor of the form,

$$k_t = \sqrt{N_t} \; / \; z_t$$

be justified.

(14) See, e.g., L. Klein, et al: Constributions of Survey Methods to Economics, chs. IV and V, especially, pp. 161-168, 212, and 232 Colombia, 1954.

Assumption (3): Individual disturbances are similar within groups, but differ among groups, they vary over time:

$$\lambda_{ijt} = \lambda_{it} \qquad (59)$$

Then the factor k becomes:

$$k_t = \sqrt{N_t / \sum_i n_{it} \; \lambda_{it}}$$

as in (54). In the special case:

$$\lambda_{ijt} = z_{it}$$

we are in the same position as in assumption (1). Formula (56) applies.

Assumption (4): Individual disturbances have constant and similar variances within groups, but differ among groups:

$$\lambda_{ijt} = \lambda_i \qquad (60)$$

This can be substituted in (54) to obtain k. For example we might obtain information on λ_i from some sample survey, and assume them to remain constant. Alternatively if we assume them proportionate to d_i defined in (35), we can substitute the estimates of these latter.

Assumption (5): Group disturbances have constant and equal variances:

$$\text{Var}\ (u_{it}) = \sigma^2 \qquad (61)$$

This means that:

$$\lambda_{ijt} = \lambda_{it} = N_{it}, \qquad \lambda_t = N_t \sum_i n_{it}^2$$

The overall function has to be multiplied by

$$k_t = 1 / \sqrt{\sum_i n_{it}^2} \qquad (62)$$

Assumption (6): All individual disturbances have constant and equal variances. Then:

$$\lambda_{ijt} = \lambda_{it} = \lambda_t = 1 \qquad (63)$$

This is the simplest, but most restrictive assumption. The relevant formula for k is (12),

$$k_t = \sqrt{N_t}$$

which had been obtained in a similar case.

Assumption (7): The direct use of overall per capita value assumes that

$$\text{Var}\ (\bar{u}_t) = \sigma^2 \tag{64}$$

This implies that

$$\lambda_{ijt} = \lambda_{it} = \lambda_t = N_t \tag{65}$$

In other words

$$\text{Var}\ (u_{ijt}) = N_t \sigma^2, \qquad \text{Var}\ (u_{it}) = \sigma^2/n_{it}$$

This means that the smaller the size of a group the larger will be its variance. The variance will be a maximum for a group of size one, namely an individual. As the size of population increases, the variance of a group will remain the same so long as its relative size does not change, while individual variances increase. Such an implied hypothesis has to be justified.

XI - The linear Form; General Case:

Suppose that the micro-relationships between the economic variables y and z are linear, assuming the general form (16). As had been shown in Assumptions (A), some simplification of that function can be introduced by assuming a certain degree of similarity of micro-structures. Such a similarity might be assumed only for averages over groups, or for individuals within a group or the whole economy. We shall investigate first the group functions derived under the various alternatives of assumptions (A).

If assumption (A/1) is satisfied, the aggregation over members of one group gives:

$$y_{it} = \frac{1}{N_{it}} \sum_j a_{ij} + \frac{1}{N_{it}} \sum_j b_{ij}\, z_{ijt} + u_{it}$$

$$= a_{io} + b_{io}\, z_{it} + \frac{1}{N_{it}} \sum_j (b_{ij} - b_{io})\, z_{ijt} + u_{it}$$

where a_{io} and b_{io} are the average group parameters defined by (28). It follows that the last term but one represents the covariance of the b's and z's. Denoting by ρ_{it} the corresponding correlation coefficient, and by σ_{bi} and V_{it} the standard deviations of b and z, we have :

$$\frac{1}{N_{it}} \sum_{j} (b_{ij} - b_{io})\, z_{ijt} = \rho_{it} . \sigma_{bi} . V_{it}$$

If it could be assumed that the correlation coefficient is stable over t, so that $\rho_{it} = \rho_i$ for all t, we can write

$$b_{i1} = \rho_{it} . \sigma_{bi} \qquad (66)$$

If follows that the covariance will be proportionate to the standard deviation V of the z's.

In fact, there is evidence from consumers' sample surveys pointing to such a possibility. If we accept the relative-income hypothesis, we can think of the correlation coefficient as measuring the relationships between the b's and the relative incomes z_{ijt} / z_{it}. Substituting from (39), we can rewrite the covariance term in terms of var V. Using v and introducing assumption (D/1), we can write the <u>group functions</u> as :

$$y_{it} = a_{io} + (b_{io} + b_{i1} v_t)\, z_{it} + u_{it} \qquad (67)$$

But using V and assuming (E/1) they become

$$y_{it} = a_{io} + b_{io} z_{it} + b_{i1} V_t + u_{it} \qquad (68)$$

As mentioned before assuming v_t for all groups implies a variable V_{it} and vice versa.

Assumption (A/2) means that all $b_{ij} = b_{io} = b_i$ say, and the group functions are :

$$y_{it} = a_i + b_i z_{it} + u_{it} \qquad (69)$$

This function can also be assumed for those cases of (A/1) where other variants of (D) or (E) apply. For example, if we assume (D/2) or (D/3) we can write :

$$b_i = b_{io} + b_{i1} v_i \quad , \quad \text{or,} \quad b_i = b_{io} + b_{i1} v$$

then put $a_{io} = a_i$. On the other hand, if we assume (E/2) or (E/3), we can write:

$$a_i = a_{io} + b_{i1} V_i , \quad \text{or,} \quad a_i = a_{io} + b_{i1} V$$

and put $b_{io} = b_i$. In either case we obtain an equation similar to (69) , with the appropriate interpretation of the parameters.

Finally, assumption (A/3) of similar structures over the whole economy leads to the simple group function :

$$y_{it} = a + b z_{it} + u_{it} \tag{70}$$

Let us now discuss the implications of these four group functions with respect to the <u>overall function</u>.

The following notation will be used :

$$\left.\begin{aligned}
\alpha_0 &= \sum_i a_{io} m_{io} & \alpha_1 &= \sum_i a_{io} m_{i1} \\
\beta_{10} &= \sum_i b_{io} m_{io} d_{io} & \beta_{20} &= \sum_i b_{i1} m_{io} d_{io} \\
\beta_{11} &= \sum_i b_{io} m_{i1} d_{io} & \beta_{21} &= \sum_i b_{i1} m_{i1} d_{io} \\
\beta_{12} &= \sum_i b_{io} m_{io} d_{i1} & \beta_{22} &= \sum_i b_{i1} m_{io} d_{i1} \\
\beta_{13} &= \sum_i b_{io} m_{i1} d_{i1} & \beta_{23} &= \sum_i b_{i1} m_{i1} d_{i1} \\
\beta'_{10} &= \beta_{10}/ D, & \beta'_{20} &= \beta_{20}/ D \\
\gamma_0 &= \sum_i b_{i1} m_{io} & \gamma_1 &= \sum_i b_{i1} m_{i1}
\end{aligned}\right\} \tag{71}$$

where D is the value of D_t defined in (35) in the special case of assumptions (B/2) and (C/2), namely

$$D = \sum_i d_i \; m_i \;; \qquad D_t = \sum_i d_{io} \; n_{it} + \sum_i d_{i1} \, f_t \, n_{it} \qquad (72)$$

Case (1) :

Micro-structures different, assumptions (A/1) and (D/1). The group functions are (67). Four sub-classes can be distinguished, on the basis of alternatives of assumptions (B) and (C).

Case (1/i): Assumptions (B/1) and (C/1) :

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_{10} + \beta_{11} x_t + \beta_{12} f_t + \beta_{13} x_t f_t) \; z_t / D_t$$
$$+ (\beta_{20} + \beta_{21} x_t + \beta_{22} f_t + \beta_{23} x_t f_t) \, v_t \, z_t / D_t + \bar{u}_t \qquad (73)$$

Assuming all variables envolved are observable, we still have to estimate ten rather than two (as in the micro-function) or three (as in the group function) parameters. Further the multiplicative way in which the variables enter, raise certain statistical difficulties. These difficulties are further complicated if the expressions d_{it} and n_{it} in (22) and (31) contain stochastic elements. On the other hand, the linear transformations given by (71) put some conditions on the matrix of transformation itself. It can be easily seen that if the macro-parameters are to be linearly independent then g should be at least 4. For smaller values of g we have to solve (71) for linear restrictions on the parameters β_{1k} (k = 0,1,2,3). This will be illustrated later for the case g = 2.

