THE USE OF AVERAGE AND DISTRIBUTION DATA TO CHARACTERIZE MICRO FUNCTIONS

by

Thomas M. Stoker

Massachusetts Institute of Technology

WP #1207-81 August 1981

Thomas M. Stoker is Assistant Professor, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139.

1. Introduction

In formulating and estimating macroeconomic models, researchers tend to either ignore or just mention the problems of aggregation over individuals. Instead, linear relationships are fit to aggregate data, with the coefficients interpreted as representing individual behavior. As the aggregate data often fails to reject the form of such equations, the view that the aggregation problem is of secondary importance is confirmed.

Such an inference, however, is not justified. The true relationship between aggregate data is the result of averaging individual (micro) behavior across the population distribution for each time period.[1] A linear aggregate relationship can arise in a number of situations. A sufficient condition for aggregate linearity is when each agent is characterized by a common linear model, which is also necessary if the distribution can change over time without restriction.[2] If individual behavior is nonlinear, then a linear aggregate relation can result if the movement of the distribution obeys certain restrictions (c.f. Stoker (1980)). Usually, the form of such restrictions will depend on the nonlinear form of micro behavior.[3] However if the distribution movement obeys a simple form of trending (known as linear probability movement), the true aggregate relation will be linear regardless of the form of individual behavior.

These points indicate a difficulty in the interpretation and application of estimated macroeconomic equations. When the movement of the population distribution is limited, it is possible for a linear model to describe the observed configuration of aggregate data quite closely, although the underlying individual behavior is nonlinear in form. As shown later, in this situation the macro coefficient estimates will depend not only on the form of micro behavior, but also on the actual pattern of distribution changes underlying the aggregate data. Consequently, the use of such coefficients for forecasting or policy analysis implicitly assumes the extension of past patterns of distribution movements. In most applications there is no reason to believe that such patterns will continue --

on the contrary, many studies concern economic events (such as oil price shocks) or policies (such as tax changes) which likely involve a substantive change in the underlying population distribution.[4]

In order to assess the extent of this problem in an application, a characterization of micro behavior, or at least a test of micro linearity, is necessary. The above discussion indicates that such a characterization is not possible with aggregate data alone, and requires some additional type of data input.

If the additional data is in individual, cross-section form (for one time period), then it is possible to test for linearity directly, as indicated in Stoker (1981). Even if linearity fails, conditions of asymptotic sufficiency imply that cross-section OLS coefficients provide the proper effects for short-term forecasting. But while providing one method for testing aggregation conditions, this approach is not without flaws. First, the quality of the individual data may be questionable, or it may be definitionally incompatible with the average data. Second, the actual movement of the underlying population distribution is not observed.

The purpose of this paper is to discuss a different approach to testing micro linearity. Often, in addition to the average macroeconomic data, there exists data on the population distribution over time, usually via observations on the proportions of agents in a cellular breakdown of the distribution. Here we indicate how the data on proportions can be used together with the average data to test for linear micro behavior.

Perhaps an example will best illustrate our conclusions. Suppose that the relationship between average food consumption and average income is of interest, and that our theory states that food consumption is a stable function of income for individual families (ignoring prices and other economic variables for simplicity). As data, we have time series of average food consumption, average income and the proportion of agents in $N > 2$ cells of the income distribution (say, those agents with incomes less than \$5,000, between \$5,000 and \$10,000,

etc.). How can these data be used to test the linearity of the micro food consumption-income relationship?

Here we show that regressing average food consumption on a constant, average income and N-2 of the proportions can indicate the micro behavioral structure. In particular, the coefficients on the proportions indicate the deviations of the micro food consumption-income function from linearity. If any of the proportion coefficients is significantly different from zero, micro linearity is rejected.

Whether micro linearity is rejected or not, more can be learned from further scrutiny of the above test. If we fail to reject linearity, and the constant-average income-proportions moment matrix is of full rank N, then micro linearity is strongly confirmed. If the matrix is of less than full rank, then the distribution movement is restricted, and the linearity of the micro food consumption function is not assured. If, assuming that average income is not constant over time, the moment matrix is of smallest rank 2, then the distribution obeys linear probability movement, and the micro behavior, while possibly linear, could be of any form. In these latter cases, the only method of characterizing micro behavior is with individual data.

If micro linearity is rejected, then the same technique can be used to test whether individual food consumption is a more general function of income, say a linear function of income and income squared.[5] This is done by regressing average food consumption on a constant, average income, the second moment of the income distribution, and N-3 proportions; -- proceeding as before. If linearity here is rejected, one can continue to add statistics of the income distribution and test for linearity in up to N-2 statistics, the most these data could distinguish. In this way the micro behavior can be characterized, as much as the available evidence will permit.

We begin the exposition by introducing the notation to be used, starting with the case of a discrete population distribution. We next investigate the regression of the observed proportions on the average explanatory variables,

which will later serve as auxiliary regressions, and which also indicate the relative gain in fit (via residual variance) of aggregate (time-series) regressions over individual (cross-section) regressions. We then discuss the linearity test, showing our various conclusions. Next, changes are indicated in the interpretation of the test when the underlying distribution is continuous. Finally, we close with a summary and discussion of additional topics.[6]

2. Preliminaries

We begin by assuming a dependent quantity x, which is determined by an independent quantity A through a micro behavioral function $x = x(A)$. A is assumed to be a scalar variable for simplicity, and the function $x(A)$ is assumed unchanging over time $t = 1,\ldots,T$. We assume that the population distribution in each year consists of N discrete cells, with p_{jt} denoting the proportion of agents in cell j in year t. Each agent in cell j has $A = A_j$, where without loss of generality $A_i \neq A_j$, for $i \neq j$, and $x = x_j = x(A_j)$. It is convenient to form the vectors $\underline{A}' = (A_1,\ldots,A_N)$, $\underline{x}' = (x_1,\ldots,x_N)$ and $\underline{p}_t' = (p_{1t},\ldots,p_{Nt})$.

The mean of A in year t is denoted μ_t, where $E_t(A) = \mu_t = \underline{A}'\underline{p}_t$. The mean of x in year t is denoted $E_t(x)$, where $E_t(x) = \underline{x}'\underline{p}_t$. We assume that $T \geq N$, $N > 2$ and that μ_t is not constant for all time periods. Finally, in Section 4 we will assume that $\bar{x}_t = E_t(x) + v_t$ is observed, where v_t is normally distributed with mean 0 and variance σ^2, uncorrelated with $\underline{p}_t$ and uncorrelated over time.

