

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: <http://www.nber.org/books/kmen80-1>

Publication Date: 1980

Chapter Title: Aggregation and Disaggregation of Nonlinear Equations

Chapter Author: Harry H. Kelejian

Chapter URL: <http://www.nber.org/chapters/c11697>

Chapter pages in book: (p. 135 - 152)

# Aggregation and Disaggregation of Nonlinear Equations

*HARRY H. KELEJIAN*

DEPARTMENT OF ECONOMICS  
UNIVERSITY OF MARYLAND  
COLLEGE PARK, MARYLAND

## 1. Introduction

Economic models are often formulated and estimated in terms of aggregated data. In such cases, it is not difficult to show that the aggregated (or macro-) variables do not relate to each other in the same manner as their disaggregated (or micro-) counterparts.<sup>1</sup> Indeed, if the underlying micro-relationships are nonlinear, the very existence of macro-relationships which relate only macro-variables to each other is called into question.<sup>2</sup> Nevertheless, economists frequently concern themselves with issues relating to validation of macro-models, most of which contain nonlinearities.

In this paper we suggest that in many cases it may be reasonable to assume that the independent variables of micro-econometric models are stochastic. If so, macro-models relating aggregated dependent variables to each other and to the aggregated independent variables can be defined in terms of the moments of the joint distribution of the aggregated dependent variables conditional upon the aggregated independent variables. Therefore, it would be reasonable to assume, in this stochastic framework, that macro-econometric models exist. However, our example in the single equation case will demonstrate that these macro-models will typically be very complex unless very restrictive assumptions are made. This raises the

<sup>1</sup> See, e.g., Theil (1954) and Akdeniz & Milliken (1975). For related issues see Orcutt, Watts, & Edwards (1968) and Kelejian (1972).

<sup>2</sup> See, e.g., F. Fisher (1969), Zarembka (1968), and Green (1964).

possibility that macro-models considered in practice are merely approximations to these "true" models, and hence the question is raised as to which approximation may be best for the purpose at hand or for the particular range of values of the variables involved. This also suggests that the issue of validation concerning macro-models should focus on the purpose for which the model is considered, rather than on the correctness of the model specifications. As we will indicate, these two concerns need not coincide even in the absence of cost considerations relating to the operation of the model. However, if one wishes to focus on model specifications in a single equation or systems framework, our results suggest that the objects of analysis should be the functional forms of the exogenous variables and the higher moments of the disturbance terms.

If, in our stochastic framework, one can assume that macro-models exist, the problem of finding conditions under which they can be disaggregated arises. This "inverse" problem is also considered, but only in the context of a single equation or a reduced form equation model. More specifically, conditions are given under which a correspondence is one to one between a micro-equation which is nonlinear in the independent variables and disturbance terms, but identical for all micro-units, and the resulting macro-equation. The problem of using macro-data and a given macro-equation to estimate the implied micro-equation is therefore discussed, and some results presented.<sup>3</sup> However, these results are somewhat academic in the sense that the conditions of the correspondence are (expectedly) strong, and consequently, aside from a few cases, it may not be advisable to infer such micro-behavior in practice. Ironically, our results on inversion seem to drive a wedge between macro-models and revealed micro-behavior.

We begin our analysis in terms of single equation models in Section 2. Specifically, the single equation macro-model is defined in Section 2.1; an illustration along with certain limiting results is given in Section 2.2; the inverse problem is defined and results concerning estimation are given in Section 2.3. Generalizations of the single equation results to the systems framework are given in Section 3, and conclusions are presented in Section 4.

<sup>3</sup> Although there are similarities, the approach taken here is quite different than that taken by Houthakker (1955-1956) in his study of the production function. Essentially, under certain conditions, Houthakker demonstrated that if production within each firm of an industry takes place according to a fixed coefficients production function, but the coefficients vary from firm to firm in a regular but nonstochastic way, a macro-production function will exist which relates total output to total input. Several authors (Levhari, 1968; Sato, 1970) extended Houthakker's results in a variety of ways. However, their results, as well as Houthakker's, were in a non-stochastic framework and directly related to aggregation problems associated with production functions. In addition, they did not consider problems associated with empirical implementation.

## 2. Single Equation Models

### 2.1. THE FRAMEWORK OF ANALYSIS

In this section we define a single equation micro-regression model for each of  $N$  micro-decision units. The independent variables and the disturbance terms are assumed to be stochastic. To keep the presentation as general as possible, the parameters are also assumed to be stochastic. Interpretations for the stochastic regressor assumption are given. A rule for aggregation is then specified, and a macro-function is defined in terms of the conditional distribution of the aggregated dependent variable given the aggregated independent variables. Special cases are considered in the following sections.

#### 2.1.1. Basic Specifications

Consider a microeconomic regression model specified as

$$Y_{it} = K_i(X_{it}, A_{it}, U_{it}), \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where  $Y_{it}$  is the value of the dependent variable at time  $t$  corresponding to the  $i$ th micro unit,  $X_{it}$  is a vector of stochastic independent variables at time  $t$  corresponding to the  $i$ th unit,  $A_{it}$  is a corresponding vector of parameters which may be random, and  $U_{it}$  is, possibly, a vector of disturbance terms. Taking  $A_{it}$ , in general, to be stochastic, we assume that the joint distribution of  $X_{1t}, \dots, X_{Nt}, U_{1t}, \dots, U_{Nt}, A_{1t}, \dots, A_{Nt}$ , and the functions  $K_i(\cdot; \cdot)$  are such that the dependent variables  $Y_{it}, i = 1, \dots, N$ , have finite first moments. The assumption of finite moments is a reasonable one in that most economic variables have finite range.

Consider the assumption that the independent variables are stochastic. Aside from seasonal and time trend dummy variables, the researcher will not generally be able to predict with perfect certainty the future values of the independent variables in a model such as (1). This will be the case whether the independent variables relate to essentially noneconomic events, such as rainfall, temperature, and other environmental characteristics, or to economic characteristics associated with the individual decision units. Therefore, at least in a Bayesian framework, the assumption that the independent variables are stochastic may be a reasonable one.