Case (1/ii) : Assumptions (B/1) and (C/2):

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_{10} + \beta_{11} x_t) \; z_t / D_t$$
$$+ (\beta_{20} + \beta_{21} x_t) \, v_t \, z_t / D_t + \bar{u}_t \qquad (74)$$

Here we allow for variation in population distribution, but assume group per capita incomes to move proportionately.

Case (1/iii): Assumptions (B/2) and (C/1)

$$y_t = \alpha_o + (\beta_{10} + \beta_{12} f_t)\ z_t/D_t + (\beta_{20} + \beta_{22} f_t)\ v_t z_t / D_t + \bar{u}_t \qquad (75)$$

In this case, as in the preceding one, there will be no linear restrictions on the β's, for any $g \geqslant 2$.

Case (1/iv): Assumptions (B/2) and (C/2):

$$y_t = \alpha_o + (\beta'_{10} + \beta'_{20} v_t)\ z_t + \bar{u}_t \qquad (76)$$

This is quite similar to the group functions. But it is valid only under very restrictive assumptions.

Case (2) :

Micro-structures different; assumptions (A/1) and (E/1). The group function is (68), and the following forms of the overall function can be distinguished :

Case (2/1) : Assumptions (B/1) and (C/1) :

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_{10} + \beta_{11} x_t + \beta_{12} f_t + \beta_{13} x_t f_t)\ z_t/D_t + (\gamma_o + \gamma_1 x_t)\ V_t + \bar{u}_t \qquad (78)$$

Case (2/ii): Assumptions (B/1) and (C/2):

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_{10} + \beta_{11} x_t)\ z_t/D_t + (\gamma_o + \gamma_1 x_t)\ V_t + \bar{u}_t \qquad (78)$$

Case (2/iii): Assumptions (B/2) and (C/1):

$$y_t = \alpha_o + (\beta_{10} + \beta_{12} f_t)\ z_t/D_t + \gamma_o V_t + \bar{u}_t \qquad (79)$$

Case (2/iv): Assumptions (B/2) and (C/2):

$$y_t = \alpha_o + \beta'_{10} z_t + \gamma_o V_t + \bar{u}_t \qquad (80)$$

Again it is case (iv) which resembles the group functions, though it is valid under very restrictive assumptions.

Case (3) :

The group function is of the form (69), which arises under : assumption (A/2); or assumption (A/1) together with (D/2), (D/3), (E/2) or (E/3). Aggregation gives:

$$y_t = \sum_i a_i n_{it} + \sum_i b_i z_{it} n_{it} + \bar{u}_t$$

This is similar to case (2) with the exclusion of term in V_t. The following four sub-cases can be distinguished.

Case (3/i) : Assumptions (B/1) and (C/1):

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_o + \beta_1 x_t + \beta_2 f_t + \beta_3 x_t f_t) \; z_t / D_t + \bar{u}_t \qquad (81)$$

The β's are the same as β_{1k} defined in (71), dropping the subscript from the β's and the b's.

Case (3/ii): Assumptions (B/1) and (C/2):

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_o + \beta_1 x_t) \, z_t / D_t + \bar{u}_t \qquad (82)$$

Case (3/iii): Assumptions (B/2) and (C/1):

$$y_t = \alpha_o + (\beta_o + \beta_2 f_t) \; z_t / D_t + \bar{u}_t \qquad (83)$$

Case (3/iv) : Assumptions (B/2) and (C/2) :

$$y_t = \alpha_o + \beta'_o z_t + \bar{u}_t \qquad (84)$$

Apart from this last sub-case, the overall function differs from the group functions, and the variable z_t does not appear alone in any term, at least it has to be divided by D_t.

Case (4) :

Micro-structures are similar for the whole economy. The group functions (70) have the same structure, and the same applies to the overall function :

$$y_t = a + b' z_t + \bar{u}_t \qquad (85)$$

This holds true irrespective of the alternatives of assumptions (B) and (C).

To conclude, the assumption that aggregate demand is a linear function of aggregate function only does not follow from the micro-theory, except under very restrictive assumptions, as in (84) and (85). Cases (1) and (2) lead to very complicated, but still linear, functions. It could be argued that case (3) might give a plausible approximation, especially as it holds under a number of alternative assumptions. Further, since the basic idea of grouping is to represent the behaviour of similar individuals by that of a single group, assumption (A/2) seems to be a reasonable starting point. Since the distribution of population among groups is usually subject to change, we have to use (81) or at least (82).

In both cases it is z_t / D_t rather than z_t which appears as an explanatory variable. Information on x_t can be easily obtained from the analysis of population data, usually more readily available. One the other hand, the behaviour of z_{it} through f_t can be estimated from other sources, e.g., sample survey data, or income-tax statistics. It should be remembered that the degree of approximation is relatively high in this latter case, which raises the problem of treating the errors in observation in the explanatory variables.

XII - The Linear Form; Special Case $g = 2$:

The present paper has been initiated by the need for a study of the aggregation problem in an economy composed of two groups:

$i = 1$, for rural population
$i = 2$, for urban population

It is worthwhile to study this special case, since it offers certain simplifications and envolves some restrictions on the parameters, as mentioned in the previous section.

Assumptions (A), (D), (E) and (F) do not raise any particular problems. Let us therefore consider assumption (B). Suppose we put $x_t = n_{2t}$. Then (31) can be explicitly written as follows:

$$n_{1t} = 1 - n_{2t} = 1 - x_t \qquad n_{2t} = x_t \tag{86}$$

Hnece,

$$m_{10} = 1, \qquad m_{11} = -1, \qquad m_{20} = o, \qquad m_{21} = 1$$

Assumption (B/2) means that x_t in this case is a constant = x_o, say, which can be considered as the value of x_t at the base period. Then (33) gives :

$$m_1 = 1 - m_2 = 1 - x_o, \qquad m_2 = x_o \qquad (87)$$

Assumption (C) can be expressed in terms of one of the group incomes, e.g., z_{1t}. Let D_{it} be the ratios of z_{it} to z_{1t}. Then :

$$D_{it} = 1, \qquad D_{2t} = (do + d_1 f_t) = z_{2t}/ z_{1t}$$

$$D_t = 1 + \left[(d_0 - 1) + d_1 f_t\right] n_{2t}$$

where n_{2t} is equal to either x_1 or x_o, according to the variant of of assumption (B) assumed. Hence,

$$z_{1t} = z_t / D_t, \qquad z_{2t} = (do + d_1 f_t) \; z_t/ D_t \qquad (88)$$

In other words,

$$d_{10} = 1, \qquad d_{11} = 0, \qquad d_{20} = d_o, \qquad d_{21} = d_1$$

The variable z_t/ D_t figuring in the overall functions of the previous section is, in fact, z_{1t} under the present assumption. If (C/2) is assumed, we ignore f and write :

$$d = z_{2t}/ z_{1t} \qquad D_t = 1 + (d - 1) n_{2t}$$

Equations (88) become :

$$z_{1t} = z_t/ D_t, \qquad z_{2t} = d (z_t/ D_t) \qquad (89)$$

which means that:

$$d_{10} = 1, \qquad d_{11} = 0, \qquad d_{20} = d, \qquad d_{21} = 0$$

If n_{it} are constant as in (87), D_t will be also constant.

The cases considered in the previous section are reduced to the following :

Case (1) :
=========

Group functions are (67); (A/1) and (D/1)

Case (1/i) : Assumptions (B/1) and (C/1).