In the example referred to in the Introduction, x denotes family food consumption, A denotes income and $\underline{p}_t$ a proportional breakdown of the income distribution. $\underline{A}$ is the vector of average income over cells, and $\underline{x}$ is the vector of average food consumption over cells, where, for the moment, each vector is assumed constant over time. μ_t is average income and $E_t(x)$ and $\bar{x}_t$ are true and observed average food consumption, respectively. The macro function is the relationship between $E_t(x)$ and μ_t, induced by the movement of the population distribution $\underline{p}_t$.

In order to characterize this relationship, we require some results of Stoker (1980), concerning the orthogonal decomposition of $\underline{p}_t$. Let $\underline{i}$ denote the N vector of units, and write $\underline{A}$ uniquely as $\underline{A} = \bar{A}\underline{i} + \tilde{\underline{A}}$, where $\bar{A} = \Sigma A_j/N$, and $\tilde{\underline{A}}'\underline{i} = 0$. In Stoker (1980), it was shown that $\underline{p}_t$ could be written uniquely as

$$\underline{p}_t = \frac{1}{N}\underline{i} + D(\mu_t)\tilde{\underline{A}} + \tilde{\underline{p}}_t \qquad (2.1)$$

where $D(\mu_t) = (\mu_t - \bar{A})/\tilde{A}'\tilde{A}$, and $\underline{i}'\tilde{p}_t = \underline{A}'\tilde{p}_t = 0$.

The first two terms of this decomposition are the same for any density vector obeying $E_t(A) = \mu_t$, with all differences in such distributions appearing in the $\tilde{p}_t$ term, which is orthogonal to $\underline{i}$ and $\tilde{A}$.[7] $\underline{x}$ is decomposed uniquely as

$$\underline{x} = \bar{x}\underline{i} + b\tilde{A} + \underline{x}_n \tag{2.2}$$

where $\bar{x} = \Sigma x(A_j)/N$, $b = \underline{x}'\tilde{A}/\tilde{A}'\tilde{A}$ and $\underline{x}_n'\underline{i} = \underline{x}_n'\tilde{A} = 0$. From (2.1) and (2.2) we have that

$$E_t(x) = \underline{x}'\underline{p}_t = \bar{x} + bD(\mu_t)\tilde{A}'\tilde{A} + \underline{x}_n'\tilde{p}_t = (\bar{x} - b\bar{A}) + b\mu_t + \underline{x}_n'\tilde{p}_t \tag{2.3}$$

It is useful to illustrate the structure of (2.1) and (2.2) graphically. In Figure 1 we have graphed a possible $x = x(A)$ function for $N = 5$. The first two terms of (2.2) represent the points lying on line a-a, the fitted line from "regressing" $x(A_j)$ on A_j and a constant. The numbered "residuals" (deviations) are the component of $\underline{x}_n$. Clearly, $x = x(A)$ is a linear function of A if and only if $\underline{x}_n$ is zero, in which case (2.2) and (2.3) become:

$$\underline{x} = \bar{x}\underline{i} + b\tilde{A} = (\bar{x} - b\bar{A})\underline{i} + b\underline{A} = a\underline{i} + b\underline{A}$$

$$E_t(x) = \underline{x}'\underline{p}_t = a + b\mu_t \tag{2.4}$$

In Figure 2 we have plotted a possible $\underline{p}_t$, $t=1,\ldots,T$, sequence for $N = 3$ and $T = 5$. In Figure 2a, we show the unit simplex $S = \{\underline{i}'\underline{p} = 1 \mid \underline{p} \geq 0\}$, the $T = 5$ distribution points, and the vectors $\underline{i}$ and $\tilde{A}$. In Figure 2b, we show the unit simplex S in two dimensions. Referring to eqn. (2.1), the $\frac{1}{3}\underline{i}$ term places the $\underline{p}_t$ vectors in the unit simplex. The term $D(\mu_t)\tilde{A}$ indicates points along the line a-a. The $\tilde{p}_t$ terms indicate the orthogonal deviations of the actual density points $\underline{p}_t$, from the line a-a. Further, the invariance of the first two terms is reflected in the fact that the position on a-a corresponds to the value of μ_t -- in Figure 2b, if $\underline{p}^*$ is such that $\underline{A}'\underline{p}^* = \mu_2$, then $\underline{p}^*$ must lie on line c-c.

As mentioned in the Introduction, a linear macro relation can result with a nonlinear micro function if the underlying distribution movement is restricted. The extreme form of distribution movement, termed linear probability movement (LPM), occurs if $\tilde{\underline{p}}_t$ follows a linear function of μ_t; i.e. there exist vectors $\underline{c}$ and $\underline{d}$ such that $\tilde{\underline{p}}_t = \underline{c} + \underline{d}\mu_t$. This implies that the density vectors $\underline{p}_t$ lie along the same line in S, such as b-b in Figure 2b. In this case, the macro relation (2.3) appears as:

$$E_t(x) = \underline{x}'\underline{p}_t = \bar{x} + bD(\mu_t)\underline{\tilde{A}}'\underline{\tilde{A}} + \underline{x}_n'(\underline{c} + \underline{d}\mu_t)$$

$$= (\bar{x} - b\bar{A} + \underline{x}_n'\underline{c}) + (b + \underline{x}_n'\underline{d})\mu_t$$

$$= a^* + b^*\mu_t$$

Note that $\underline{x}_n$ is not necessarily zero, i.e. this result holds for any nonlinear $x = x(A)$ function. This completes the presentation of notation and preliminary results.

In Section 3, we show how data on μ_t and $\underline{p}_t$ can be used to investigate the possibility of restricted distribution movement. In Section 4 we show how $\underline{x}_n$ can be estimated from data on $\bar{x}_t$, μ_t and $\underline{p}_t$, and in particular, how $\underline{x}_n = 0$ can be tested, the case of linear micro behavior.

3. Studying Distribution Movement

In this section we assume that we observe $\underline{p}_t$ and μ_t for t = 1,...,T. Our interest is in studying whether the movement of $\underline{p}_t$ is restricted. This is trivial if the $\underline{p}_t$ series strictly obeys LPM, as the difference $\underline{p}_{t'} - \underline{p}_t$ must be $(\mu_{t'} - \mu_t)\underline{d}^*$, for all t,t' with $\underline{d}^*$ a constant vector. But what are the implications if the $\underline{p}_t$ series is "close" to LPM, and how should we define "close"?

We begin this inquiry by supposing that $\underline{c}$ and $\underline{d}$ are arbitrary vectors of constants, and defining the vector $\underline{u}_t$ as:

$$\underline{u}_t = \tilde{\underline{p}}_t - \underline{c} - \underline{d}\mu_t = \underline{p}_t - \frac{1}{N}\underline{i} - D(\mu_t)\tilde{\underline{A}}_t - \underline{c} - \underline{d}\mu_t$$

Without loss of generality, we can assume that $\underline{c}'\underline{i} = \underline{d}'\underline{i} = \underline{c}'\tilde{\underline{A}} = \underline{d}'\tilde{\underline{A}} = 0$, so that $\underline{u}_t$ lies in the N-2 dimensional subspace of R^N orthogonal to $\underline{i}$ and $\tilde{\underline{A}}$.