A variation on a scenario offered by Lipsey (1960) may also be given as an interpretation of the stochastic regressor assumption as it relates to the independent variables describing the economic characteristics of the decision units. For purposes of illustration, assume that all of the independent variables in (1) relate to such economic characteristics and that these variables

are independent of the disturbance terms and of the parameters. Then the scenario is that, in each period, general economic conditions determine certain parameters of the joint distribution of the independent variables (e.g., their means) while the actual values of these variables are in part determined by events particular to each cross sectional unit—e.g., information discrepancies, attitudinal differences, adjustment costs, and various random occurrences. As an example, suppose the amount of labor hired by the  $i$ th unit at time  $t$ , say  $L_{it}$ , is the only such economic variable. Suppose also that the amount hired depends, in part, upon the real wage rate. Then, in this scenario, we would be assuming that the economy determines the wage variable and, in doing so, determines, say, the expected value of  $L_{it}$ . The actual value of this variable, however, will deviate from its mean, and this deviation would be assumed to be stochastically distributed.

Returning to our general model (1), assume that observations on  $X_{it}$  and  $Y_{it}$  are not available. Instead, observations are available only on the aggregates<sup>4</sup>

$$Y_t = \sum_{i=1}^N Y_{it}, \quad X_t = \sum_{i=1}^N X_{it}. \quad (2)$$

Our problem is to define a meaningful (marco-) relationship between  $Y_t$  and  $X_t$ .

Let  $f_{it}(X_{it}, A_{it}, U_{it} | X_t)$  be the conditional density of  $X_{it}$ ,  $A_{it}$ , and  $U_{it}$  given  $X_t$  which is implied by the joint distribution of  $X_{1t}, \dots, X_{Nt}, A_{1t}, \dots, A_{Nt}, U_{1t}, \dots, U_{Nt}$ .<sup>5</sup> Then the macro-function is defined as

$$\begin{aligned} E(Y_t | X_t) &= \sum_{i=1}^N E[K_i(X_{it}, A_{it}, U_{it}) | X_t] \\ &= \sum_{i=1}^N \int_{R_{it}} K_i(X_{it}, A_{it}, U_{it}) f_{it}(X_{it}, A_{it}, U_{it} | X_t) dX_{it} dA_{it} dU_{it} \\ &= \sum_{i=1}^N \Gamma_{it}(X_t) = \Gamma_t(X_t), \end{aligned} \quad (3)$$

where  $R_{it}$  denotes the corresponding region of integration,

$$\Gamma_{it}(X_t) = \int_{R_{it}} K_i(X_{it}, A_{it}, U_{it}) f_{it}(X_{it}, A_{it}, U_{it} | X_t) dX_{it} dA_{it} dU_{it},$$

<sup>4</sup> It will become apparent that, aside from minor modifications, all of the results in this paper, except those dealing with the inverse problem in Section 2.3, will hold for the more general aggregates  $Y'_t = \sum d_{1i} Y_{it}$  and  $X'_t = \sum d_{2i} X_{it}$ . The results in Section 2.3 depend upon the condition  $d_{2i} = 1, i = 1, \dots, N$ .

<sup>5</sup> For ease of presentation we will not distinguish between the random variables and the values they take on.

and, therefore,  $\Gamma_t(X_t)$  is a time function of the elements of  $X_t$ .<sup>6</sup> Clearly, a sufficient condition for the "macro-" function  $\Gamma_t(X_t)$  to be independent of  $t$ ,  $\Gamma_t(X_t) = \Gamma(X_t)$ , is that the variables involved be stationary. However, an example below will demonstrate that stationarity is not a necessary condition for this result. The example will also demonstrate that, even under rather strong assumptions, the macro-function will typically not be tractable.

## 2.2. AN ILLUSTRATION

In this section an example is given in which a macro-function is calculated from the specifications of the micro-model. In terms of the results obtained, the nature of some of the complications concerning macro-model specifications is indicated. In addition, a certain limiting case is considered in which one would expect the macro-function to simplify. It is shown that this case corresponds to an unconditional expectations approach to macro-model formulation which has been suggested by Ando (1971) and Goldfeld & Quandt (1976). This correspondence is shown not to depend upon the particular example considered.

A mathematically "convenient" form of the micro-model (1) is specified in Section 2.2.1. The corresponding macro-model is calculated in Section 2.2.2, and the limiting results are derived in Section 2.2.3.

### 2.2.1. The Model

Consider the following specialization of the micro-model (1)

$$Y_{it} = a_i + b_i X_{it} + c_i \exp(d_i X_{it} + U_{it}), \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (4)$$

where  $Y_{it}$ ,  $X_{it}$ , and  $U_{it}$  are as defined in (1), but  $X_{it}$  and  $U_{it}$  are now assumed to be scalars, and  $a_i$ ,  $b_i$ ,  $c_i$ , and  $d_i$  are nonstochastic parameters whose values vary with  $i$  but not  $t$ . Assume that  $U_{it}$  is normally distributed with mean zero,  $E(U_{it}) = 0$ , constant variance,  $E(U_{it}^2) = \sigma_u^2$ , and cross moments  $E(U_{it} U_{jt}) = 0$ , for  $t \neq s$ , and/or  $i \neq j$ . Assume also that the independent variable  $X_{it}$  is normally distributed with moments  $E(X_{it}) = \eta_i$ , and  $E[(X_{it} - \eta_i)(X_{jt} - \eta_j)] = \sigma_{ij}$ , so that  $\sigma_{ii}$  denotes the variance of  $X_{it}$ . Finally, assume that  $X_{it}$  is independent of  $U_{jt}$  for all  $i$  and  $j$ .

