WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

STATISTICAL AGGREGATION ANALYSIS: CHARACTERIZING MACRO FUNCTIONS WITH CROSS SECTION DATA

by

Thomas M. Stoker

WP 1084-79

October 7, 1979 Revised: April 3, 1980
(revision of #1025)

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

STATISTICAL AGGREGATION ANALYSIS: CHARACTERIZING MACRO FUNCTIONS WITH CROSS SECTION DATA

by

Thomas M. Stoker

WP 1084-79

October 7, 1979 Revised: April 3, 1980
(revision of #1025)

ABSTRACT

This paper investigates the use of individual cross section data to describe macro functions. Necessary and sufficient conditions (denoted AS) are found for OLS slope coefficients from a cross section to consistently estimate the first derivatives of the macro function. AS embodies both sets of aggregation assumptions known; linear aggregation and sufficient statistics, and thus represents generalized aggregation conditions. A methodology is given for estimating second order derivatives of the macro function from cross section data for distributions of the exponential family, which extends to higher order derivatives. Finally, a general test of linear aggregation schemes is described.

KEYWORDS

Cross Section Least Squares Regression

Aggregation Theory

Asymptotic Sufficiency

Linear Aggregation

Sufficient Statistics

Exponential Family

* Thomas M. Stoker is Assistant Professor, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. The author wishes to thank J. Green, D. Schmalensee, D. McFadden, J. Hausmann, P. Krugman, and E. Kuh for helpful comments.

Statistical Aggregation Analysis: Characterizing Macro Functions with Cross Section Data*

1. INTRODUCTION

It is common practice in the study of macroeconomic relations to derive a model among the relevant variables based on individual behavior, and then estimate the model's parameters using time series data on the averages of those variables across the population. Such estimated relations are justified as describing the behavior of an individual with "representative" values of the predictor variables.

In general, the true macro relation between averaged data results from the process of integrating (averaging) the true individual behavioral function over the distribution of its arguments in the population. Even in the simplest consumption function regression of average consumption on average income, one is only capturing the statistical relation between two summary statistics of the underlying consumption-income distribution. Unless saving behavior is virtually identical across individuals or the structure of the income distribution can be simply represented, an average consumption-average income regression will not adequately describe the structure relating average consumption to the income distribution.[1] In this sense, any macro function in the form of an individual behavioral relation is likely to ignore important distributional influences.[2]

A consistent model of such a macro relation thus requires both the specification of the individual behavioral function and the population distribution of its arguments. Only if the analyst resorts to restrictive assumptions provided by aggregation theory, such as linearity in the individual behavioral function, can the requirement of fully specifying the behavioral function and distribution forms be relaxed. In addition, just stating such restrictions and proceeding to estimate with average data only provides a weak basis for the macro relation form, as any violation of the underlying restrictions will alter it. For instance, if the true individual function is nonlinear, then in general the macro relation between averages will differ in form from the individual function, with the true macro relation form heavily dependent on the actual distribution of individual variables.

Often there are available cross section data - individual data on the components of the averages - for one or more time periods of the study. If these data represent a random sampling of the population, then in principle both the micro behavioral relation and the underlying population distribution can be empirically characterized. However, this process is likely to be imprecise, leaving large portions of the observed data configuration unexplained.[3]

The initial purpose of this paper is to discover when simple statistical analysis applied to cross section data - namely ordinary least squares (OLS) regression analysis - can reveal partial information about the true macro relation without recourse to specific micro functional form or distribution form assumptions. We find that the slope coefficients from an OLS regression on cross section data will consistently estimate the first derivatives of the true macro function if and only if a certain property holds, called asymptotic sufficiency (AS) of the average predictor variables for the average dependent variable. This is shown in Section 3, after the notation and basic assumptions are given in Section 2.

Because of the importance of aggregation theory in the consistent formation of a macro function, we next investigate the relation of AS to the two major blocks of aggregation assumptions appearing in the literature; the linear (exact and consistent) aggregation approaches of economics[4] and the theory of sufficiency in the statistical literature.[5] These approaches are reviewed in conjunction with some illustrative examples in Section 4. We find that the AS property contains both types of aggregation assumptions as special cases, and thus AS represents a generalized aggregation condition. Next a characterization theorem for AS is proven which shows linear aggregation (which uses only functional form assumptions) and sufficient statistics (which uses only distribution assumptions) as polar cases under which AS holds, with intermediate cases showing the trade-offs required between making distribution and functional form assumptions under AS.