Our object is to choose $\underline{c}$ and $\underline{d}$ so as to describe $\tilde{\underline{p}}_t$ as an LPM process as closely as possible. Regressing each component of $\tilde{\underline{p}}_t$ (or $\underline{p}_t$) on μ_t and a constant appears to be a good idea, as the components of $\underline{u}_t$ are each made small, but why is this the proper metric? We now justify this procedure by a least squares argument pertaining to the macro variable $E_t(x)$.

Suppose that x = x(A) is a given function, i.e. $\underline{x} = \bar{x}\underline{i} + b\tilde{\underline{A}} + \underline{x}_n$ is given. Define x_t^f as:

$$x_t^f = \underline{x}'(\underline{p}_t - \underline{u}_t) = (\bar{x} - b\bar{A} + \underline{x}_n'\underline{c}) + (b + \underline{x}_n'\underline{d})\mu_t$$

x_t^f is the fitted mean x, utilizing the invariant coefficients of $\underline{p}_t$ plus $\underline{c}$ and $\underline{d}$. By definition we have that

$$E_t(x) - x_t^f = \underline{x}_n'\underline{u}_t$$

and so

$$(E_t(x) - x_t^f)^2 = (\underline{x}_n'\underline{u}_t)^2 \leq |\underline{x}_n|^2 \, |\underline{u}_t|^2$$

by the Cauchy-Schwartz inequality. Summing over t=1,...,T gives:

$$\Sigma(E_t(x) - x_t^f)^2 \leq |\underline{x}_n|^2 \, \Sigma \, |\underline{u}_t|^2 \qquad (3.1)$$

The bound (3.1) on the average sum of squares will be "tightest" (for a given $\underline{x}$) if we choose $\underline{c}$ and $\underline{d}$ to minimize $\Sigma|\underline{u}_t|^2$. But this is computationally equivalent to choosing the components of $\underline{c}$ and $\underline{d}$ from OLS regressions for $j=1,\ldots,N$, of $\tilde{p}_{jt}$ on μ_t and a constant, $t=1,\ldots,T$. Clearly, the same estimated residuals $\hat{\underline{u}}_t$ are obtained from OLS regressions for $j=1,\ldots,N$ of p_{jt} on μ_t and a constant. Figure 3 reproduces Figure 2b, where now line b-b represents the fitted distribution vectors, and the residual vectors $\hat{\underline{u}}_1,\ldots,\hat{\underline{u}}_5$ are noted.

In addition to justifying OLS as above, the bound (3.1) is informative about the difference between the fit of an unweighted cross-section regression of $x(A_j)$ on A_j, and a time-series regression of $E_t(x)$ on μ_t. Let x_t° be the fitted value at time t from regressing (the true) $E_t(x)$ on μ_t and a constant. Then, since x_t^f is a linear function of μ_t, (3.1) can be expanded to:

$$\Sigma(E_t(x) - x_t^\circ)^2 \leq \Sigma(E_t(x) - x_t^f)^2 \leq |\underline{x}_n|^2 \, \Sigma \, |\underline{u}_t|^2 \qquad (3.2)$$

By considering the Gauss-Markov Theorem, it is easy to see that if $\hat{\underline{c}}$ and $\hat{\underline{d}}$ are chosen by OLS as indicated above, that $x_t^\circ = x_t^f$, $t=1,\ldots,T$, and so the left inequality of (3.2) can be replaced by equality. Define the average aggregate residual variance as:

$$ARV = \frac{\Sigma(E_t(x) - x_t^\circ)^2}{T}$$

Now, in Section 2 it was indicated that the components of $\underline{x}_n$ could be viewed as residuals from an unweighted regression of $x(A_j)$ on A_j, $j=1,\ldots,N$. If a cross-section data base consisted of M agents in each of the N cells, then the average residual variance from regressing x on A and a constant would be:

$$CRV = |\underline{x}_n|^2/N$$

Thus (3.2) can be rewritten as:

$$ARV \lesssim CRV \cdot N \cdot \frac{\Sigma|\hat{u}_t|^2}{T} \tag{3.3}$$

Now, if the micro function $x = x(A)$ were exactly linear, then CRV = 0 and ARV = 0. Otherwise, we have

$$\frac{ARV}{CRV} \lesssim N \cdot \frac{\Sigma|\hat{u}_t|^2}{T} \tag{3.4}$$

Equation (3.4) gives the relative gain in fit between an (unweighted) cross-section regression and a time-series regression due to aggregation, when the underlying micro function is nonlinear. Note that if the distribution follows LPM exactly, then $|\hat{u}_t| = 0$ for all t, and ARV = 0, independent of the value of CRV.

Unfortunately (3.4) is not a strong theoretical tool, as there are circumstances when the right-hand side is relatively large. But for "reasonable" values of N, T and $\hat{u}_t$, (3.4) indicates a substantial gain in fit from aggregation. For example, suppose that N = 10, T = 20, and that the standard deviation of the proportion residuals $\hat{u}_{jt}$ is .01 (one percent of the population). In this case $\Sigma|\hat{u}_t|^2 \lesssim .02$, and

$$\frac{ARV}{CRV} \lesssim .01$$

so that the time series standard error is less than one-tenth the cross-section standard error. Whether one views these numbers as reasonable or not, the main importance of (3.4) is that it holds for any (nonlinear) function $x = x(A)$, and using data on proportions to get the $\hat{u}_t$, the right-hand side of the bound is calculable, and so may indicate the extent of standard error differences due to aggregation.[8] Do bear in mind, however, that CRV represents unweighted (between) cross-section residual variance and that ARV represents the true aggregate variance, omitting aggregate disturbances.

To close this section, we raise one more issue concerning the interpretation of the regressions of proportions on μ_t and a constant. For purposes of the linearity test, presented next, we see these regressions only as auxiliary regressions, absent a formal model of distribution movement. If, however, it is believed that past distributional trends and policy effects should follow a linear process in μ_t up to a stochastic component, then a multivariate regression model is appropriate to estimate the trend parameters. The only assumption crucial to the next section is that $E_t(x)$ has been formed from the observed proportions, and not just the parametric model. Otherwise, an errors-in-variables problem appears with respect to the proportion data.

4. Testing Micro Linearity

The true model determining $\bar{x}_t$, t=1,...,T is

$$\bar{x}_t = \underline{x}'\underline{p}_t + v_t \tag{4.1}$$

where $\underline{x}$ represents fixed coefficients of the distribution proportions $\underline{p}_t$, and v_t is the disturbance introduced earlier. For now we assume that the TxN matrix P with t^{TH} row $\underline{p}_t'$ is of full rank N. Later this assumption is relaxed, and its implications discussed.