Let the aggregates  $Y_t$  and  $X_t$  be defined as in (2). Then our assumptions concerning  $X_{it}$  imply that  $X_t$  is normally distributed with mean  $\mu_X = \sum_i \eta_i$ ,

<sup>6</sup> Chipman (1975, 1976) has also noted that a macro-relationship can be defined in this way and that the resulting relationship provides a best approximation to the micro-model in terms of mean squared errors. For a related study see Zellner (1969).

and variance  $\sigma_X^2 = I'VI$ , when  $I'$  is the  $1 \times N$  row vector of unit elements,  $I' = (1, 1, \dots, 1)$ , and  $V$  is the variance-covariance matrix of  $X_1, \dots, X_N$ . Letting  $C_i$  denote the covariance of  $X_t$  and  $X_{it}$ , it is easily shown that

$$C_i = \sum_{j=1}^N \sigma_{ji}. \quad (5)$$

It follows that the conditional distribution of  $X_{it}$  given  $X_t$  is normal with mean

$$\begin{aligned} E(X_{it}|X_t) &= \eta_i + (C_i/\sigma_X^2)(X_t - \mu_X) \\ &= B_{1i} + B_{2i}X_t, \end{aligned} \quad (6)$$

where  $B_{1i} = \eta_i - \mu_X(C_i/\sigma_X^2)$ , and  $B_{2i} = C_i/\sigma_X^2$ . The conditional variance of  $X_{it}$  given  $X_t$  is

$$\text{var}(X_{it}|X_t) = \sigma_{ii}(1 - R_i^2), \quad (7)$$

where  $R_i^2 = (C_i^2/\sigma_X^2\sigma_{ii})$ .

### 2.2.2. The Macro-Function

We are now in a position to calculate the macro-function. Specifically, (2), (4), and our assumption that  $X_{it}$  and  $U_{jt}$  are independent for all  $i$  and  $j$  imply that

$$E(Y_t|X_t) = \sum_{i=1}^N [a_i + b_i E(X_{it}|X_t) + c_i E[\exp(d_i X_{it})|X_t] E[\exp(U_{it})]]. \quad (8)$$

Since  $U_{it}$  is normally distributed, and the conditional distribution of  $X_{it}$  given  $X_t$  is normal, (8) can be evaluated by using (6), (7), and a standard result in the literature concerning the expectation of an exponential involving a normal variable.<sup>7</sup> Specifically,

$$\begin{aligned} E(Y_t|X_t) &= A_1 + A_2 X_t + \sum_{i=1}^N q_i \exp(K_i X_t) \\ &= g(X_t), \end{aligned} \quad (9)$$

where  $A_1 = \sum_{i=1}^N [a_i + b_i B_{1i}]$ ,  $A_2 = \sum_{i=1}^N b_i B_{2i}$ ,  $q_i = c_i \exp(d_i B_{1i} + \frac{1}{2} d_i^2 \text{var}(X_{it}|X_t) + \sigma_{ii}^2/2)$ , and  $K_i = d_i B_{2i}$ . Thus the mean relationship for  $Y_t$  is a function of  $X_t$ , and so a macro-function exists. However, it is evident from (9) that the macro-function will typically not be tractable because the exact specification will involve a functional form that is "unusual," and there are more parameters than typically assumed. Indeed, an explicit analytical

<sup>7</sup> Let  $z$  be normal with mean  $\mu_z$  and variance  $\sigma_z^2$ . Then  $E(e^{az}) = \exp(a\mu_z + \frac{1}{2}a^2\sigma_z^2)$ —see Mood, Graybill, & Boes (1974, p. 165).

form of the macro-function may not exist for certain distributional and functional form assumptions concerning the micro-relationship. This suggests that if macro-models are viewed as aggregations of micro-relationships, they should be viewed as approximations, rather than as exact specifications.

One can argue that in some cases the approximations may be "close", and so a concern with macro-model validation via its specifications is appropriate in that gross violations will be detected. Abstracting from issues relating to cost of operation and model maintenance, the implicit assumption underlying such an argument is that, in the typical uses to which models are put, models which are correctly specified will "outperform" models which are not. This assumption, however, will generally not hold for models such as those typically considered in practice for at least one of the major model uses, namely forecasting. As an illustration, suppose the conditional mean of a dependent variable,  $z_t$ , given an independent variable,  $w_t$ , is  $h(w_t; b)$ , where  $b$  is a parameter, and  $h(w_t; b)$  is a function which is nonlinear in  $b$ .<sup>8</sup> Then it is easily shown that  $h(w_t; b)$  is the minimum mean squared error predictor of  $z_t$  based on  $w_t$ . In practice, however, the parameters of a model are typically assumed not to be known, and so a parameter such as  $b$  would have to be estimated. Let  $\hat{b}$  be any estimator of  $b$ . Then, unfortunately, the predictor which would be suggested by the "true" model, namely  $h(w_t; \hat{b})$ , cannot also be assumed to have such optimal forecasting properties. Therefore, other predictors exist, which correspond to incorrectly specified models and which have mean squared errors of forecasts which are less than that corresponding to the model forecast,  $h(w_t; \hat{b})$ . Indeed, Kelejian & Vavrichek (1978) give an example in which the mean squared error of forecast based on a linear approximation to a "true" nonlinear model is less than or equal to that of the nonlinear model whose parameters have been estimated by a full information, maximum likelihood technique, for all values of the independent variable considered. This result and the results above concerning the complex nature of the exact specifications of macro-models suggest, at least for the case of forecasting, that macro-models be validated (or evaluated) in terms of their purpose rather than in terms of the exactness of their specifications.

Finally, we point out a similarity between the macro-model (9) and the micro-model (4). Note from (4) that the mean micro-relationship can be expressed as

$$E(Y_{it} | X_{it}) = a_i + b_i X_{it} + r_i \exp(d_i X_{it}), \quad (10)$$

<sup>8</sup> This assumption of nonlinearity is reasonable if one recognizes that many macro-models are simultaneous equation systems which have reduced forms that are nonlinear in the parameters. In addition to this, many macro-models contain nonlinearities in the endogenous variables, which serve to compound the difficulties—see, e.g., Goldfeld & Quandt (1972, Chapter 8).



where  $r_i = c_i \exp(\sigma_u^2/2)$ . From the first line of (9) note that the mean macro-relationship can be expressed as

$$E(Y_i | X_i) = \sum_{i=1}^N [A_{1i} + A_{2i}X_i + q_i \exp(K_i X_i)], \quad (11)$$

where  $A_{1i}$  and  $A_{2i}$  are the  $i$ th terms in the summation defining  $A_1$  and  $A_2$ :  $A_{1i} = a_i + b_i B_{1i}$ ,  $A_{2i} = b_i B_{2i}$ . In comparing (10) and (11) we see that the mean macro-relationship may be interpreted as a sum of mean micro-relationships which differ only in their parameter values and which are *all evaluated* at  $X_i$ .<sup>9</sup>

### 2.2.3. A Limiting Case

Consider again the micro-model (4), but assume, in addition to our other assumptions, that  $a_i = a$ ,  $b_i = b$ ,  $c_i = c$ , and  $d_i = d$  for all  $i$  so that the non-linear function is identical over the micro-units; in addition, assume that  $\sigma_{ij} = 0$  for  $i \neq j$ , and  $\eta_i = \eta$  for all  $i$  so that  $X_{it}$  and  $X_{jt}$  are independent for all  $i \neq j$  and have the same distribution. Finally, assume that the macro-variables of interest are the sample averages  $\bar{Y}_t = Y_t/N$ ,  $\bar{X}_t = X_t/N$ —see (2).