The macro-parameters in (71) become :

$$\left.\begin{aligned} \alpha_o &= a_1 & \alpha_1 &= (a_2 - a_1) \\ \beta_{10} &= b_{10} & \beta_{20} &= b_{11} \\ \beta_{11} &= (d_o\, b_{20} - b_{10}) & \beta_{21} &= (d_o\, b_{21} - b_{11}) \\ \beta_{12} &= 0 & \beta_{22} &= 0 \\ \beta_{13} &= d_1\, b_{20} & \beta_{23} &= d_1\, b_{21} \\ \gamma_o &= b_{11} & \gamma_1 &= b_{21} - b_{11} \end{aligned}\right\} \quad (90)$$

It is clear that these parameters satisfy the following linear restrictions

$$\begin{aligned} \beta_{k2} &= 0, & \beta_{k3} &= (d_1/ d_o)\,(\beta_{k1} + \beta_{ko}) \\ \gamma_o &= \beta_{20} & \gamma_1 &= [\beta_{21} + (1 - d_o)\,\beta_{20}] \,/\, d_o \end{aligned} \quad (91)$$

where k = 1 and 2. Notice, however, that the restrictions on the γ's are irrelevant since they replace β_{2k}.

Le us introduce the following composite variables:

$$x_{ot} = x_t$$

$$\left.\begin{aligned} x_{1t} &= (d_o + d_1\, x_t\, f_t)\, z_t / d_o\, D_t \\ x_{2t} &= (d_o\, x_t + d_1\, x_t\, f_t)\, z_t / d_o\, D_t \\ x_{3t} &= x_{1t}\, v_t \\ x_{4t} &= x_{2t}\, v_t \end{aligned}\right\} \quad (92)$$

The overall function is :

$$y_t = \alpha_o + \alpha_1 x_o + \beta_{10} x_{1t} + \beta_{11} x_{2t}$$

$$+ \beta_{20} x_{3t} + \beta_{21} x_{4t} + \bar{u}_t \qquad (93)$$

<u>Case (1/ii)</u> : Assumptions (B/1) and (C/2):

In this case we have to put $d_1 = 0$ and $d_o = d$ in (90) and (91). This means that the same restrictions hold with the exception that:

$$\beta_{k3} = 0 \qquad (94)$$

Hence we do not need to introduce the transformations (92); in fact if we do, they will be redundant, since $d_1 = 0$. The overall function is exactly the same as (74), valid for the general case. It is exactly similar to (93) if we define (92) for the case $d_1 = 0$.

<u>Case (1/iii)</u>: Assumptions (B/2) and (C/1)

In this case we can apply (75) as in the general case with the following interpretation of the parameters:

$$\left.\begin{array}{llll}
\alpha_o = a_1 + x_o (a_2 - a_1) & & \alpha_1 = 0 \\
\beta_{10} = b_{10} + x_o (d_o b_{20} - b_{10}) & & \beta_{11} = 0 \\
\beta_{12} = d_1 b_{20} x_o & & \beta_{13} = 0 \\
\beta_{20} = b_{11} + x_o (d_o b_{21} - b_{11}) & & \beta_{21} = 0 \\
\beta_{22} = d_1 b_{21} x_o & & \beta_{23} = 0 \\
\gamma_o = b_{11} + x_o (b_{21} - b_{11}) & & \gamma_1 = 0
\end{array}\right\} \qquad (95)$$

Apart from the zero-value conditions, the only linear restriction is that of :

$$\gamma_o = \beta_{20} + \beta_{22} (1 - d_o)/ d_1 \qquad (96)$$

which again is irrelevant since γ_o replaces the coefficients β_{2k}.

Case (1/iv) : Assumptions (B/2) and (C/2)

Defining $\beta'_{ko} = \beta_{ko} / D$, as in (71), where

$$D = 1 + x_o (d - 1) \qquad (97)$$

we can use (76) with the parameters defined according to (95). Notice that in the present case, $d_o = d$ and $d_1 = 0$, which implies that $\beta_{k2} = 0$ $(k = 1, 2)$. The restriction (96) is irrelevant.

Case (2) :

Assumptions (A/1) and (E/1)

Case (2/i): Assumptions (B/1) and (C/1)

Equations (77) can be rewritten as follows :

$$y_t = (\alpha_o + \alpha_1 x_{ot}) + \beta_{10} x_{1t} + \beta_{11} x_{2t} + (\gamma_o + \gamma_1 x_{ot}) V_t + \bar{u}_t \qquad (98)$$

where the coefficients are defined according to (90), and the variables x_{ot}, x_{1t} and x_{2t} are defined according to (92).

Case (2/ii): Assumptions (B/1) and (C/2):

The overall function is exactly the same as (78), which is similar to (98) with the definition of the x's taking into consideration that $d_1 = 0$.

Case (2/iii): Assumptions (B/2) and (C/1):

The overall function is the same as (79), with the coefficients satisfying (95).

Case (2/iv) : Assumptions (B/2) and (C/2):

Equation (80) applies with the definitions of parameters as in (95).

Case (3) :

Assumption (A/1) together with (D/2), (D/3), (E/2) or (E/3), or assumption (A/2). The overall functions can be easily derived from those of the general case, taking into consideration the conditions of case (1) of the present section. Thus :

Case (3/i) : Assumptions (B/1) and (C/1)

$$y_t = (\alpha_o + \alpha_1 x_{ot}) + \beta_o x_{1t} + \beta_1 x_{2t} + \bar{u}_t \qquad (99)$$

where the variables x are defined by (92) and the coefficients β_k are the same as β_{1k} in (90), (k = 0, 1).

Case (3/ii): Assumptions (B/1) and (C/2):

The overall function is the same as (82), with the same interpretation of coefficients as in the previous sub-case.

Case (3/iii) : Assumptions (B/2) and (C/1):

Equation (83) can be used with coefficients defined by (95), β_{1k} being replaced by β_k.

Case (3/iv): Assumptions (B/2) and (C/2)

Equation (84) applies with $\beta'_o = \beta_{10}/ D$, defined according to (95) and (97).

Case (4) : Assumption (A/3), for any (B) or (C) The overall function is (85).

In spite of the simplifications offered by the condition that g = 2, the variety of overall functions is rather startling. To ensure homoscedasticity we still have to apply the transformations suggested in section IX. Problems of estimation are not completely solved this way, owing to the possibilities of introducing observation errors through such transformations.

XIII - The Parabolic Form :

Suppose that y is a quadratic function of z, such that the micro-functions, according to assumption (A/1) become :

$$y_{ijt} = a_{ij} + b_{ij} z_{ijt} + c_{ij} z^2_{ijt} + u_{ijt} \qquad (100)$$

It has been shown that the assumption of completely different structures leads - even in the linear case - to a number of complications; and that it would be quite reasonable to assume similarity within groups:

$$y_{ijt} = a_i + b_i z_{ijt} + c_i z_{ijt}^2 + u_{ijt} \qquad (101)$$

Using the notation for the variance of income, (39), and knowing that

$$\frac{1}{N_{it}} \sum_j z_{ijt}^2 = z_{it}^2 + V_{it}^2 \qquad (102)$$

We can obtain the group functions under assumption (A/2) leading to (101):

$$y_{it} = a_i + b_i z_{it} + c_i (z_{it}^2 + V_{it}^2) + u_{it} \qquad (103)$$

The overall function is:

$$y_t = \sum_i a_i n_{it} + \sum_i b_i n_{it} z_{it} + \sum_i c_i n_{it} (z_{it}^2 + V_{it}^2) + \bar{u}_t$$

This envolves:

1. Population distribution which is generally variable over time, in a known way. Hence it is necessary to consider both alternatives of assumption (B), of section IX.

2. Group incomes and their squares. The variability introduced by assumption (C/1) leads to quite complicated formulae for incomes and squares. It seems reasonable to assume (C/2) which is a good approximation, at least over limited periods.

3. Group variances, which were assumed under assumption (E) to satisfy certain restrictions. If it is considered better to assume these restrictions as satisfied by the coefficient of variation, it would be preferable to introduce assumption (D) instead. This means that we replace V by vz as in (39).

Accordingly the following cases can be distinguished

Case (1): Assumptions (A/2), (B/1), and (C/2):

The following notation for the macro-parameters will be used:

$$\alpha_0 = \sum_i a_i m_{io} \qquad \alpha_1 = \sum_i a_i m_{i1}$$

$$\alpha_2 = \sum_i a_i m_{io} (V_i^2 + 1) \qquad \alpha_3 = \sum_i a_i m_{i1} (V_i^2 + 1)$$

$$\left.\begin{array}{ll} \beta_o = \sum_i b_i m_{io} d_i & \beta_1 = \sum_i b_i m_{i1} d_i \\ \gamma_o = \sum_i c_i m_{io} & \gamma_1 = \sum_i c_i m_{i1} \\ \gamma_2 = \sum_i c_i m_{io} d_i^2 & \gamma_3 = \sum_i c_i m_{i1} d_i^2 \\ \gamma_4 = \sum_i c_i m_{io} (v_i^2 + 1) & \gamma_5 = \sum_i c_i m_{i1} (V_i^2 + 1) \end{array}\right\} \quad (104)$$

The notation for parameters envolving v or V will remain the same in the cases where they are equal for all i.