When the average predictor variables are sufficient statistics for the parameters of the underlying distribution, the true macro function can be nonlinear in form. In Section 5 we present a methodology for estimating all higher order derivatives of the true macro relationship from cross section moments, when the distribution is a member of the exponential family. We present explicitly the formulae for second order derivatives. Finally, these formulae give rise to a general test of linear aggregation approaches, relying only on the existence of certain population moments.

In short, this paper investigates the use of simple statistical techniques as applied to cross section data to characterize the true macro relation, termed "statistical aggregation analysis" in the title. These techniques provide information about the macro function based on relatively weak assumptions, which can either be used to judge specific modeling assumptions or pooled with averaged data in a joint estimation process.

2. PRELIMINARIES

For a complete discussion of the issues addressed in this paper, a very general specification of the population structure underlying a macro relation is required. However, in order to direct attention to the distributional influences on macro relations, which provide the focus of results in Section 3.1, we present slightly simplified background assumptions and notation. In Section 3.2 these assumptions are relaxed and the results reinterpreted.

We begin by assuming that there is a large population of individuals in T time periods with periods indexed by $t=1,\ldots,T$. There are N^t individuals in period t, indexed by $n=1,\ldots,N^t$. For each agent n in period t, there is a vector of personal attributes A_n^t. For given t, A_n^t is assumed to capture all differences in individual agents, whether observable or not. Also, for each agent n in period t there is a dependent quantity x_n^t, which is determined by A_n^t via

$$x_n^t = f(A_n^t) \tag{2.1}$$

f, the individual behavioral relation, is assumed here to not vary with t, an unnecessary restriction which is relaxed in Section 3.2.

Now for each t the set $\{A_n^t | n=1,\ldots,N^t\}$ may be considered as a random sample from a distribution with density $p(A|\theta^t)$.[6] $\theta^t = (\theta_1^t,\ldots,\theta_L^t)'$ is an L-vector of parameters which account for all changes in the underlying distribution $p(A|\theta^t)$ over time t. We denote the parameter space of θ^t as Γ, where $\Gamma = \{\theta\varepsilon R^L | p(A|\theta)$ is a density$\}$, where R^L is L-dimensional Euclidean space.

For each period t, the following average statistics are observed

$$\bar{x}^t = \frac{\sum_{n=1}^{N^t} x_n^t}{N^t} ; \qquad \bar{V}_m^t = \frac{\sum_{n=1}^{N^t} v_m(A_n^t)}{N^t}; \qquad m=1,\ldots,M \tag{2.2}$$

where $v_m(A_n^t)$, $m=1,\dots,M$ are observable functions of A_n^t. The vector $(v_1(A_n^t),\dots,v_m(A_n^t))'$ is denoted as $v(A_n^t)$ and the vector $(\bar{V}_1^t,\dots,\bar{V}_M^t)'$ as $\bar{V}^t$. Our primary interest here is in the relation between $\bar{x}^t$ and $\bar{V}^t$, referred to as the macro relation, which arises from the micro functional form f and the distributional form p. We now proceed to characterize this relation.

We first make an assumption concerning the population structure.

ASSUMPTION A1: All first and second order moments of x_n^t and $v(A_n^t)$ exist given t, and the covariance matrix of $v(A_n^t)$ is nonsingular.

As notation, denote

$$E(x|\theta^t) = \int f(A)p(A|\theta^t)\partial A = \phi(\theta^t)$$

$$E(v(A)|\theta^t) = g(\theta^t) = \mu_v^t$$

$$E((x - \phi(\theta^t))^2|\theta^t) = \sigma_{xx}^t \tag{2.3}$$

$$E((x - \phi(\theta^t))(v(A) - \mu_v^t)|\theta^t) = \Sigma_{xv}^t$$

$$E((v(A) - \mu_v^t)(v(A) - \mu_v^t)'|\theta^t) = \Sigma_{vv}^t$$

In (2.3) the means of x_n^t and $v(A_n^t)$ given t are written as functions of θ^t. In order to ascertain the large sample relationship between $\bar{x}^t$ and $\bar{V}^t$, we reparameterize $E(x|\theta^t) = \phi(\theta^t)$ in terms of μ_v^t. For this we require

ASSUMPTION A2: $L = M$, and $\mu_v^t = g(\theta^t)$ is invertible in θ^t. Moreover, the range of g, i.e. $\{g(\theta)|\theta\varepsilon\Gamma\}$ contains an open convex set $\Phi\subset R^M$, with the realized values $\mu_v^1 = g(\theta^1),\dots,\mu_v^T = g(\theta^T)$ interior points of Φ.