Under these assumptions it is not difficult to show that, for any  $i$ , the macro-function is

$$\begin{aligned} E[\bar{Y}_t | \bar{X}_t] &= E[Y_{it} | \bar{X}_t] \\ &= a + b\bar{X}_t + p_n \exp(d\bar{X}_t) \equiv G_n(\bar{X}_t), \end{aligned} \quad (12)$$

where  $p_n = c \exp(\frac{1}{2}d^2 \sigma_{ii}^2(1 - N^{-1}) + \sigma_u^2/2)$ , and where  $G_n(\bar{X}_t)$  is defined by the second line of (12).<sup>10</sup> Thus, in this case the macro-function  $G_n(\bar{X}_t)$  simplifies and is in the same parametric family as the *mean* micro-function.

If we now assume that  $N \rightarrow \infty$ , it is not difficult to show that  $\text{Plim } \bar{X}_t = \eta$ , and  $\text{Plim } G_n(\bar{X}_t) = G(\eta)$ , where

$$G(\eta) = a + b\eta + c \exp(d\eta + \frac{1}{2}d^2 \sigma_{ii}^2 + \sigma_u^2/2). \quad (13)$$

Similarly, it can be shown that

$$E(\bar{Y}_t) = G(\eta). \quad (14)$$

That is, in this limiting case, the macro-function can be obtained by simply calculating the *unconditional* expectation of  $\bar{Y}_t$  and then equating the sample average  $\bar{X}_t$  to the corresponding moment  $\eta$ .

<sup>9</sup> That is, (11) can be obtained from (10) by replacing  $a_i$ ,  $b_i$ ,  $r_i$ ,  $d_i$  and  $X_{it}$ , by, respectively,  $A_{1i}$ ,  $A_{2i}$ ,  $q_i$ ,  $K_i$ , and  $X_i$ , and then summing.

<sup>10</sup> For a result which is similar to (12) but which is derived in a different setting and under different assumptions see Ramsey (1972).

This result generalizes. For instance, consider again the general model (1), but assume it is identical for all of the macro-units. Then a retracing of the steps leading to (14) will demonstrate that the macro-function relating the sample average  $\bar{Y}_t$  to the elements of  $\bar{X}_t$  can be obtained in terms of the *unconditional* expectation if the vectors  $X_{it}$ ,  $A_{it}$ , and  $U_{it}$  are identically and independently distributed over the micro-units, and the number of micro-units increases beyond limit,  $N \rightarrow \infty$ . Although their framework was different, this unconditional expectations approach was taken by Ando (1971) and then by Goldfeld & Quandt (1976) in their studies of aggregation concerning switching models.

One more point concerning the example in Section 2.2.1 should be noted. The assumptions we made concerning the joint distribution of  $X_{1t}, \dots, X_{Nt}$  imply that the conditional distribution of  $X_{it}$  given  $\bar{X}_t$  has parameters that depend only on the given value of  $\bar{X}_t$ ,  $\sigma_{it}$ , and  $N$ . Therefore, the result in (12) that  $G_N(\bar{X}_t)$  does not explicitly involve  $t$  holds regardless of our assumptions concerning the unconditional mean of  $X_{it}$ , e.g.,  $\eta$  could be a time function. In other words, the macro-function can be stationary in the sense that its form and parameters do not vary over time even if the micro-independent variables are nonstationary.

### 2.3. THE INVERSE PROBLEM

In this section the problem of inferring a mean micro-function corresponding to an assumed macro-function, which may be nonlinear, is defined, and some results concerning estimation are given. However, as one would expect for such purposes, the assumptions of the model are strong, and consequently, one may not wish to "invert" many of the macro-models considered in practice. Nevertheless, the results of this section suggest that the issue of inversion concerning nonlinear econometric macro-models is a meaningful one, and one which could be considered in certain restricted cases.

The specifications of the model are given in Section 2.3.1; the inverse problem is formally defined in Section 2.3.2; selected problems concerning estimation are discussed in Sections 2.3.3 and 2.3.4. Finally, an example illustrating some of the principles is given in Section 2.3.5.

#### 2.3.1. Basic Specifications

Consider the following variation of the micro-model (1):

$$Y_{it} = \sum_{j=1}^K G_j(X_{it})H_j(U_{jit}), \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (15)$$

where  $Y_{it}$  and  $X_{it}$  (a vector) denote the values at  $t$  of the dependent and independent variables corresponding to the  $i$ th micro unit,  $U_{jit}$ ,  $j = 1, \dots, K$ , are disturbance terms, and  $G_j(\cdot)$  and  $H_j(\cdot)$ ,  $j = 1, \dots, K$ , may, in general, be nonlinear functions of their respective arguments.<sup>11</sup> As in previous sections we again assume that data are available only on the macro-variables,  $Y_t$  and  $X_t$ , defined in (2).

Assume that  $U_{jit}$  is identically and independently distributed over the micro-units, that it is independent of  $X_{st}$  for all  $j, i, t$ , and  $s$ , and that

$$E[H_j(U_{jit})] = \alpha_j. \quad (16)$$

Also assume that  $X_{it}$  is identically and independently distributed over the micro-units and that its distribution is *known, including its parameters*. Finally, denote the conditional density of  $X_{it}$  given  $X_t$ , which may be non-stationary, as  $f_t(X_{it}|X_t)$ .