Under these assumptions, we see that BLUE estimates of $\underline{x}$ can be obtained by regressing $\bar{x}_t$ on $\underline{p}_t$ (without a constant). Of more central interest, though, is the relationship between $\bar{x}_t$ and μ_t, the mean of A. From our previous development, (4.1) can be rewritten as

$$\bar{x}_t = (a + \underline{x}_n'\hat{\underline{c}}) + (b + \underline{x}_n'\hat{\underline{d}})\mu_t + \underline{x}_n'\hat{\underline{u}}_t + v_t \tag{4.2}$$

with $\underline{x} = a\underline{i} + b\underline{A} + \underline{x}_n$, and $\hat{\underline{c}}$, $\hat{\underline{d}}$ and $\hat{\underline{u}}_t$ the estimated coefficient and residual vectors from the auxiliary proportion regressions $\underline{\tilde{p}}_t = \underline{c} + \underline{d}\mu_t + \underline{u}_t$. Since $\hat{\underline{u}}_t$ is uncorrelated with both μ_t and the constant by construction, estimating

$$\bar{x}_t = a^* + b^*\mu_t + e_t \tag{4.3}$$

by least squares yields $\hat{a}^*$ and $\hat{b}^*$, where

$$E(\hat{a}^*) = a + \underline{x}_n'\hat{\underline{c}}$$

$$E(\hat{b}^*) = b + \underline{x}_n'\hat{\underline{d}}$$

Thus, the estimates of a^* and b^* depend on three factors; the behavioral coefficients a, b, the nonlinearity of $x = x(A)$ (through $\underline{x}_n$) and the movement of the underlying distribution (through $\hat{\underline{c}}$ and $\hat{\underline{d}}$).

This illustrates the general notion that if microeconomic behavior is nonlinear in form, the estimated coefficients of a linear macro function (such as (4.3)) will depend on the empirical history of distribution movements. Moreover,

it is easy to see that a forecast from such an equation will be unbiased only if the future distribution movement extends the pattern of past distribution movement.[9] Since forecasts often involve evaluating effects of different policies (such as tax changes) or of outside economic shocks, assuming such regularity in future distribution changes may not be warranted. In this situation, micro linearity is virtually mandatory for confidence in forecasts from linear macro equations.

Testing micro linearity in our framework amounts to testing whether $\underline{x}_n = 0$. With reference to equation (4.2), we could include the proportion residuals $\hat{\underline{u}}_t$ as regressors in the macro equation, but this suffers from collinearity through the relations $\underline{i}'\hat{\underline{u}}_t = \underline{A}'\hat{\underline{u}}_t = 0$. This implies that $\underline{f}'\underline{x}_n$ is estimable only if $\underline{f}'\underline{i} = \underline{f}'\underline{A} = 0$. But this suffices for our purposes, since by construction we have $\underline{i}'\underline{x}_n = \underline{A}'\underline{x}_n = 0$. Therefore, BLUE estimates of $\underline{x}_n$ (and a^*, b^*) are obtained by performing the regression of $\bar{x}_t$ on μ_t, $\hat{\underline{u}}_t$ and a constant, constraining the $\hat{\underline{u}}_t$ coefficients to obey $\underline{i}'\underline{x}_n = \underline{A}'\underline{x}_n = 0$. (The details of this procedure are presented in the Appendix.) $\underline{x}_n = 0$ is then tested by a conventional F test. Moreover, by the mechanics of least squares, we can replace $\hat{\underline{u}}_t$ by $\underline{p}_t$ in the above procedure, dispensing with the need to calculate $\hat{\underline{u}}_t$.

A much simpler technique is to omit two of the proportions, including only N-2, which we collect here as a vector $\underline{p}_t^*$. That is, we estimate by least squares the model

$$\bar{x}_t = a' + b'\mu_t + \underline{x}_n^{*\prime}\underline{p}_t^* + v_t \tag{4.4}$$

The new coefficients $\underline{x}_n^*$ are unconstrained, and are connected to the original coefficients $\underline{x}_n$ via

$$M\underline{x}_n = \begin{bmatrix} 0 \\ 0 \\ \underline{x}_n^* \end{bmatrix}$$

where M is a nonsingular matrix that depends on $\underline{A}$ (see the Appendix).

Therefore $\underline{x}_n = 0$ if and only if $\underline{x}_n^* = 0$, and so the standard F test of $\underline{x}_n^* = 0$ provides a test of micro linearity. This procedure avoids having to perform a constrained regression, and provides a test of micro linearity which does not require knowledge of the vector $\underline{A}$. $\underline{A}$ must be known to transform the estimates of $\underline{x}_n^*$ to estimates of $\underline{x}_n$ (for constructing M), but not for testing $\underline{x}_n = 0$, the case of most interest.

The problems of restrictions on distribution movement discussed earlier manifest themselves empirically in a reduced rank of the matrix P. If the rank of P is N-1 (since μ_t is not constant over time), then there exists $\underline{\alpha}$ such that $\underline{\alpha}'\underline{\hat{u}}_t = 0$ for all t, with $\underline{\alpha}'\underline{i} = \underline{\alpha}'\underline{A} = 0$. In this case $\underline{f}'\underline{x}_n$ is estimable in the constrained regression if and only if $\underline{f}'\underline{\alpha} = \underline{f}'\underline{i} = \underline{f}'\underline{A} = 0$. Thus, the nonlinear coefficients are not all identified. This situation clearly imparts a similar lack of identification to the coefficients in the unconstrained regression technique (see Appendix for details). The extreme form of this problem occurs when the rank of P is 2, which implies that $\underline{\hat{u}}_t = 0$ for all t. In this case no linear combination of the $\underline{x}_n$ coefficients is estimable. This is nothing more than the case of LPM discussed earlier.

This discussion points out that difficulties in discerning nonlinear micro-economic structure arise from a classical source, namely multicollinearity. The extreme forms of collinearity or distribution restrictions discussed above forbid identification of certain linear combinations of $\underline{x}_n$. Even if the rank of P is N, limited movement of the distribution in a direction $\underline{\alpha}$ will result in $\underline{\alpha}'\underline{x}_n$ not being estimated precisely. From an analysis of the proportion residuals $\underline{\hat{u}}_t$, it may be possible to discern which combinations of $\underline{x}_n$ will not be estimated precisely, and a priori considerations used to mitigate the problem. But if the data is weak, there is no easy empirical resource. If the problem is severe (as with LPM), either more aggregate data must be obtained, or attention turned to a different data source, such as cross-section data on individuals.