Assumption A2 is mainly made for convenience, and is relaxed somewhat in Section 3.2.

Performing this inversion, we can reparameterize $p(A|\theta^t)$ as $p^*(A|\mu_v^t) = p(A|g^{-1}(\mu_v^t))$, so that mean x_n^t in period t appears as

$$E(x|\theta^t) = \phi(g^{-1}(\mu_v^t)) = \phi^*(\mu_v^t) \qquad (2.4)$$

ϕ^* represents the correct large sample relationship between $\overline{x}^t$ and $\overline{V}^t$, because by the Weak Law of Large Numbers;[7]

$$\underset{N^t\to\infty}{\text{plim}}\ \overline{x}^t = \phi^*(\mu_v^t); \quad \underset{N^t\to\infty}{\text{plim}}\ \overline{V}^t = \mu_v^t \qquad (2.5)$$

so that if N^t is large,

$$\overline{x}^t \simeq \phi^*(\overline{V}^t) \qquad (2.6)$$

represents the correct macro relation between $\overline{x}^t$ and $\overline{V}^t$ over all time periods.

Our final background assumption is

ASSUMPTION A3: $\nabla_{\mu_v}\phi^*$ exists for all $\mu_v \varepsilon \Phi$ where ∇ denotes the gradient operator.

In addition to the macro data (2.2), we also observe a random sample of K agents in a particular period t°; a cross section data base. We index members of this sample by k=1,...,K, and therefore have as data $x_k^{t_\circ}$, $v(A_k^{t_\circ})$, k=1,...,K.[8] We assume that K is smaller than $N^{t_\circ}$, but still large enough to employ large sample statistical results.[9] In this paper our main concern is what can be learned from this sample about ϕ^*, the macro function. In particular, in the next section we establish necessary and sufficient conditions for the slope coefficients $\hat{b}_K$ obtained from regressing $x_k^{t_\circ}$ on $v(A_k^{t_\circ})$, k=1,...,K (and a constant) to consistently estimate the derivatives of ϕ^*; $\nabla_{\mu_v} \phi^*$. By standard methods, we have that

$$\operatorname*{plim}_{K\to\infty} \hat{b}_K = (\Sigma^{t_\circ}_{vv})^{-1} \Sigma^{t_\circ}_{xv} \tag{2.7}$$

This concludes the presentation of the basic framework and notation.

As stated in the introduction, if f and p are known, then an integration process (in principle) yields the correct macro relation ϕ^*, whose parameters could then be estimated using average data observed over time t. However, in general, f and p will not be known with certainty. Even if a form f is suggested by economic theory, ϕ^* will depend in form on the choice of p, unless f satisfies consistent (linear) aggregation restrictions.[10] In any realistic model indicating differences between individuals, A_n^t and/or $v(A_n^t)$ will be a large vector, usually making certainty about the form of f or p unwarranted.[11] Recall $p(A|\theta^t)$ is the joint distribution of all relevant individual attributes. To reiterate, the overall aim of this paper is to study the conditions under which a cross section data base, as a reflection of both f and p structures, can through simple statistical techniques provide information about ϕ^*.

Our general notation provides for a distinction between the underlying behavioral attributes A_n^t and the observable variables $v(A_n^t)$. If x_n^t depends directly on $v(A_n^t)$ through f; i.e. there exists f^* such that $f(A_n^t) = f^*(v(A_n^t))$, then no such distinction is required. However, included in our general analysis are situations where $\bar{V}^t$ is the relevant predictor of $\bar{x}^t$ through assumptions on p only.[12] In general A_n^t in used to represent all dimensions on which individuals differ, and therefore linear models with random coefficients, standard disturbance terms, etc., can all be embodied in this framework, in addition to the predictor variables $v(A_n^t)$.

For concreteness, suppose that consumer demand is studied. x_n^t can represent the demand for a particular commodity by family n in year t, $v_1(A_n^t)$ family income, $v_2(A_n^t)$ family size and $v_3(A_n^t)$ a qualitative variable indicating whether the family has a rural residence. $\bar{x}^t$ is average quantity demanded in year t, $\bar{V}_1^t$ average income, $\bar{V}_2^t$ average family size and $\bar{V}_3^t$ the percentage of families with rural residences.[13] Our framework covers both aggregation schemes where x_n^t is functionally related to $v(A_n^t)$ or where $\bar{V}^t$ describes movements in the underlying distribution sufficiently to determine $\bar{x}^t$ movements over time.