The assumptions above essentially imply that the micro-units are identical to each other except for random deviations in the values of the variables involved. Furthermore, because these random deviations are assumed to be independent of each other, the behavior of each micro-unit is independent of that of all the others.<sup>12</sup> Although this framework may not strictly hold for any of the aggregated models considered by economists, it may be a reasonable approximation for some relationships for certain subgroups of the population, e.g., the yield equation for small farmers as a function of their factor input decisions.

### 2.3.2. The Problem Defined

The assumptions above imply that the conditional distribution of  $X_{it}$  given  $X_t$  is independent of  $i$ . In light of this, the macro-function relating  $Y_t$  to the elements of  $X_t$  can be expressed for any  $i$  as

$$\begin{aligned} E(Y_t|X_t) &= \sum_{i=1}^N E \left[ \sum_{j=1}^K G_j(X_{it}) \alpha_j | X_t \right] \\ &= \int_R N \left[ \sum_{j=1}^K \alpha_j G_j(X_{it}) \right] f_t(X_{it}|X_t) dX_{it} \\ &= D_t(X_t), \end{aligned} \quad (17)$$

<sup>11</sup> Note that a linear model is a special case of (15).

<sup>12</sup> It is interesting to note that the assumption of known parameters concerning the distribution of  $X_{it}$  is not quite as restrictive as one might first expect. For example, if the form of the distribution of  $X_{it}$  is specified (e.g., normality), its parameters can be estimated in terms of a time series of observations on the macro-variables  $X_t$  if  $N$  is *known*. If  $N$  is not known, the results can be given parametrically in terms of  $N$ . For example, since  $X_t = \sum_i X_{it}$ , the mean value of each element of  $X_t$  will be  $N$  times the mean value of the corresponding element of  $X_{it}$ . A similar conclusion holds with respect to the higher moments.

where  $R$  denotes the region of integration, and  $D_t(X_t)$  is the mean macro-function.

Now let

$$Q(X_{it}) = N \sum_{j=1}^K \alpha_j G_j(X_{it}); \tag{18}$$

that is, from (15) and (16),  $Q(X_{it})$  is proportional to the mean micro-function, where the proportionality factor is the number of micro-units,  $N$ . Then from (17) we have

$$\int_R Q(X_{it}) f_t(X_{it} | X_t) dX_{it} = D_t(X_t). \tag{19}$$

The inverse problem is then defined to be the problem of inferring  $Q(X_{it})$  from a knowledge of the conditional distribution,  $f_t(X_{it} | X_t)$ , and the macro-function,  $D_t(X_t)$ . That is, given the conditional density and the macro-function, can a function be found whose first moment is  $D_t(X_t)$  for all values of the elements of  $X_t$ ?<sup>13</sup>

The elements of  $X_t$  can be viewed as parameters of the conditional density  $f_t(X_{it} | X_t)$ . It then follows from (19) that  $f_t(X_{it} | X_t)$  and  $D_t(X_t)$  will uniquely determine  $Q(X_{it})$  if  $f_t(X_{it} | X_t)$  is complete with respect to the elements of  $X_t$ .<sup>14</sup> For instance, assume that  $f_t(X_{it} | X_t)$  is of the form

$$f_t(X_{it} | X_t) \propto \exp(A_1(X_{it}, t) + A_2(X_t, t) + X_t' C X_{it}), \tag{20}$$

where  $\propto$  denotes proportionality,  $A_1(X_{it}, t)$  and  $A_2(X_t, t)$  are known functions of the elements of  $X_{it}$  and  $X_t$  which may involve  $t$ ,  $C$  is a square matrix of known constants, and the range of possible values assumed by  $X_{it}$  does not involve  $X_t$ .<sup>15</sup> Then if  $Q_1(X_{it})$  and  $Q_2(X_{it})$  satisfy (19), we must have

$$\int_R (Q_1(X_{it}) - Q_2(X_{it})) f_t(X_{it} | X_t) dX_{it} = 0, \tag{21}$$

for all  $X_t$ . Let  $Q_3(X_{it}) = Q_1(X_{it}) - Q_2(X_{it})$ . Then, from (20) and (21), we have

$$\int_R [Q_3(X_{it}) \exp(A_1(X_{it}, t))] \exp[X_t' C X_{it}] dX_{it} = M(\delta_t) = 0, \tag{22}$$

where  $\delta_t = X_t' C$ , and  $M(\delta_t)$  is the Laplace transform of the product of the functions in brackets in (22). Since the transform is unique,  $M(\delta_t) = 0$  for all  $X_t$  implies  $Q_3(X_{it}) \exp(A_1(X_{it}, t)) \equiv 0$ , which in turn implies  $Q_3(X_{it}) \equiv 0$ .

<sup>13</sup> As an illustration, if the macro-function is known to be of the constant elasticity of substitution (CES) form, what must the corresponding micro-function be?

<sup>14</sup> For a discussion of completeness see Kendall & Stuart (1961, pp. 190-195). The following discussion is an adaptation of their material and is given here because all readers may not be familiar with the concepts involved.

<sup>15</sup> Clearly, a special case of (20) is the multivariate normal.

### 2.3.3. Estimation

If  $f_i(X_{it}|X_t)$  is complete,  $Q(X_{it})$  in (19) is unique and therefore identified. Computationally,  $Q(X_{it})$  can be obtained exactly if it can be specified parametrically. If the functional form of  $Q(X_{it})$  is not known, an approximation to  $Q(X_{it})$  can be obtained—i.e., the approximation will be meaningful since  $Q(X_{it})$  is identified.

To see this, suppose, first, that the functional form of  $Q(X_{it})$  is known and contains  $\Lambda$  parameters denoted by the vector  $B$ . To make this dependence on the vector  $B$  explicit, we rewrite  $Q(X_{it})$  as  $Q(X_{it}; B)$ . Now let

$$\Pi_t(X_t; B) = \int_R Q(X_{it}; B) f_i(X_{it}|X_t) dX_{it}. \quad (23)$$

and note that, since  $f_i(X_{it}|X_t)$  is assumed to be known, the value of  $\Pi_t(X_t; B)$  can, in principle, be calculated for an assumed value of the vector  $B$ . Then from (19) we have, for the true value of the vector  $B$ ,

$$\Pi_t(X_t; B) = D_t(X_t). \quad (24)$$

Suppose now that we have a sample of size  $T \geq \Lambda$  on the elements of  $X_t$ ,  $t = 1, \dots, T$ , and we know the macro-function  $D_t(X_t)$ —including its parameters. Then, from (24), these observations may be expressed as

$$\Pi(X, B) = D(X), \quad (25)$$

where  $\Pi(X, B)$  and  $D(X)$  are  $T \times 1$  vectors whose  $t$ th elements are, respectively,  $\Pi_t(X_t; B)$  and  $D_t(X_t)$ . It follows from (25) that we may obtain  $B$  in terms of  $X_t$ ,  $t = 1, \dots, T$ , if  $\text{rank}(\partial \Pi'(X; B)/\partial B) = \Lambda$ . This condition holds, in general, if  $f_i(X_{it}|X_t)$  is complete, otherwise the solution to (19) would not be unique.