So far our discussion has centered on the use of only a single statistic μ_t to describe the movement of $\bar{x}_t$. If either linearity of micro behavior is rejected, or if other variables (other than A) are thought to affect x, then the micro linearity of x in several variables can be studied. In particular, suppose that the basic model is $x = x(A,B)$, where each agent in cell j has $B = B_j$, and $\underline{B} = (B_1,\ldots,B_N)'$. Denote the mean of B as $\nu_t = \underline{B}'\underline{p}_t$. Now, as long as $\underline{i}$, $\underline{A}$ and $\underline{B}$ are linearly independent, a regression of $\bar{x}_t$ on a constant, μ_t, ν_t and N-3 of the proportions can be used to test linearity of $x = x(A,B)$ in both A and B. That is, we estimate

$$\bar{x}_t = a + b\mu_t + c\nu_t + \underline{x}_n^{**\prime}\underline{p}_t^{**} + v_t \qquad (4.5)$$

where $\underline{p}_t^{**}$ is the vector of N-3 proportions, and test for $x_n^{**} = 0$.

If micro linearity of x in A is rejected, then a plot of the estimates of $\underline{x}_n$ can indicate how to revise the model. As in the example cited in the Introduction, if such a plot indicates parabolic curvature, then $B = A^2$ can be chosen, with $\nu_t = \Sigma p_{jt} A_j^2$. In this case, a test of $\underline{x}_n^{**} = 0$ (of (4.5)) is a test of whether x is a quadratic function of A.

As above, in the case where rank P = N, one can test the linearity of x in at most N-2 variables (plus the constant). N-1 variables (plus the constant) will exhaust all the available degrees of freedom of the observed N proportions. Moreover, if when a certain number of distribution statistics are included, a problem of multicollinearity (imprecise estimation) arises, then including more statistics will only make the problem worse. However, subject to these considerations, a rather complete characterization of the micro behavior consistent with the aggregate data is possible by our techniques.

5. Complications of Continuous Distributions

In an empirical investigation of aggregation structure, the analyst is likely to encounter problems not covered by our basic setup. For instance, there may be variables common to all agents in each time period which affect behavior, such as prices or general economic conditions. Thus the problem of varying behavior (i.e. changing functional form of $x = x(A)$) must be faced as well as the aggregation problem.

The issues connected with changing functional form will vary from one application to another, and so we refrain from discussing them here. A related problem refers to when the basic (within cell mean) vectors $\underline{x}$ and $\underline{A}$ vary over time. This can occur if the actual variation in the predictor variable A is continuous, but the distribution is only observed as cell proportions over time. In this section we indicate potential biases in our test of micro linearity when the true underlying distribution is continuous.

For this discussion we must expand our notation slightly. As before, $j=1,\ldots,N$ indexes a cellular division of the population, with p_{jt} the proportion of agents in cell j at time t. A is a continuous random variable distributed across the population, with $A_{jt} = E_t(A|j)$, $j=1,\ldots,N$, denoting the within cell means of A at time t. $x = x(A)$ is a stable behavioral function, and $x_{jt} = E_t(x(A)|j)$ $j=1,\ldots,N$ are the within cell means of x at time t. As before we form the vectors $\underline{p}_t$, $\underline{A}_t = (A_{1t},\ldots,A_{Nt})'$ and $\underline{x}_t = (x_{1t},\ldots,x_{Nt})'$. The overall means of A and x at time t are given as $\mu_t = E_t(A) = \underline{A}_t'\underline{p}_t$, $E_t(x) = \underline{x}_t'\underline{p}_t$. As before we assume that $\underline{p}_t$, μ_t and $\bar{x}_t = E_t(x) + v_t$ are observed for each time period.

If $\underline{A}_t$ and $\underline{x}_t$ are constant over all time periods then the previous analysis applies, with a test of (between cell) micro linearity provided by the estimation of equation (4.4). But suppose that this estimation is performed when $\underline{x}_t$ and $\underline{A}_t$ vary -- is the test of (between cell) linearity still valid? We show below that such estimation is subject to biases, possibly invalidating the linearity test, with the source of bias a familiar one to students of aggregation theory.

We must first reinterpret the coefficients a, b and $\underline{x}_n^*$ of (4.4). Ideally, we would like a characterization of the entire function $x = x(A)$, or at least a description of the relation between $\underline{x}_t$ and $\underline{A}_t$ for each time period. This is not possible by our previous method (with $\underline{A}_t$ varying), since the orthogonal decomposition (2.1) changes each period. Instead, since the function $x = x(A)$ is unchanging over time, we consider the relation between the time averages of the within cell means of $\underline{x}$ and $\underline{A}$, defined as $\underline{x}^m = \Sigma_t \underline{x}_t/T$ and $\underline{A}^m = \Sigma_t \underline{A}_t/T$. The (between cell) "average" micro relation is

$$\underline{x}^m = a\underline{i} + b\underline{A}^m + \underline{x}_n \tag{5.1}$$

which provides an interpretation of $\underline{x}_n^*$ of (4.4). We can define the overall means of $\underline{x}^m$ and $\underline{A}^m$ as $E_t(x^m) = \underline{x}^{m\prime}\underline{p}_t$; $E_t(A^m) = \mu_t^m = \underline{A}^{m\prime}\underline{p}_t$; and pretend that μ_t^m and $\bar{x}_t^m = E_t(x^m) + v_t$ are observed. The model corresponding to (4.4) with (5.1) as a basis is then:

$$\bar{x}_t^m = a' + b'\mu_t^m + \underline{x}_n^*\underline{p}_t^* + v_t \tag{5.2}$$

If $\bar{x}_t^m$ and μ_t^m were actually observed, then estimates of the "average" parameters a', b' and x_n^* would be provided by least squares, and $\underline{x}_n^* = 0$ could be tested. If $\underline{x}_n^* = 0$ were rejected, then we clearly must reject that $x = x(A)$ is a linear function.

The actual estimation is performed with $\bar{x}_t$ and μ_t, instead of $\bar{x}_t^m$ and μ_t^m. With (5.2) as the "true" model, the model relating the observed data is

$$\bar{x}_t = a' + b'\mu_t + \underline{x}_n^*\underline{p}_t^* + \gamma_t + v_t \tag{5.3}$$

where $\gamma_t = [E_t(x) - E_t(x^m) - b'(\mu_t - \mu_t^m)]$. Estimating (4.4) with the observed data omits γ_t, and thus correlation between γ_t and μ_t or $\underline{p}_t^*$ would lead to biased estimates, with the test of $\underline{x}_n^* = 0$ invalid.