3. MICRO REGRESSIONS AND MACRO FUNCTIONS

3.1 The Basic Results

In this section we characterize the conditions under which the micro slope coefficients $\hat{b}_K$ of (2.7) will consistently estimate the first derivatives of the macro function ϕ^* with respect to μ_V. For the majority of this section we consider only the time period t_o, and so the time superscripts are omitted.

Of central importance to this inquiry is the conditional expectation of $\bar{x}$ given $\bar{V}$, denoted $\tilde{x}$

$$\tilde{x} = E(\bar{x}|\bar{V}) \qquad (3.1)$$

In general, $\tilde{x}$ is a function of 2M+1 arguments, $\bar{V}$, μ_V and N, so that we write

$$\tilde{x} = \tilde{x}(\bar{V}, \mu_V, N) \qquad (3.2)$$

$\tilde{x}$ is required to obey some regularity conditions, as summarized in

ASSUMPTION A4: $\tilde{x}$ exists and is continuous and differentiable in $\bar{V}$, and $\nabla_{\bar{V}}\tilde{x}$[14] approaches a finite limit $G(\mu_V) \neq 0$ as N approaches infinity and $\bar{V}$ approaches μ_V.

We can obtain the following result concerning the large sample behavior of $\bar{x}$, $\bar{V}$, and $\tilde{x}$

Lemma 3.1:

a) Under Assumptions A1 and A2, we have that as N increases

$$\operatorname*{plim}_{N\to\infty} \bar{x} = \phi^*(\mu_v); \qquad \operatorname*{plim}_{N\to\infty} \bar{V} = \mu_v$$

and that the asymptotic distribution of

$$\sqrt{N} \begin{pmatrix} \bar{x} - \phi^*(\mu_v) \\ \bar{V} - \mu_v \end{pmatrix}$$

is multivariate normal with mean zero and variance covariance matrix

$$\begin{pmatrix} \sigma_{xx} & \Sigma_{xv}' \\ \Sigma_{xv} & \Sigma_{vv} \end{pmatrix}$$

b) Under Assumptions A1, A2 and A4, as $N\to\infty$

$$\sqrt{N}\,(\tilde{x}(\bar{V},\mu_v N) - \phi^*(\mu_v))$$

converges in distribution to

$$\sqrt{N}\,(\bar{V} - \mu_v)' G(\mu_v)$$

Proof: Part a) is a standard application of the Weak Law of Large Numbers and the Central Limit Theorem.[15] Part b) is shown in the Appendix.

QED.

We are now in a position to show the first important result:

Theorem 3.2: Consider the micro slope coefficients $\hat{b}_K$ obtained by regressing x_k on $v(A_k)$ (and a constant) in a cross section random sample. Under Assumptions A1, A2, and A4, we have $\operatorname{plim}_{K\to\infty} \hat{b}_K = G(\mu_v)$.

Proof: Multiply $\sqrt{N}\,(\bar{x} - \phi^*(\mu_v))$ by $\sqrt{N}\,(\bar{V} - \mu_v)$ and take the expectation, giving

$$E(N\,(\bar{x} - \phi^*(\mu_v))(\bar{V} - \mu_v)) = \Sigma_{xv}$$

which expands as

$$\Sigma_{xv} = E(N(\tilde{x} - \phi^*(\mu_v))(\bar{V} - \mu_v)) + E(N(\bar{x} - \tilde{x})(\bar{V} - \mu_v))$$

$$= E(N(\tilde{x} - \phi^*(\mu_v))(\bar{V} - \mu_v))$$

where the second term vanishes by first conditioning on $\bar{V}$ and then taking the overall expectation. We also clearly have that

$$E(N(\bar{V} - \mu_v)(\bar{V} - \mu_v)') = \Sigma_{vv}$$

Applying Lemma 3.1 b), we obtain the equality

$$\lim_{N\to\infty} E(N(\tilde{x} - \phi^*(\mu_v))(\bar{V} - \mu_v))$$

$$= \lim_{N\to\infty} E(N(\bar{V} - \mu_v)(\bar{V} - \mu_v)')G(\mu_v)$$

or, in view of the above developments

$$\Sigma_{xv} = \Sigma_{vv} G(\mu_v)$$

which, from (2.7) and the assumption that Σ_{vv} is nonsingular, gives

$$\operatorname{plim} \hat{b}_K = G(\mu_v)$$

QED.