### 2.3.4. An Approximation

The calculations involved in solving (25) may be extensive. Therefore, consider the problem of estimating a polynomial approximation to  $Q(X_{it})$ . For ease of presentation assume that  $X_{it}$  is a scalar. Then assume

$$Q(X_{it}) \doteq \sum_{j=0}^r b_j X_{it}^j. \quad (26)$$

Substituting (26) into (19) we have

$$\begin{aligned} \int_R \sum_{j=0}^r b_j X_{it}^j f_i(X_{it}|X_t) dX_{it} &= \sum_{j=0}^r b_j \int_R X_{it}^j f_i(X_{it}|X_t) dX_{it} \\ &= \sum_{j=0}^r b_j \Pi_{jt}(X_t) \doteq D_t(X_t) \end{aligned} \quad (27)$$

where

$$\Pi_{jt}(X_t) = \int_R X_{it}^j f_t(X_{it}|X_t) dX_{it}.$$

It follows from the last two lines of (27) that, given  $D_t(X_t)$ , the  $r + 1$  parameters,  $b_j, j = 0, \dots, r$ , can be estimated from a time series of observations on  $X_t, t = 1, \dots, T$ , where  $T \geq r + 1$  if  $\Pi_{0t}(X_t), \dots, \Pi_{rt}(X_t)$  are linearly independent. For example, the  $T$  observations on  $X_t$  can be expressed as

$$\Pi b \doteq D \tag{28}$$

where  $\Pi$  is the  $T \times r + 1$  matrix whose  $t, j$  element is  $\Pi_{jt}(X_t)$ ,  $b' = (b_0 \cdots b_r)$ , and  $D' = (D_1(X_1) \cdots D_T(X_T))$ . Again, if  $f_t(X_{it}|X_t)$  is complete, one would expect the rank of  $\Pi$  to be  $r + 1$ . Thus from (28) we obtain our estimate of  $b, \hat{b}$ , as

$$\hat{b} = (\Pi' \Pi)^{-1} \Pi' D. \tag{29}$$

It should be clear from (27) that the results of this polynomial approach will only be approximate unless the macro-function  $D_t(X_t)$  is expressible as

$$D_t(X_t) = \sum_{j=0}^r b_j E(X_{it}^j | X_t) \tag{30}$$

where  $E(X_{it}^j | X_t)$  is the conditional mean of  $X_{it}^j$ —that is, unless  $D_t(X_t)$  is expressible as a linear combination of the (linearly independent) conditional moments of  $X_{it}$ . The similarity of (30) to a polynomial approximation of a function is evident.

2.3.5. *An Example*

In order to illustrate in a simple manner some of the concepts in this section consider a multivariate generalization of our earlier example in Section 2.2. Specifically, in the micro-model (4) consider the case in which  $X_{it}$  is an  $M \times 1$  vector,  $b_i = b$  and  $d_i = d$ , where  $b$  and  $d$  are  $1 \times M$  vectors, and  $a_i = a$  and  $c_i = c$ , where  $a$  and  $c$  are scalars. Assume that  $X_{it}$  is identically and independently distributed over the micro-units, with mean vector  $EX_{it} = \eta$ , and variance-covariance matrix  $E(X_{it} - \eta)(X_{it} - \eta)' = V_x$  for  $i = 1, \dots, N$ . Again take  $X_t = \sum_{i=1}^N X_{it}$ . Then the conditional distribution of  $X_{it}$  given  $X_t$  is readily shown to be normal with mean vector  $E(X_{it}|X_t) = (1/N)X_t$ , and variance-covariance matrix  $V_* = V_x(1 - 1/N)$ .

Under these conditions,  $\Pi_t(X_t; B)$  in (23) would be

$$\Pi_t(X_t; B) = \gamma_1 + \gamma_2 X_t + \gamma_3 \exp((1/N)dX_t + \frac{1}{2}dV_*d'), \tag{31}$$

where  $\gamma_1 = Na$ ,  $\gamma_2 = b$ ,  $\gamma_3 = c \exp(\sigma_u^2/2)$ , and where  $B'$  would now denote the  $2M + 2$  vector  $(\gamma_1, \gamma_2, \gamma_3, d)$ . In (31) let  $Z_t = \exp((1/N)dX_t + \frac{1}{2}dV_*d')$ .

Then the rank condition discussed in reference to (25) will be satisfied since the  $t$ th row of  $\partial\Pi(X; B)/\partial B'$  would be

$$\partial\Pi_t(X_t; B)/\partial B' = (1, X_t', Z_t, \gamma_3(\partial Z_t/\partial d)), \quad (32)$$

where

$$\partial Z_t/\partial d = Z_t[(1/N)X_t' + dV_*].$$

Now consider the moments of the polynomial approximation to  $Q(X_{it})$  as discussed in reference to (26) and (27). In particular, suppose the approximation is

$$P(X_{it}) = A_1'X_{it} + X_{it}'A_2X_{it}, \quad (33)$$

where  $A_2$  is symmetric and nonsingular. Then the conditional moments corresponding to the terms  $\Pi_{jt}(X_t)$  in (27) are

$$E[P(X_{it})|X_t] = (1/N)A_1'X_t + (1/N^2)X_t'A_2X_t + \alpha, \quad (34)$$

where  $\alpha = \text{Trace}(A_2V_*)$ . Clearly, thinking of (34) as a regression model, its  $1 + M + \frac{1}{2}M(M + 1)$  regressors are linearly independent so that the rank of  $\Pi$  in (28) would be full.