We can learn more about the size of this problem by writing γ_t as:

$$\gamma_t = \sum_{j=1}^{N} p_{jt}(x_{jt} - x_j^m - b'(A_{jt} - A_j^m))$$

γ_t is thus the average error in predicting variations in the within cell means of x with variations in the within cell means of A and the average slope coefficient b'. If such errors tend to average to zero (or a constant over time) or if large errors occur in small probability cells (i.e. x(A) is linear over heavily populated cells), then γ_t will be small and biases relatively unimportant. Alternatively, if we can consider the function $x = x(A)$ to have constant slopes b_j, $j=1,\ldots,N$ within each cell over the range of variations of A_{jt}, $j=1,\ldots,N$, $t=1,\ldots,T$, then γ_t appears as

$$\gamma_t = \sum_{j=1}^{N} p_{jt}(b_j - b')(A_{jt} - A_j^m)$$

which is the average product of deviations in the slopes of x(A) and the deviations of A_{jt}. The form of this term should be familiar to readers of H. Theil, whose consistent linear aggregation conditions require the overall covariance of predictor variables and slope coefficients to vanish.[10] Again if large deviations occur in small probability cells, γ_t is likely to be small.

The above discussion with regard to the size of γ_t essentially bounds the possible biases, with actual biases depending on the correlation between γ_t and μ_t or $\underline{p}_t$. At any rate, we should expect that if x(A) is a relatively smooth function, then the γ_t term will be smaller for finer cellular partitions (larger N) of the domain of A, since the time variation of each term of $\underline{x}_t$ and $\underline{A}_t$ should decrease.

One final point is that even if the estimated coefficients of (4.4) are biased, multicollinearity in the $\underline{p}_t$ data still indicates restricted distribution movement as before. In this case the estimation is subject to both the above bias problems and the more fundamental lack of identifiability of the nonlinear micro effects.

6. Summary and Extensions

In this paper we have introduced a new method of studying aggregation problems in empirical work. The key feature of this approach is the formal incorporation of distribution movement into macroeconomic relations.

We have shown explicitly that ignoring the aggregation problem in estimating macroeconomic relations can lead to biases (if micro behavior is nonlinear) which make forecasting or policy analysis from such equations problematic. A method was provided for studying these issues through testing for micro linearity, when the distribution movement is sufficiently varied.

A limited amount of variation in the underlying distribution will force the estimates of the nonlinear effects to be imprecise, and in the extreme case of restricted distribution movement, certain nonlinear effects will not be identified. Such problems arise because of the equivalence of distribution restrictions and multicollinearity in our procedure. Moreover, we have found a theoretical link between average time series variation and individual cross-section variation, by studying the consequences of limited distribution movement.

Although it is hoped that these considerations provide some headway into studying aggregation issues, it is clear that our basic (static) framework is much too simplistic to address many current problems of macroeconomics, as well as production and demand analysis. The framework must be extended to accommodate a number of additional problems, such as distributed lags, simultaneity and external effects in the behavior of individual agents. We conclude the paper with a few cursory remarks on these areas, in order to motivate further study.

For each of these topics, it is likely that aggregation problems can be very severe. Consider first the case where agents behave with respect to a number of lags in a certain variable. A moment's reflection will reveal that if such behavior is nonlinear, then the number of (averaged) lags in the true macro relation can be either greater than, equal to, or less than the number of lags considered by individual agents. Here linearity tests must be directed toward

the identification of the length of individual lag structures as opposed to aggregate lag structures.

Usual textbook examples of simultaneity in econometrics, such as supply and demand interplay, must be re-examined as to whether they arise from individual behavior or aggregate consequences of individual behavior. The task of modeling such detailed interactions is likely to be monumental, and must address the issues of whether the population distribution is endogenous.

The problems of external economies or diseconomies in individual behavior lie in a rather fundamental change in the definition of the correct macroeconomic relation. Although there is substantial theoretical discussion and some progress on empirical techniques in this area, the issues of distribution and behavior change have not been resolved. This topic, however, may be the most important, as many current theories of the macroeconomy are based on the use of aggregate data by individual decision makers.

APPENDIX: OMITTED FORMULAE

In this Appendix we present some formulae relevant to the estimation of Section 4. Although our discussion utilizes $\hat{\underline{u}}_t$ as predictors, bear in mind that for estimating the nonlinearity effects $\underline{x}_\eta$, it is equivalent to replace $\hat{\underline{u}}_t$ by $\underline{p}_t$.

Constrained Estimation of Equation (4.2)

Least squares subject to an identifying constraint is a standard technique (c.f. Scheffe, The Analysis of Variance or Theil, Principles of Econometrics), and so here we present only the formulae for the estimator. Let X be the T x 1 matrix with t^{TH} element $\bar{x}_t$, W be the T x (N+2) matrix with t^{TH} row $[1,\mu_t,\hat{\underline{u}}_t]$, H be the matrix of constraints

$$H = \begin{bmatrix} 0 & 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & A_1 & A_2 & \dots & A_N \end{bmatrix}$$

and $\eta = [a^*,b^*,\underline{x}_\eta]'$ be the parameters to be estimated. Then we seek $\hat{\eta}$ which uniquely solves

$$W'W\hat{\eta} = W'X \qquad ; \qquad H\hat{\eta} = 0$$

$\hat{\eta}$ is given as

$$\hat{\eta} = (W'W + H'H)^{-1} W'X$$

The Transformation of $\underline{x}_\eta^*$ to $\underline{x}_\eta$

In this section we present the matrix M for transforming $\underline{x}_\eta^*$ values to $\underline{x}_\eta$ values. We assume that the first two components of $\hat{\underline{u}}_t$ are dropped in order to estimate $\underline{x}_\eta^*$. Partition $\hat{\underline{u}}_t$ and $\underline{x}_\eta$ into vectors with 2 and N-2 components as $\hat{\underline{u}}_t = (\hat{\underline{u}}_t^1,\hat{\underline{u}}_t^*)$ and $\underline{x}_\eta = (\underline{x}^1,\underline{x}^2)$. Let K denote the matrix

$$K = \begin{bmatrix} \underline{i}' \\ \underline{A}' \end{bmatrix} = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$$

where K_1 is a 2x2 matrix, and K_2 a 2x(N-2) matrix.