Case (1/i): Assumption (E/1):

$$y_t = (\alpha_o + \alpha_i \; x_t) + (\beta_o + \beta_1 x_t) \; z_t/D_t + (\gamma_o + \gamma_1 x_t) \; V_t^2 + (\gamma_2 + \gamma_3 \; x_t) \; z_t^2/D_t^2 + \bar{u}_t \quad (105)$$

where, as before,

$$D_t = \sum_i d_i n_{it}$$

Case (1/ii): Assumption (D/1)

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_o + \beta_1 x_t) \; z_t/D_t + (\gamma_2 + \gamma_3 x_t) \; (1 + v_t^2) \; z_t^2/D_t^2 + \bar{u}_t \quad (106)$$

Case (1/iii): Assumptions (E/2) or (E/3):

$$y_t = (\alpha_2 + \alpha_3 \; x_t) + (\beta_o + \beta_1 x_t) \; z_t/D_t + (\gamma_2 + \gamma_3 x_t) \; z_t^2/D_t^2 + \bar{u}_t \quad (107)$$

Case (1/iv): Assumptions (D/2) or (D/3):

$$y_t = (\alpha_o + \alpha_1 \; x_t) + (\beta_o + \beta_1 x_t) \; z_t/D_t + (\gamma_4 + \gamma_5 x_t) \; z_t^2/D_t^2 + \bar{u}_t \quad (108)$$

It is interesting to notice that (107) is exactly similar to (108) apart from the interpretation of the coefficients α and γ.

Case (2): Assumtpions (A/2), (B/2) and (C/2);

In this case $D_t = D$ is a constant, hence we define the coefficients of z or its square as in (104) after division by D or D^2. Further, since m_{i1} = 0, we can obtain the equations for the alternative sub-cases distinguished in case (1) by putting $x_t = 0$ in (105) - (108). The reader can easily find out the expressions required. Notice that for sub-cases (iii) and (iv), the overall function will be exactly similar to the micro-functions, while in sub-cases (i) and (ii), it is similar to the group functions, but without restricting the coefficient of V^2 to be equal to that of z^2, in (i).

Consider now the implications of assumption (A/3). The micro-functions are:

$$y_{ijt} = a_o + b_o z_{ijt} + c_o z_{ijt}^2 + u_{ijt} \tag{109}$$

the group functions would be:

$$y_{it} = a_o + b_o z_{it} + c_o (z_{it}^2 + V_{it}^2) + u_{it} \tag{110}$$

and the overall function becomes:

$$y_t = a_o + b_o z_t + c_o \sum_i n_{it} (d_{it}^2 z_t^2 + V_{it}^2) + \bar{u}_t$$

We define the following variables:

$$E_t = \sum_i n_{it} d_{it}^2, \qquad F_t = \sum_i n_{it} d_{it}^2 (1 + v_{it}^2) \tag{111}$$

and the following parameters:

$$\begin{aligned} a_1 &= a_o + c_o \sum_i m_{io} V_i^2 & c_1 &= c_o (1 + v^2) \\ a_2 &= c_o \sum_i m_{i1} V_i^2 & c_2 &= c_o E \\ a_3 &= a_o + c_o V_i^2 & c_3 &= c_o F = c_1 E \end{aligned} \tag{112}$$

where

$$E = \sum_i m_i d_i^2 \qquad F = \sum_i m_i d_i^2 (1 + v_i^2) \tag{113}$$

The following cases can be distinguished:

Case (3): Assumptions (A/3), (B), (C/1); or (A/3), (B/1), (C/2).

Case (3/i): Assumption (E/1):

$$y_t = a_o + b_o z_t + c_o (E_t z_t^2 + V_t^2) \tag{114}$$

Case (3/ii): Assumption (E/2)

The overall function is the same as (114), with

$$V_t^2 = \sum_i n_{it} V_i^2 = \sum_i m_{io} V_i^2 + \sum_i m_{i1} V_{i1}^2 x_t \tag{115}$$

Using the last expression, we can dispense with the V's and rewrite the equation as follows:

$$y_t = a_1 + c_2 x_t + b_o z_t + c_o (E_t z_t^2) + \bar{u}_t \qquad (116)$$

If (B/2) is assumed this last equation becomes:

$$y_t = a_1 + b_o z_t + c_o (E_t z_t^2) + \bar{u}_t \qquad (117)$$

<u>Case (3/iii):</u> Assumption (E/3):

$$y_t = a_3 + b_o z_t + c_o (E_t z_t^2) + \bar{u}_t \qquad (118)$$

which differs from (117) in the interpretation of the constant term.

<u>Case (3/iv):</u> Assumption (D/2)

$$y_t = a_o + b_o z_t + c_o (F_t z_t^2) + \bar{u}_t \qquad (119)$$

which is similar to (118), since under assumption (D/1) we have:

$$F_t = (1 + v_t^2) E_t$$

<u>Case (3/vi):</u> Assumption (D/3)

$$y_t = a_o + b_o z_t + c_1 (E_t z_t^2) + \bar{u}_t$$

which is similar to (117) apart from the definition of the parameters a and c.

<u>Case (4):</u> Assumptions (A/3), (B/3), (B/2) and (C/2).

In this case (111) is replaced by (113):

<u>Case (4/i):</u> Assumption (E/1):

$$y_t = a_o + b_o z_t + c_2 z_t^2 + c_o V_t^2 + \bar{u}_t \qquad (120)$$

which is similar to the group function, but the coefficients of V^2 and z^2 are different.

<u>Case (4/ii):</u> Assumption (E/2):

$$y_t = a_1 + b_o z_t + c_2 z_t^2 + \bar{u}_t \qquad (121)$$

This resembles, formally, the micro-equations but the parameters a and c differ in meaning.

Case (4/iii): Assumption (E/3)

$$y_t = a_3 + b_o z_t + c_2 z_t^2 + \bar{u}_t$$

which differs from (121) in the constant term only.

Case (4/iv): Assumption (D/1)

$$y_t = a_o + b_o z_t + c_2 (1 + v_t^2) z_t^2 + \bar{u}_t \qquad (122)$$

which is similar to the corresponding group functions.

Case (4/v): Assumption (D/2) or (D/3)

$$y_t = a_o + b_o z_t + c_3 z_t^2 + \bar{u}_t$$

which is similar to (121).

Thus, even under the most restrictive conditions, the simple case of assumption (3) leads to formulae different from the micro-functions. The constant income and population distributions hypotheses do not guarantee the interpretation of the macro-parameters as being those of the micro function. On the other hand, the types of functions studied in this section do not raise any specific problems for the case g = 2. Nor do they envolve any new problems as far as the treatment of disturbances is concerved.