### 3. A Systems Generalization

In this section we specify a simultaneous equation model which relates to the micro decision units. We then define the aggregated endogenous and exogenous variables, and use the structural equation approach<sup>16</sup> to specify a simultaneous equation model relating the aggregated variables to each other. We discuss the implications of our results for model "validation." Finally, an inverse result is given which relates to reduced form equations.

The models are specified and their interpretations are given in Section 3.1. The inverse result is given in Section 3.2.

#### 3.1. AGGREGATION

Consider the following  $M$  equation micro-system

$$F_i(Y_{it}, X_{it}) = U_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (35)$$

where  $Y_{it}$  is the  $M \times 1$  vector of endogenous variables at time  $t$  corresponding to the  $i$ th micro-unit,  $X_{it}$  is the corresponding  $L \times 1$  vector of stochastic

<sup>16</sup> A nice presentation of this approach is given by Goldberger (1964, pp. 383–387).

exogenous variables,  $U_{it}$  is the  $M \times 1$  vector of disturbance terms, and  $F_i$  is an  $M \times 1$  vector of, possibly, nonlinear functions in the elements of  $Y_{it}$  and  $X_{it}$ . In order to avoid unnecessary complications, assume that the joint distribution of  $X_{it}$  and  $U_{it}$  is stationary.

As before, define the aggregated variables as the unweighted sums

$$Y_t = \sum_{i=1}^N Y_{it}, \quad X_t = \sum_{i=1}^N X_{it}. \quad (36)$$

Then the conditional density of  $Y_t$  given  $X_t$ , say  $f_1(Y_t|X_t)$ , is determined by the equations of (35) and the joint distribution of the vectors  $X_{1t}, \dots, X_{Nt}$ ,  $U_{1t}, \dots, U_{Nt}$ . Taking the structural equation approach to model formulation and assuming (very reasonably) finite moments, we have

$$E[B_1 Y_t + B_2 G_t | X_t] = \int_{R_1} (B_1 Y_t + B_2 G_t) f_1(Y_t | X_t) dY_t = H(X_t), \quad (37)$$

where  $R_1$  denotes the region of integration,  $B_1$  and  $B_2$  are, respectively,  $M \times M$  and  $M \times M_1$  matrices of parameters,  $G_t = G(Y_t, X_t)$  is an  $M_1 \times 1$  vector of nonlinear functions of the elements of  $Y_t$  which may also contain  $X_t$ , and  $H(X_t)$  is an  $M \times 1$  vector of functions of the elements of  $X_t$ . In this approach the structure of the matrices,  $B_1$  and  $B_2$ , in terms of zero and nonzero elements, and the element functions of  $G_t$  would be taken in a manner suggested by economic theory. However, subject to the finite moment assumption, the result in (37) holds for *every* set of assumed interrelationships as described by  $B_1$ ,  $B_2$ , and  $G_t$ .

Equation (37) implies that a macro-model may be specified as

$$B_1 Y_t + B_2 G_t = H(X_t) + \Psi_t, \quad (38)$$

where  $E(\Psi_t | X_t) = 0$ . The model in (38) is a typical one with two exceptions. The first is that the elements of  $H(X_t)$  will generally not be linear in  $X_t$ , nor will they be functions whose coefficients are known since these coefficients will, in part, depend upon  $B_1$ ,  $B_2$ , and the moments of the joint distribution of  $U_{it}, \dots, U_{Nt}$ . Second, assuming they are finite, the second and higher conditional moments of the macro-disturbance vector  $\Psi_t$  given  $X_t$  will generally involve  $X_t$ . Therefore,  $\Psi_t$  will not in general be homoscedastic.

The implications of these results for model "validation" are straight forward. In practice the joint specification of the economic interrelationships as described by  $B_1$ ,  $B_2$ , and  $G_t$ , need not be consistent with the specification of the exogenous vector of variables, nor with the assumed higher moments of the disturbance vector,  $\Psi_t$ . Therefore, questions of model specification should relate to the functional forms of the exogenous variables, and to the higher moments of the disturbance vector *conditional* upon the assumed specifications of  $B_1$ ,  $B_2$ , and  $G_t$ . In our framework problems concerning



the mutual consistency of the assumed specifications of  $B_1$ ,  $B_2$ , and  $G_t$  do not arise.<sup>17</sup> The reason for this is that, subject to the finite moments assumption, there exists a macro-model corresponding to every joint specification of  $B_1$ ,  $B_2$ , and  $G_t$ .

### 3.2. AN INVERSE RESULT

Assume that the micro-system in (35) does not explicitly involve  $i$  and that  $X_{it}$  and  $U_{it}$  are independent for all  $i$  and  $j$ , and both are identically and independently distributed over the micro-units. Assume also that (35) uniquely defines  $Y_{it}$  in terms of  $X_{it}$  and  $U_{it}$  for all relevant values of the variables involved. Express this dependence as  $Y_{it} = Z(X_{it}, U_{it})$ , where  $Z(X_{it}, U_{it})$  is an  $M \times 1$  vector of functions of  $X_{it}$  and  $U_{it}$ . Then the conditional means of the elements of  $Y_{it}$  with respect to  $X_{it}$  is given by the vector  $E(Y_{it}|X_{it}) = K_1(X_{it})$ , where

$$K_1(X_{it}) = \int_{R_2} Z(X_{it}, U_{it}) f_2(U_{it}) dU_{it}, \quad (39)$$

where  $f_2$  represents the joint density of  $U_{it}$ , and  $R_2$  denotes the region of integration.