Now, if J denotes the matrix

$$J = \begin{bmatrix} K_1 & K_2 \\ 0 & I_{N-2} \end{bmatrix}$$

then J is nonsingular and

$$J\underline{\hat{u}}_t = \begin{bmatrix} 0 \\ 0 \\ \underline{\hat{u}}_t^* \end{bmatrix}$$

Therefore

$$\underline{x}_n'\underline{\hat{u}}_t = (\underline{x}_n'J^{-1})(J\underline{\hat{u}}_t) = \underline{x}_n^{*\prime}\underline{\hat{u}}_t^*$$

where $\underline{x}_n^* = \underline{x}_2 - K_2'K_1^{-1}\underline{x}_1$. Solving for $\underline{x}_n$ from $\underline{x}_n^*$ amounts to solving the system

$$\begin{bmatrix} K_1 & K_2 \\ -K_2'K_1^{-1} & I_{N-2} \end{bmatrix} \underline{x}_n = \begin{bmatrix} 0 \\ 0 \\ \underline{x}_n^* \end{bmatrix}$$

so that

$$M = \begin{bmatrix} K_1 & K_2 \\ -K_2'K_1^{-1} & I_{N-2} \end{bmatrix}^{-1}$$

With the above development, the remark in Section 4 on the identifiability of $\underline{x}_n^*$ is trivial: Suppose $\underline{\alpha}'\underline{\hat{u}}_t = 0$ for t=1,...,T, where $\underline{\alpha}'\underline{i} = \underline{\alpha}'A = 0$, then

$$0 = \underline{\alpha}'\underline{\hat{u}}_t = (\underline{\alpha}'J^{-1})(J\underline{\hat{u}}_t) = \underline{\beta}\,\underline{\hat{u}}_t^*$$

so that $\underline{\beta}'\underline{x}_n^*$ is not identified.

FOOTNOTES

1. In this paper we use the terms "aggregate" and "average" interchangeably, although the issues (and results) apply to studies of "total" variables. From a practical standpoint, we refer to usual macroeconomic relations, such as consumption and investment functions, as well as "microeconomic" studies of average production or demand.

2. See Lau (1980) for a proof of this result. The linearity condition is known as an exact aggregation condition; see also Muellbauer (1975,1977) and Gorman (1953), among others.

3. A well-known example of this is found in the consistent aggregation approach of Theil (1954), where agents behave with respect to differing linear micro functions, and aggregate linearity is assured by the assumption that the covariance between slope coefficients and predictor variables vanishes. Note that our use of the term "linear micro behavior" refers to a common linear model for each individual agent. We return to this point in Footnote 10.

4. These problems with estimated macroeconomic relations have been discussed in partial form before. Certain sections of Theil (1954) and the chapter on aggregation in Allen (1965) indicate awareness of these problems. The essential structure of this argument is presented in the pioneering article de Woolf (1941). Finally, it should be noted that this criticism is distinct from that of Lucas (1976), where it is argued that behavioral parameter change will render estimated macro equations ineffective for policy analysis.

5. As indicated in Section 4, if the within cell means of income are observed and constant, one can solve for estimates of the actual deviations from linearity of micro food consumption. These deviations can then empirically indicate how to generalize the individual (and aggregate) model.

6. There is rather substantial literature on aggregation problems in econometrics; notably Grunfeld and Griliches (1960), Kuh (1959), and Kuh and Welsh (1976) among others.

   These studies begin with the consistent aggregation approach of Theil (1954), which does not focus explicitly on the distributional aspects, as does our approach. Certain aspects, such as the "synchronization" effect noted by Grunfeld and Griliches, are present in our current analysis (see Section 3, Note 8).

7. Note that (2.1) does not restrict the form of the distribution vector p_t, other than to reflect the condition $\mu_t = \underline{A}'\underline{p}_t$.

8. The inequalities (3.2) and (3.4) illustrate the "synchronization" effect of Grunfeld and Griliches (1960) in our format.

9. Notice that this is true even if $\hat{\underline{c}} = \hat{\underline{d}} = 0$.

10. Note that in the consistent aggregation framework (Theil (1954)) if the <u>within cell</u> covariances of slope coefficients and predictor variables do not vanish, the between cell micro relationship can be nonlinear.

REFERENCES

Allen, R. G. D., Mathematical Economics, 2nd ed., London: Macmillan & Co., Ltd., 1965.

de Woolf P., "Income Elasticity of Demand, A Microeconomic and a Macroeconomic Interpretation," Economic Journal 51 (April 1941):104-45.

Gorman, W. M., "Commodity Preference Fields," Econometrica 21 (January 1953):63-80.

Grunfeld, Y. and A. Griliches, "Is Aggregation Necessarily Bad?," Review of Economics and Statistics, 42, 1960, pp. 1-13.

Kuh, E., "The Validity of Cross-Sectionally Estimated Behavior Equations in Time Series Applications," Econometrica, April 1959.

Kuh, E. and R. Welsh, "The Variances of Regression Coefficient Estimates Using Aggregate Data," Econometrica, February 1976.

Lau, L. J., "Existence Conditions for Aggregate Demand Functions," forthcoming in Econometrica, 1980.

Lucas, R., "Econometric Policy Evaluation: A Critique," in Brunner and Meltzer, eds., The Phillips Curve and Labor Markets, Amsterdam: North-Holland Publishing Company, 1976.

Muellbauer, J., "Aggregation, Income Distribution and Consumer Demand," Review of Economic Studies 42 (October 1975):525-43.

Muellbauer, J., "Commodity Preferences and the Representative Consumer," Econometrica 44 (1977):979-99.

Stoker, T., "Aggregation and Assumptions on Macro Functions," MIT Sloan School of Management Working Paper No. WP1177-80, November 1980.

Stoker, T., "The Use of Cross Section Data to Characterize Macro Functions," MIT Sloan School of Management Working Paper No. 1084-79, Revised April 1981.

Theil, H., Linear Aggregation of Economic Relations, Amsterdam: North-Holland Publishing Company, 1954.