XIV. The Linear Logarithmic Form:

In some case the micro-demand function is assumed to be a linear function in the logarithm of income, rather than income itself.

$$y_{ijt} = a_{ij} + b_{ij} \log z_{ijt} + u_{ijt} \qquad (123)$$

If assumption (A/1) is satisfied, the group functions can be obtained by means of (28):

$$y_{it} = a_{io} + b_{io} \frac{1}{N_{it}} \sum_j \log z_{ijt} + \frac{1}{N_{it}} \sum_j (b_{ij} - b_{io}) \log z_{ijt} + u_{it}$$

The second term in this expression is the logarithm of the geometric mean g_{it}, of the group income distribution, as defined by (17). According to the notation of Table (1), we have:

$$\log g_{it} = \log z_{it} + \log r_{it}.$$

On the other hand, the use of the logarithms is another way of expressing relative incomes. It follows that the variance θ_{it}^2 of the distribution of logarithms is a monotonically increasing function of the variance of the relative income distribution v_{it}^2, which is the square of the coefficient of variation $(V_{it}/z_{it})^2$. In fact, under assumption (F) of a log-normal distribution, we have by (48)

$$\theta_{it}^2 = \log (v_{it}^2 + 1)$$

Hence, the third term in the expression for y_{it} is approximated by:

$$\frac{1}{N_{it}} \sum_j (b_{ij} - b_{io}) \log z_{ijt} = b_{i1} \; \theta_{it}$$

where b_{i1} is defined as in (66), with θ_{it} being the (constant) correlation coefficient between the b's and the logarithms of z's. The group functions become

$$y_{it} = a_{io} + b_{io} (\log z_{it} + \log r_{it}) + b_{i1} \theta_{it} + u_{it} \qquad (124)$$

Under assumption (F) of a log-normal distribution of incomes, we can use (46) to rewrite this equation as:

$$y_{it} = a_{io} + b_{io} (\log z_{it} - \tfrac{1}{2}\theta_{it}^2) + b_{i1} \theta_{it} + u_{it} \qquad (125)$$

The aggregation of (124) gives:

$$y_t = \sum_i a_{io} n_{it} + \sum_i b_{io} n_{it} (\log z_{it} + \log r_{it}) + \sum_i b_{i1} n_{it} \theta_{it} + \bar{u}_t$$

Suppose that:

$$\frac{b_{io} n_{it} \log z_{it}}{b_{io} n_{it}} = \log A + \sum_i n_{it} \log z_{it}$$

where A is a constant. This means that the logarithms of the geometric means of the z's are proportionate; these means being weighted by b_{io} n_{it} and by n_{it} respectively. According to table (1), this latter mean can be expressed as:

$$\sum_i n_{it} \log z_{it} = \log z_t + \log \delta_t$$

$$\log \delta_t = \sum_i n_{it} \log d_{it}$$

where δ_t is defined according to (21).

If structures within groups are similar, assumption (A/2), the group functions (124) become

$$y_{it} = a_{io} + b_{io}\ (\log z_{it} + \log r_{it}) + u_{it} \tag{126}$$

Finally if all structures are similar, we drop the subscripts i from the parameters a and b. The following cases can be distinguished:

<u>Case (1):</u> Assumptions (A/1) and (D/1):

The assumption that $v_{it} = v$ implies that

$r_{it} = r_t = R_t$ and, $\theta_{it} = \theta_t$

where R is defined according to (24).

<u>Case (1/i):</u> Assumptions (B/1) and (C/1) or (C/2).

The overall function is:

$$y_t = (\alpha_o + \alpha_1\ x_t) + (\beta_o + \beta_1\ x_t)(\log z_t + \log \delta_t + \log R_t) + (\gamma_o + \gamma_1\ x_t)\ \theta_t + \bar{u}_t \tag{127}$$

In this and the following equations, the macro-parameters are defined as follows:

$$\left.\begin{aligned}
\alpha_o &= \sum_i (a_{io} + b_{io} \log A)\ m_{io} \\
\alpha_1 &= \sum_i (a_{io} + b_{io} \log A)\ m_{i1} \\
\alpha_2 &= \alpha_o + (\log \delta) \sum_i b_{io}\ m_{io} \\
\alpha_3 &= \alpha_o + \sum_i (b_{io} \log r_i + b_{i1}\ \theta_i)\ m_{io} \\
\alpha_4 &= \alpha_o + \sum_i (b_{io} \log r_i + b_{i1}\ \theta_i)\ m_{i1} \\
\alpha_5 &= \alpha_3 + (\log \delta) \sum_i b_{io}\ m_{io} \\
\beta_o &= \sum_i b_{io}\ m_{io} \qquad \beta_1 = \sum_i b_{io}\ m_{i1} \\
\gamma_o &= b_{i1}\ m_{io} \qquad \gamma_1 = \sum_i b_{i1}\ m_{i1}
\end{aligned}\right\} \tag{128}$$

Case (1/ii): Assumptions (B/2) and (C/1)

$$y_t = \alpha_o + \beta_o (\log z_t + \log \delta_t + \log R_t) + \gamma_o \theta_t + \bar{u}_t \qquad (129)$$

Case (1/iii): Assumptions (B/2) and (C/2).

In this case δ_t becomes constant, and when substituted in (128) gives

$$y_t = \alpha_2 + \beta_o (\log z_t + \log R_t) + \gamma_o \theta_t + \bar{u}_t \qquad (130)$$

Case (2): Assumptions (A/1) and (D/2) or (D/3).

In this case:

$r_{it} = r_i$ or r, $\qquad \theta_{it} = \theta_i$ or θ

Case (2/i): Assumptions (B/1) and (C/1) or (C/2)

$$y_t = (\alpha_3 + \alpha_4 x_t) + (\beta_o + \beta_1 x_t)(\log z_t + \log \delta_t) + \bar{u}_t \qquad (131)$$

Case (2/ii): Assumptions (B/2) and (C/1)

$$y_t = \alpha_3 + \beta_o (\log z_t + \log \delta_t) + \bar{u}_1 \qquad (132)$$

Case (2/iii): Assumptions (B/2) and (C/2)

$$y_t = \alpha_5 + \beta_o \log z_t + \bar{u}_t \qquad (133)$$

In this sub-case, the overall function is similar to the micro-functions.

Case (3): Assumptions (A/2) and (D/1)

This case is similar to case (1), putting $b_{i1} = o$, hence $\gamma_k = o$.

Case (3/i): Assumptions (B/1) and (C/1) or (C/2):

$$y_t = (\alpha_o + \alpha_1 x_t) + (\beta_o + \beta_1 x_t)(\log z_t + \log \delta_t + \log R_t) + \bar{u}_t \qquad (134)$$

Case (3/ii): Assumptions (B/2) and (C/1)

$$y_t = \alpha_o + \beta_o (\log z_t + \log \delta_t + \log R_t) + \bar{u}_t \qquad (135)$$

Case (3/iii): Assumptions (B/2) and (C/2)

$$y_t = \alpha_2 + \beta_o (\log z_t + \log R_t) + \bar{u}_t \qquad (136)$$

which is similar to the group functions (126)

Case (4): Assumptions (A/2) and (D/2) or (D/3)

The overall function takes a form similar to (131)-(133), with $b_{i1} = o$ in the parameter values defined by (128).

Case (5): Assumption (A/3).

The group functions are:

$$y_{it} = a_o + b_o (\log z_{it} + \log r_{it}) + u_{it} \tag{137}$$

The following parameters are envolved in aggregation

$$a_1 = a_o + b_o \sum_i m_{io} \log r_i$$

$$a_2 = a_o + b_o \log \delta \qquad a_3 = a_1 + b_o \log \delta \tag{138}$$

Case (5/i): Assumptions (B/1), (C/1), and (D/1) or (D/2);

assumptions (B/1), (C/2) and (D/1) or (D/2);

and assumptions (B/2), (C/1) and (D/1)

$$y_t = a_o + b_o (\log z_t + \log \delta_t + \log R_t) + \bar{u}_t \tag{139}$$

Case (5/ii): Assumptions (B/1), (D/3) and (C/1) or (C/2); and assumptions (B/2), (C/1) and (D/2) or (D/3)

$$y_t = a_1 + b_o (\log z_t + \log \delta_t) + \bar{u}_t \tag{140}$$

Case (5/iii): Assumptions (B/2), (C/2) and (D/1)

$$y_t = a_2 + b_o (\log z_t + \log R_t) + \bar{u}_t \tag{141}$$

Case (5/iv): Assumptions (B/2), (C/2) and (D/2) or (D/3)

$$y_t = a_3 + b_o \log z_t + \bar{u}_t \tag{142}$$

It might be noticed, in conclusion that apart from the term in θ included in case (1), the term envolving z is in fact the logarithm of the overall geometric mean. As shown in table (1) and equations (17) and (25):

$$\log g_t = \log z_t + \log p_t = \log z_t + \log \delta_t + \log R_t$$

The cases where δ_t or R_t are constant lead to changes in these terms, with a corresponding change in the constant terms, or terms in x_t. In fact we could maintain the general form in all equations.