Consider, now, the problem of inferring  $K_1(X_{it})$  from an assumed knowledge of the macro-system in (38). Specifically, suppose the macro-system in (38) is known, including the distribution of the disturbance terms. Then, assuming that the parameters are also known (or have been consistently estimated), the macro-model could, in principle, be used to determine the "macro-" conditional mean vector,  $E(Y_t|X_t) = K_2(X_t)$ , where  $K_2(X_t)$  is an  $M \times 1$  vector of functions of  $X_t$ .<sup>18</sup> Then, from (36) and (39), we have

$$\begin{aligned} K_2(X_t) &= \sum_{i=1}^N E[Z(X_{it}, U_{it})|X_t] \\ &= \sum_{i=1}^N \int_{R_3} \int_{R_2} Z(X_{it}, U_{it}) f_2(U_{it}) f_3(X_{it}|X_t) dU_{it} dX_{it} \\ &= N \int_{R_3} K_1(X_{it}) f_3(X_{it}|X_t) dX_{it}, \end{aligned} \quad (40)$$

<sup>17</sup> Variations on this scheme are possible but not recommended. As an example, one could focus questions of specifications on the matrix  $B_1$  conditional on assumed specifications of  $B_2$ ,  $G_t$ , the exogenous variables, and the disturbance terms. However, in practice this formulation would imply, among other things, that the researcher is willing to commit himself to the functional form specifications of the exogenous variables but not to the economic interrelationships involving the endogenous variables.

<sup>18</sup> One method for doing this would be stochastic simulation—see Howrey & Kelejian (1970); another method would be the one suggested by Kelejian & Madan (1977).

where  $f_3(X_{it}|X_t)$  is the conditional density of  $X_{it}$  given  $X_t$ , and  $R_3$  is the corresponding region of integration. Clearly, our discussion in Section 2.3 implies that  $K_1(X_{it})$  is identified if this conditional density is complete and its parameters along with  $N$  are known.

#### 4. Conclusions

We have argued that the issue with macro-econometric models is not their existence, but their complexity and their uniqueness. Therefore, we have suggested that researchers accept the approximate nature of their models, and evaluate them in terms of their usefulness. In this respect we have indicated that the usefulness of a model need not relate to the exactness of its specifications. However, if attention is to focus on model specifications, our results suggest that the object of analysis should be the functional forms of the exogenous variables and the higher moments of the disturbance terms.

We have also shown that under certain conditions it is possible to infer a nonlinear micro-function from macro-data and the corresponding macro-function. However, the conditions are strong, and so we suggested that in a majority of cases such inferences should not be made. As a corollary we suggest that hypotheses which relate to microeconomic behavior may be of limited value for purposes of macro-model specification. However, to end on a positive note, we recall that the usefulness of a macro-model need not relate to the exactness of its specifications.

#### ACKNOWLEDGMENTS

I would like to thank, but not implicate, D. Bradford, Z. Griliches, S. Hymans, J. Kmenta, M. Olson, J. Ramsey, and B. Vavricheck for helpful comments on an earlier draft of this paper. I would also like to thank, but again not implicate, Barbara Bergmann of Maryland for numerous discussions concerning issues in aggregation in which the problem of inferring the "micro-scenario" behind a macro-phenomena was often raised.

#### REFERENCES

- Akdeniz, F., & Milliken, G. The relationship between macro and micro parameters. *International Economic Review*, 1975, 16, 511-515.
- Ando, Albert. On a problem of aggregation. *International Economic Review*, 1971, 12, 306-310.
- Chipman, J. S. The aggregation problem in econometrics. *Advances in Applied Probability*, 1975, 7, 72-83.
- Chipman, J. S. Estimation and aggregation in econometrics. In M. Z. Nashed (Ed.), *Generalized inverses and applications*. New York: Academic Press, 1976. Pp. 549-768.
- Fisher, F. The existence of aggregate production functions. *Econometrica*, 1969, 37, 553-577.

- Goldberger, A. S. *Econometric theory*. New York: Wiley, 1964.
- Goldfeld, S., & Quandt, R. *Nonlinear methods in econometrics*. Amsterdam: North-Holland Publ., 1972.
- Goldfeld, S., & Quandt, R. Aggregation and structural shifts. In S. Goldfeld and R. Quandt (Eds.) *Studies in nonlinear estimation*. Cambridge, Mass.: Ballinger Publishing Co., 1976. Pp. 77-96.
- Green, H. A. J. *Aggregation in economic analysis, an introductory survey*, Princeton, N. J.: Princeton Univ. Press, 1964.
- Houthakker, H. The Pareto distribution and the Cobb-Douglas production function in Activity analysis. *Review of Economic Studies*, 1955-1956, **23**, 27-31.
- Howrey, E. P., & Kelejian, H. H. Dynamic econometric models: Simulation versus analytical solutions. In T. Naylor (Ed.), *Computer simulation experiments with models of economic systems*. New York: Wiley, 1970. Pp. 299-319.
- Kelejian, H. H. Information lost in aggregation: A Bayesian approach. *Econometrica*, 1972, **40**, 19-26.
- Kelejian, H. H., & Madan, D. The estimation of a policy response in a nonlinear system. *Economic Statistics Papers*, No. 27, University of Sydney, 1977.
- Kelejian, H., & Vavrichek, B. An evaluation of the forecasting performance of macro econometric models with special emphasis on size. Paper presented at the American Statistical Association meetings, San Diego, Cal., August 1978.
- Kendall, M., & Stuart, A. *The advanced theory of statistics*. Vol. 2. New York: Hafner, 1961.
- Levhari, D. A note on Houthakker's aggregate production function in a multifirm industry. *Econometrica*, 1968, **36**, 151-154.
- Lipsey, R. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862-1957: A further analysis. *Economica*, 1960, **27**, 1-31.
- Mood, A., Graybill, F., & Boes, D. *Introduction to the theory of statistics*. (3rd ed.) New York: McGraw-Hill 1974.
- Orcutt, G., Watts, H., & Edwards, J. Data aggregation and information loss. *American Economic Review*, 1968, **58**, 773-787.
- Ramsey, J. B. Limiting functional forms for market demand curves. *Econometrica*, 1972, **40**, 327-342.
- Sato, K. Micro and macro constant-elasticity-of-substitution production functions in a multifirm industry. *Journal of Economic Theory*, 1970, **1**, 438-453.
- Theil, H. *Linear aggregation of economic relations*, Amsterdam: North-Holland Publ., 1954.
- Zarembka, P. A note on consistent aggregation of production functions. *Econometrica*, 1968, **36**, 419-420.
- Zellner, A. On the aggregation problem: A new approach to a troublesome problem. In K. A. Fox, J. K. Sengupta, and G. V. L. Narasimham (Eds.), *Economic models, estimation and risk programming: Essays in honor of Gerhard Tintner*. New York: Springer-Verlag, 1969. Pp. 365-373.