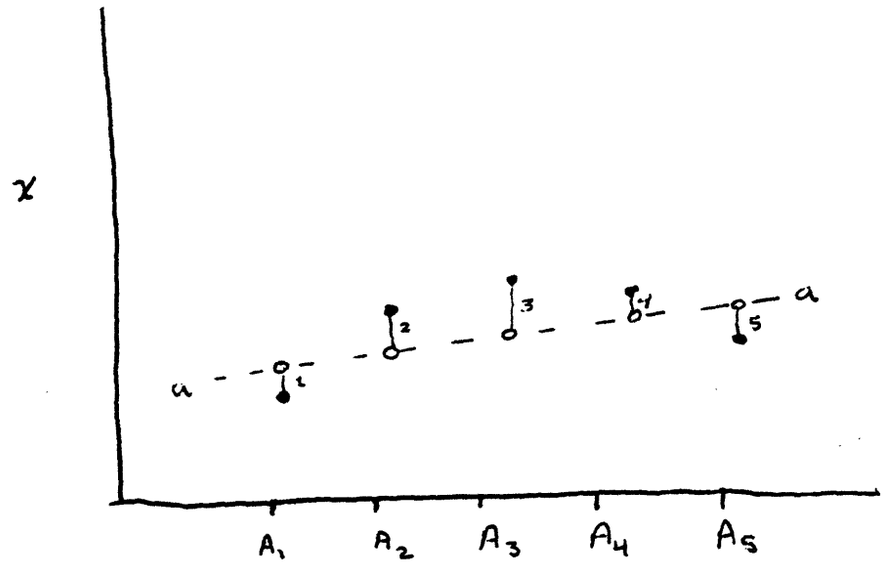
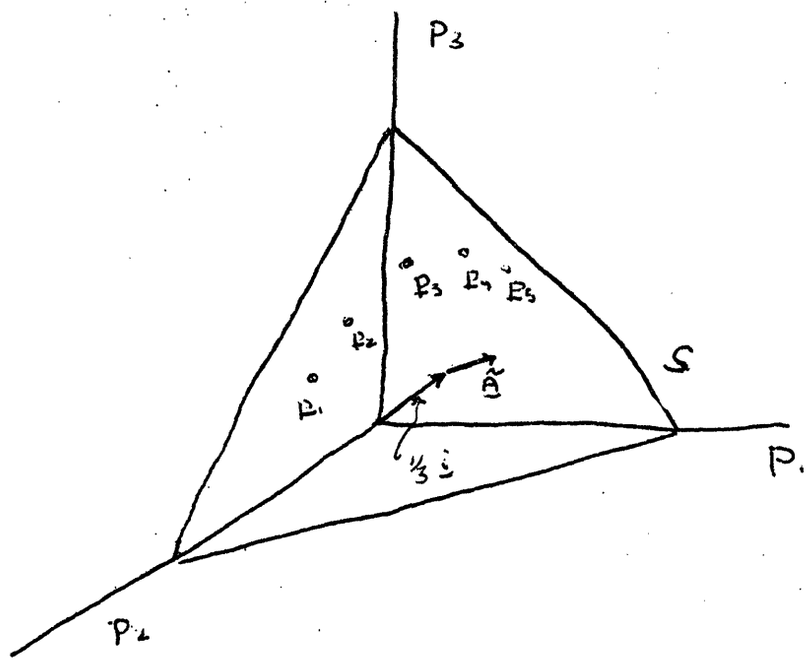
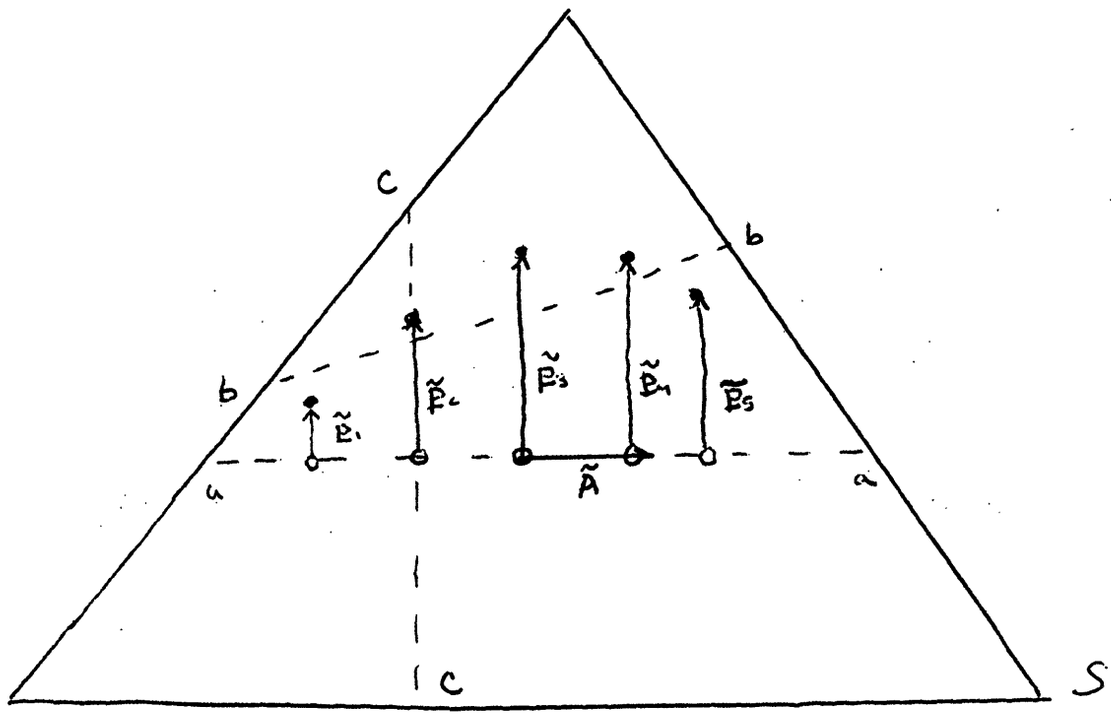


FIGURE 1: Illustration of the $\underline{x}$ Decomposition (2.2)



a



b

FIGURE 2: Illustration of the $\underline{p}_t$ Decomposition (2.1)

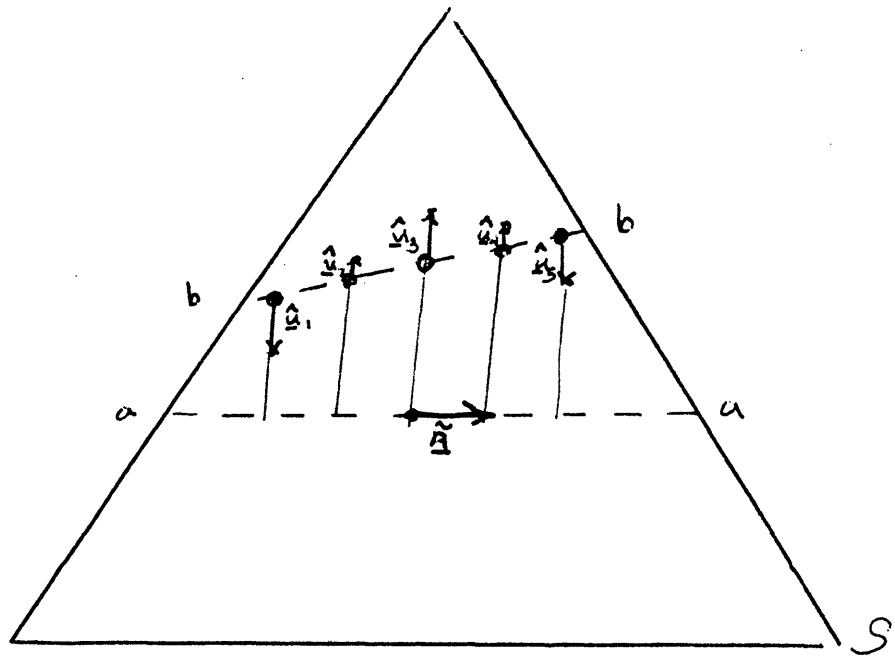


FIGURE 3: Illustration of Probability Regressions