XV. The Exponential Function:

An assumption frequently made in demand studies is that demand can be represented by an exponential function in income:

$$y = k.m^{z}$$

For purposes of statistical estimation, this function is expressed in an inverse logarithmic form. Thus, assuming different economic structures, the micro-functions are:

$$\log y_{ijt} = a_{ij} + b_{ij} z_{ijt} + u_{ijt} \qquad (143)$$

Neither form yields itself to straightforward processes of aggregation. The simple process of summation does not lead to a manageable aggregate of incomes in the exponential form. In the inverse logarithmic case, it envolves the same difficulties of aggregating income as in the linear case, and at the same time leads to the geometric rather than the arithmetic average of y. However, no other process is capable of solving both sides to the same form. We shall apply the summation process to (143) and make use of the relations established in Table (1) of section XI to obtain simple aggregate variables.

The group functions are:

$$\log \bar{h}_{it} = a_{io} + b_{io} z_{it} + b_{i1} V_{it} + u_{it} \qquad (144)$$

where,

$$\log h_{it} = \log y_{it} + \log s_{it}$$

s being the ratio of the geometric to the arithmetic means. The R. H.S. is derived in a way similar to that of (68) in the linear case. The question might be raised as to whether there is any relation-ship between the V_{it} and the s_{it}.

As had been indicated before, the log- normal assumption has established a relation between the variance and the ratio between the geometric and arithmetic means. Therefore, it can be expected that if the V's follow a certain pattern, the ratios s are liable to follow a similar pattern. This means that the alternatives of assumption (E) should have their implications for S_{it}. Such an assumption would help to deal with cases in which information on the latter is scanty. It would be expected that such is the general case, unless frequent surveys were available.

The secondary process of aggregation leads to the overall geometric mean, which according to (18), (24) and (25) satisfies:

$$\log \bar{h}_t = \log y_t + \log q_t = \log y_t + \log \epsilon_t + \log S_t$$

with,

$$\log S_t = \sum_i n_{it} \log s_{it}$$

The overall function can be considered with respect to alternatives of assumption (E), as in case (2) of section XI. Further, if we assume similar structures within the groups or for the whole economy, we can put $b_{i1} = o$ in (144) which means that the terms in V vanish.

Case (1): Assumptions (A/1) and (E/1)

In this case we assume that $V_{it} = V_t$ and, $s_{it} = s_t = S_t$. The most general form of the function is that corresponding to (77), satisfying (B/1) and (C/1):

$$\log y_t + \log \epsilon_t + \log S_t = (\alpha_o + \alpha_1 x_t) + (\gamma_o + \gamma_1 x_t) V_t + (\beta_{10} + \beta_{11} x_t + \beta_{12} f_t + \beta_{13} x_t f_t) z_t / D_t + \bar{u}_t \quad (145)$$

Where the coefficients satisfy (71). Under the more restrictive assumptions of (B/2) and/or (C/2) we can obtain the overall function from (145) by putting x_t and/or f_t equal to zero.

Case (2): Assumptions (A/1) and (E/2), or (E/3)

We put $V_{it} = V_i$ and $s_{it} = s_i$. Further we define

$$a_i = a_{io} + b_{i1} V_i$$

as in (69). This brings the group functions to the form derived directly on the basis of (A/2). As in section XI, the general form of the overall function can be derived from (145) by omitting the terms in V_t:

$$\log y_t + \log \varepsilon_t + \log S_t = (\alpha_0 + \alpha_1 x_t) + (\beta_0 + \beta_1 x_t + \beta_2 f_t + \beta_3 x_t f_t) z_t/D_t + \bar{u}_t \qquad (146)$$

Further simplification due to assumptions (B/2)or (C/2) can be derived as in the previous case. Under assumption (B/2).

$$\log S_t = \sum_i m_i \log s_i = \text{constant}$$

which can be transfered to the constant term.

Case (3): Assumption (A/3) of similar structures leads to the overall equation.

$$\log h_t = a + b z_t + \bar{u}_t \qquad (147)$$

We can substitute for log h_{it} its appropriate expression in terms of log y_{it} as before.

For case g = 2, we have to account for the linear restrictions discussed in Section XI. On the other hand, since we have adopted a process of simple summation, we can apply the transformations suggested in section X for the purposes of estimation.

XVI. The Power Function:

The constant-elasticity assumption leads to a form in which income is raised to a power equal to the elasticity:

$$y = k.z^e$$

For purposes of statistical estimation, it is usually assumed that a double-logarithmitic form satisfies the desirable properties:

$$\log y_{ijt} = a_{ij} + b_{ij} \log z_{ijt} + u_{it} \qquad (148)$$

According to assumption (A/1), and applying rules similar to those developed in the last two sections, we can write the group functions as follows:

$$\log y_{it} + \log s_{it} = a_{io} + b_{io} (\log z_{it} + \log r_{it}) + b_{i1} \theta_{it} + u_{it} \qquad (149)$$

which corresponds to (124) and (144), If (A/2)is assumed, $b_{i1} = 0$ and the term in θ vanishes. Similarly for (A/3).

Under assumptions similar to these in section XIII, the aggregation of b_{io} ($\log z_{it} + \log r_{it}$) can be approximated by:

$$(\beta_o + \beta_1 x_t)(\log z_t + \log \delta_t + \log R_t) = (\beta_o + \beta_1 x_t) \log g_t$$

This leads to the following main cases:

<u>Case (1)</u>: Assumption (A/1)

Under the most general assumptions (B/1) and (C/1), the overall function is:

$$\log y_t + \log q_t = (\alpha_o + \alpha_1 x_t) + (\beta_o + \beta_1 x_t)(\log z_t + \log p_t) + (\gamma_o + \gamma_1 x_t)\theta_t + \bar{u}_t \qquad (150)$$

Where, as before,

$$\log q_t = \log \varepsilon_t + \log S_t, \qquad \log p_t = \log \delta_t + \log R_t$$

To both r and s we can apply similar alternative assumptions corresponding to those of v as in (D). On the other hand, if (B/2) is assumed the term in x_t vanishes. If we further assume that (C/2) is satisfied, we can put δ_t = constant, and the same might be assumed for ε_t.

Suppose, for example that

$$\varepsilon_t = \kappa \delta_t \qquad (151)$$

where κ is some constant estimate from a sample survey. Defining, $\alpha_2 = \alpha_o + \log \kappa$, we can rewrite (150) as follows:

$$\log y_t + \log S_t = (\alpha_2 + \alpha_1 x_t) + (\beta_o + \beta_1 x_t) \log g_t + (\gamma_o + \gamma_1 x_t)\theta_t - \log \delta_t + \bar{u}_t \qquad (152)$$

Further if we assume that $s_{it} = s_i$, we can write

$$\log S_t = \sum_i m_{io} \log s_i + \sum_i m_{i1} (\log s_i) x_t = \alpha_3 + \alpha_4 x_t$$

This can be added to the constant term, which furhter simplifies (152). A similar simplification can be made for the case $s_{it} = s$. In both cases the function becomes:

$$\log y_t = (\alpha'_o + \alpha'_1 x_t) + (\beta_o + \beta_1 x_t)(\log z_t + \log p_t) + (\gamma_o + \gamma_1 x_t)\,\theta_t - \log \delta_t + \bar{u}_t \qquad (153)$$

Case (2): Assumption (A/2)

In this case the terms in θ in the equations (150) - (153) can be dropped. Special cases for the particular cases (B/2) or (C/2) can be derived in the same manner.

Case (3): Assumption (A/3)

If similar structures are assumed, the overall function (150) can be reduced to:

$$\log y_t + \log q_t = a + b\,(\log z_t + \log p_t) + \bar{u}_t \qquad (154)$$

which is similar to case (2) in the particular subcase of (B/2). Further, the equation corresponding to (153) is:

$$\log y_t = a_o + a_1 x_t + b\,(\log z_t + \log p_t) - \log \delta_t + \bar{u}_t \qquad (155)$$

In all these cases there will be no restrictions imposed by $g = 2$. On the other hand, the treatment of disturbances will remain the same as before.

XVII. The Ratio Logarithmic Function:

In the analysis of the results of consumers' surveys the savings function is frequently assumed to take the following form:

$$(z_{ijt} - y_{ijt}) / Z_{ijt} = a'_{ij} + b'_{ij} \log z_{ijt} + u'_{ijt} \qquad (156)$$

where $z - y$ = savings. The underlying idea is to transform the savings function into a form where the disturbances can be assumed to be homoscedastic.(15). In what follows we shall allow the a's to differ for different individuals, while the b's remain the same for each group. This enables the savings-ratio to assume different values for different individuals, but restricts changes with the level of income for all indiividuals. The propensity to save will be different for different individuals.

(15) L.R. Klein and J.N. Morgan: "Results of Alternative Statistical Treatments of Sample Survey Data", Journal of the American Statistical Association, 1951, pp. 442 - 460.

Subtracting both sides from unity, we can obtain the consumption function corresponding to the above assumption:

$$y_{ijt}/z_{ijt} = a_{ij} + b_i \log z_{ijt} + u'_{ijt} \tag{157}$$

To obtain the group functions, we multiply both sides by z_{ijt}, then add up over j and average:

$$y_{ijt} = a_{ij}\, z_{ijt} + b_i\, z_{ijt} \log z_{ijt} + u_{ijt} \tag{158}$$

where,

$$u_{ijt} = z_{ijt}\, u'_{ijt}$$

Hence,

$$y_{it} = a_{io}\, z_{it} + \frac{1}{N_{it}} \sum_j (a_{ij} - a_{io})\, z_{ijt} + b_i\, \frac{1}{N_{it}} \sum_j z_{ijt} \log z_{ijt} + u_{it}$$

Applying a rule similar to that underlying (66) and (67), we can write the second term as

$$a_{i1}\, v_{it}\, z_{it}$$

In order to simplify the third term let us assume that income is log-normally distributed,(16), as specified by (43). Dropping the subscripts ans writing:

$$X = \log z, \qquad \exp (X) = z$$

$$A = 1/\sqrt{2\pi} \cdot \theta$$

then

$$E (z \log z) = E (Xe^{X})$$
$$= A \int_{-\infty}^{\infty} X \cdot \exp \left\{ X - \frac{1}{2\,\theta^2} (X - \mu)^2 \right\} dX$$
$$= \exp \left\{ \mu + \tfrac{1}{2}\theta^2 \right\} . A \int_{-\infty}^{\infty} X \exp \left\{ - \tfrac{1}{2\theta^2} (X - \mu - \theta^2) \right\} dX$$

The first factor is, in fact, E(z), as found before, equation (45). The second is the expected value of a normal variate with mean ($\mu + \theta^2$):

$$\therefore\ E (z \log z) = (\mu + \theta^2).\ \exp \left\{ \mu + \tfrac{1}{2}\theta^2 \right\} \tag{159}$$

Substituting the observed values for the expected values, we obtain:

(16) L.R. Klein. A Textbook of Econometrics, pp. 222-3.

$$\frac{1}{N_{it}} \sum_j z_{ijt} \log z_{ijt} = z_{it} \left[\log z_{it} + \tfrac{1}{2} \log (v_{it}^2 + 1)\right]$$

The <u>group functions</u> are

$$y_{it} = (a_{io} + a_{i1} v_{it}) z_{it} + b_i \left[\log z_{it} + \tfrac{1}{2} \log (v_{it}^2 + 1)\right] z_{it} + u_{it} \qquad (160)$$

The question now is whether we should express this in ratio form as suggested by the micro-function.

First, we notice that if it is true that the disturbance of the micro-function in ratio form is homoscedastic, and hence:

$$\text{Var}\,(u'_{ijt}) = \sigma^2, \qquad \text{Cov}\,(u'_{ijt}, z_{ijt}) = 0$$

then,

$$\text{Var}(u_{ijt}) = z_{ijt}^2 \, \sigma^2$$

This means according to (51) and (57),

$$\lambda_{ijt} = z_{ijt}^2, \qquad \lambda_{it} = z_{it}^2 \, (v_{it}^2 + 1)$$

hence by (52)

$$\text{Var}\,(u_{it}) = \frac{1}{N_{it}} \; z_{it}^2 \; (v_{it}^2 + 1) \; \sigma^2$$

Therefore, even if we transform (160) into the ratio form, we still have to divide both sides by the square root of $(v_{it}^2 + 1)/N_{it}$, before the equation is suitable for statistical estimation.

If assumption (A/2) is fully satisfied, then $a_{i1} = 0$. However, the term between brackets in (160) remains the same. It can be seen that it is equal to the logarithm of the quadratic rather than the arithmetic mean of incomes. For further aggregation, it will facilitate matters if we assume that (C/2) is always satisfied.

If (B/1) is assumed, then

$$D_t = \sum_i n_{it} \, d_i$$

will still be variable. Let us introduce the following notation:

$$W_{1t} = 1/D_t, \qquad W_{2t} = (v_t^2 + 1)^{\frac{1}{2}} \qquad (161)$$

It can be seen that $W_{2t} = 1/r_t$ for the special case of assumption (D/1). The aggregation of (160) gives the following forms of the <u>overall function</u>

<u>Case (1):</u> Assumptions (B/1), (C/2) and (D/1):

$$y_t = (\sum_i a_{io} n_{it} d_i) z_t W_{1t} + (\sum_i a_{i1} n_{i1} d_i) v_t z_t W_{1t} + (\sum_i b_i n_{it} d_i)(z_t W_{1t}) \log (z_t W_{1t} W_{2t}) + (\sum_i b_i n_{it} d_i \log d_i)(z_t W_{1t}) + \bar{u}_t$$

Substituting (31) for n_{it} and introducing obvious expressions for the macro-parameters, this equation can be rewritten as:

$$y_t = (\alpha_{10} + \alpha_{11} x_t)(z_t W_{1t}) + (\alpha_{20} + \alpha_{21} x_t)(v_t z_t W_{1t}) + (\beta_o + \beta_1 x_t)(z_t W_{1t}) \log (z_t W_{1t} W_{2t}) + \bar{u}_t \qquad (162)$$

<u>Case (2):</u> Assumptions (B/1), (C/2) and (D/2) or (D/3).

In this case the terms in v will be included in the expressions for the parameters. Thus:

$$\alpha_o = \sum_i a_{io} m_{io} d_i + \sum_i a_{i1} m_{io} d_i (v_i^2 + 1)^{\frac{1}{2}} + \sum_i b_i m_{io} d_i \log (v_i^2 + 1)^{\frac{1}{2}}$$

α_1 can be defined in a similar way, with m_{io} replaced by m_{i1}.

$$\therefore \quad y_t = (\alpha_o + \alpha_1 x_t) z_t W_{1t} + (\beta_o + \beta_1 x_t)(z_t W_{1t}) \log (z_t W_{1t}) + \bar{u}_t \qquad (163)$$

<u>Case (3):</u> Assumptions (B/2), (C/2)

In this case the terms in x_t vanish and W_{1t} will be a constant. Correcting the parameters for this constant factors we can easily obtain the overall functions corresponding to (162) and (163). Thus under assumption (D/1) we have:

$$y_t = \alpha_o z_t + \alpha_1 v_t z_t + \beta_o (z_t W_{2t}) \log (z_t W_{2t}) + \bar{u}_t \qquad (164)$$

If (D/2) or (D/3) is assumed the terms in v_t and W_{2t} vanish from the above expression. The overall function will be similar in form to the micro-function, with some difference as regards the interpretation of parameters.

Case (4): Assumption (A/3)

If similar micro-structures are assumed, the group functions become:

$$y_{it} = a\ z_{it} + b\ z_{it} \log \left[z_{it}\ (v_{it}^2 + 1)^{\frac{1}{2}}\right] + u_{it} \tag{165}$$

Let,

$$F_{1t} = \left(\sum_i n_{it}\ d_i \log d_i \right), \qquad F_{2t} = F_{1t} + \tfrac{1}{2} \sum_i n_{it}\ d_i \log (v_i^2 + 1) \tag{166}$$

If assumptions (B/2), (C/1) and (D/1) are satisfied, the overall function becomes:

$$y_t = a\ z_t + b\ (z_t\ W_{1t}) \left[F_{1t} + \log (z_t\ W_{1t}\ W_{2t})\right] + \bar{u}_t \tag{167}$$

Replacing (D/1) by (D/2) or (D/3), it becomes:

$$y_t = a\ z_t + b\ (z_t\ W_{1t}) \left[F_{2t} + \log (z_t\ W_{1t})\right] + \bar{u}_t \tag{168}$$

The overall formulae derived above, or any versions of them, have to be adjusted for the purposes of statistical estimation. Given the definition of the λ's, we have, by (53) and (51)

$$\text{Var}\ (\bar{u}_t) = \frac{1}{N_t}\ \lambda_t\ \sigma^2, \qquad \lambda_t = \sum_i n_{it}\ z_{it}^2\ (v_{it}^2 + 1) \tag{169}$$

If follows that both sides of the overall function should be multiplied by k_t as defined by (58). Only when (B/2), (C/2) and (D/3) are satisfied does the factor k_t be a simple function of z_t:

$$k_t = \sqrt{N_t}/z_t \tag{170}$$

Further, if N_t itself is constant, we can justify the simple transformation back into the ratio form as suggested by Klein.(17)

(17) L.R. Klein: A Textbook of Econometrics, p. 223

XVIII. Suggestions for Further Research:

In spite of the fact that we have taken every care to reduce the size of the problem to its simplest orders, the outcome has been found to be very much diversified as well as complicated. The aggregate function was found to envolve a large number of variables which depend on the factors determining the distribution of incomes, as well as on the form of the functional relationships. It might be found necessary to carry the research further in one or more of the following directions:

1. To consider other forms of the demand-income relationship. For example, micro-structures of the form suggested by Törnqvist, e.g.,

$$y = a\left(\frac{z - c}{z + b}\right) + u \qquad (171)$$

might be considered a better approximation of individual behavior. This choice is usually justified by the fact that functions of this type seem to hold for a range of incomes much wider than that valid for functions of the familiar types discussed in this paper.[18] Some of these alternative forms might raise other types of problems not considered in the proceding sections. This is a result of the fact that the process of aggregation depends on the explicit form of the micro-relationships.

2. To introduce other types of variables in the relationship. One important group of variables is that of prices of the given commodity or group of commodities, and competing prices. The type of aggregation envolved here does not necessarily follow the rules developed for incomes. If prices are the same for all consumers, the problem will be rather trivial. For example if:

$$y_{ijt} = a_{ij} + b_{ij}\, z_{ijt} + c_{ij}\, p_t + u_{ijt} \qquad (175)$$

the group function will include a price term equal to $c_i\, p_t$ where $c_i = \sum_j c_{ij}/N_{it}$. The overall function will include the same variable with a coefficient c equal to the overall average of the c's. But if we assume that the prices prevailing for the different regiones are different, the group functions will have $c_i\, p_{it}$ as a price term. The exact form of the overall aggregate depends on the types of statistical

(18) H. Wold: Demand Analysis. pp.3, 4, 249. John Wiley, 1953.

information available: whether we posses one single price variable for the whole economy; and how is it calculated.

3. Other types of aggregation have not been discussed. Aggregation over commodities might be necessary even when we are dealing with single commodities. For example, it might be necessary that we combine such commodities as maize and millet in one function, knowing that the one is consumed mainly by one group and that the other is consumed by another. Further, there might be need to aggregate the prices of other commodities included as explanatory variables in the micro-function.

4. Another problem arises from the fact that the sizes of groups (N_{it}) are in fact the sizes of total actual and potential consumers. This means that the changes in the sizes of total population do not reflect changes in the sizes of consumers to the same extent in all groups. For example, consumption of wheat might be more widespread in the urban areas than in the rural areas. This means that many of the rural micro-parameters are in fact equal to zero. Unless assumption (A/1), and in particular condition (28) are satisfied, there will be great difficulties in deriving the aggregate function. There is doubt as to the possibility of depending on equations satisfying assumption (A/3), not because of the possibility of differences of structures among consumers, but rather because of the existence of sub-groups, some of them following the same structure, while the remainder are not consumers at all. To illustrate the point, suppose that there are two groups only, and that all consumers in group 1 are consumers, while a certain proportion $\lambda < 1$, of group 2 are consumers. Both types are following similar structures:

$$y_{ijt} = a + b\, z_{ijt} + u_{ijt}$$

where

$$j = 1, \ldots, N_{1t} \text{ for } i = 1, \text{ and } j = 1, \ldots\ldots, \lambda N_{2t}$$

for i = 2. Summation over j in both cases gives:

$$Y_{1t} = a\, N_{1t} + b\, Z_{1t} + U_{1t}$$

$$Y_{2t} = a\, \lambda N_{2t} + b\, Z^{*}_{2t} + U_{2t}$$

Where Z^{*}_{2t} is the aggregate income of consumers in group 2. This means that the remaining members of this group, $(1 - \lambda)\, N_{2t}$ in number, follow another equation, where both a and b are both zero, and their incomes are $(Z_{2t} - Z^{*}_{2t})$. This leads us to a special situation of assumption (A/2). An explicit formulation of the problem should be made before deriving the aggregate function itself.

5. In our treatment of the stochastic disturbances we have confined our analysis to one aspect only, namely the homoscedasticity of their variances. There are still a number of problems that can and should be raised. One problem is the following: The assumptions introduced in section VIII were based on the hypothesis that we possess exact information on certain new variables (e.g., f and x) and on the corresponding relationships. If these assumptions are satisfied only in a stochastic sense, their residuals should be studied, and their effects on the total residuals in the aggregate equation should be discussed. If we allow for such a situation there is the possibility that the effects will not be additive, since they will also affect the expression D_t, appearing in the denominator in many cases. Now many questions might be raised: What will be the nature of the disturbance elements in the aggregate functions? What will be the appropriate method of statistical estimation? Is the degree of approximation introduced through the direct estimation of an aggregate function derived on less rigorous bases larger or smaller than the amount of deviation envolved as a result of applying familiar statistical techniques without accounting for the actual specifications of the random components?.

6. One consequence of the above set of problems is the justification of classifying the aggregate variables into variables containing errors and void of errors. If it is true that the aggregates are only approximations and if it is possible to express their (unknown) virtual values as functions of actual observations, how is this going to affect familiar methods of estimation; e.g., the instrumental variables method?

7. Another problem which might open a new outlook on the general approach towards the construction of macro-models is that of joint dependence. In micro-analysis we can assume that some variables can be classified as exogenous. However, if the commodity whose demand is studied has a significant contribution on incomes, we cannot maintain such as assumption at the aggregate level. Suppose now that the commodity is produced in one region and consumed in both, though mainly in the other. For example wheat is produced in the rural region but mainly consumed in the urban. How are we going to classify the variables. and what will be the appropriate methods of estimating the aggregate (grous and overall) functions?

8. Suppose that the above statistical problems do not have any significance. Suppose also that we did possess information of all micro and macro variables We could, in principles, estimate both the micro-and the macro- equation. What are the relationships between the two sets of estimates? Is there any

loss of information, or any decrease in the efficiency of the estimates? In particular suppose that for acase g = 2 we could obtain estimates of an equation of the form (93); what statements could we make with respect to the parameter $a_1 = \alpha$, or any other parameter derived from the solutions of (90)? How do these statements change from movement from one set to the other, and what is the effect of the "apparent" degrees of freedom? Also what will be the significance of linear restrictions such as those given by (91)?

The above list of problems in only a tentative one, and the careful reader can spot out many others. It helps only to emphasize the complexity of the problem and its repercussions on economic analysis itself. On the other hand it shows that an effective solution has to be obtained within the framework of an econometric rather than a purely economic analysis. Here is an invitation for further study that is extended to colleagues in this Institute and elsewhere, which is hoped to be accepted in order to explore this rather fundamental area.

I would also suggest that some simulation work be done in this field. Constructing a number of micro-structures on the basis of alternative assumptions we can derive the aggregate variables derived on their basis. This has to be done after superimposing sets of values for the random compenents, satisfying certain given distributions. Investigation of the various problems and approaches discussed here will lead us to evaluate the relative importance of each of them. Problems of prediction and predictive statements have also to be investigated, since it is prediction that is the final goal of the analysis. This simulation approach would also help to investigate the problems arising from errors in specification of the underlying economic structures. There might arise cases where one group follows one structure and another group follows another. Or it might happen that the underlying structure is different from the one assumed for empirical analysis. In any case, thinking on these lines would help us to classify the multiple econometric problems according to their importance. This would help us in approaching actual situations with some plausible a prioristic grounds that does not envolve any circularity in the probability statements qualifying our econometric findings.