From applying results of the Central Limit Theorem, we have just shown that OLS slope coefficients from a randomly sampled cross section will consistently estimate the large sample derivatives of the "average" regression function $\tilde{x} = E(\bar{x}|\bar{V})$ with respect to $\bar{V}$.[16] This is a very general result, relying only on the regularity properties of Assumptions A1, A2 and A4, which concern $\tilde{x}$ and the population distribution p.[17]

In order to relate this result to the derivatives of ϕ^*, we begin by noting the pointwise convergence of the function $\tilde{x}$ to ϕ^* implicit in Lemma 3.1 b):

$$\lim_{N\to\infty} \tilde{x}(\mu_V,\mu_V,N) = \phi^*(\mu_V) \tag{3.3}$$

where the argument $\bar{V}$ has been set to μ_V. Theorem 3.2 relates the regression coefficients to the large sample derivatives of $\tilde{x}$ with regard to the first argument only. Because of this we must be very specific about the role of $\bar{V}$, the first argument in $\tilde{x}$. To this end we introduce an M vector of dummy arguments ψ and rewrite $\tilde{x}$ as

$$\tilde{x} = \tilde{x}(\psi, \mu_V,N)\big|_{\psi=\bar{V}} \tag{3.4}$$

This allows us to discriminate changes in the first argument ψ as $N\to\infty$ from changes in the second argument μ_V, avoiding the problem of $\bar{V}$ approaching μ_V in probability as $N\to\infty$.

Using this notation, we also have

$$\nabla_{\bar{V}}\tilde{x} = \nabla_\psi\tilde{x}(\psi,\mu_V,N)\big|_{\psi=\bar{V}} \tag{3.5}$$

and pointwise convergence as in

$$\lim_{N\to\infty} \nabla_\psi\tilde{x}(\mu_V,\mu_V,N) = G(\mu_V) \tag{3.6}$$

Now, in order to remove some pathological cases from the analysis, we adopt the following assumption on $\nabla_\psi \tilde{x}$, the gradient vector with regard to the first set of arguments ψ, and $\nabla_{\mu_v} \tilde{x}$, the gradient vector with regard to the second set of arguments μ_v.

ASSUMPTION A5: $\nabla_\psi \tilde{x}$ converges uniformly[18] to a vector function $G^{**}(\psi,\mu_v)$ as $N\to\infty$. Also, $\nabla_{\mu_v} \tilde{x}$ exists and converges uniformly to a vector function $H(\psi,\mu_v)$ as $N\to\infty$.

A5 implies that $\tilde{x}$ converges to a function $\phi^{**}(\psi,\mu_v)$ as $N\to\infty$. From (3.6) and (3.3) we have that

$$\phi^{**}(\mu_v,\mu_v) = \phi^{*}(\mu_v) \tag{3.7}$$

$$G^{**}(\mu_v,\mu_v) = G(\mu_v)$$

and by the uniform convergence assumption[19]

$$\nabla_\psi \phi^{**}(\psi,\mu_v) = G^{**}(\psi,\mu_v) \tag{3.8}$$

so $$\nabla_\psi \phi^{**}(\mu_v,\mu_v) = G(\mu_v)$$

and

$$\nabla_{\mu_v} \phi^{**}(\psi,\mu_v) = H(\psi,\mu_v) \tag{3.9}$$

We can now decompose the gradient of the macro function ϕ^* with respect to μ_v (via (3.7)) as

$$\nabla_{\mu_v} \phi^{*}(\mu_v) = \nabla_\psi \phi^{**}(\mu_v,\mu_v) + \nabla_{\mu_v} \phi^{**}(\mu_v,\mu_v) = G(\mu_v) + H(\mu_v,\mu_v) \tag{3.10}$$

In view of this discussion, we have shown

Theorem 3.3: Under Assumptions A1, A2, A3, A4 and A5

$$\text{plim}\ \hat{b}_K = \nabla_{\mu_V} \phi^*(\mu_V)$$

if and only if $H(\mu_V, \mu_V) = 0$

Thus, at a given point in time, the micro regression coefficients $\hat{b}_K$ will consistently estimate the first derivatives of the macro function $\phi^*(\mu_V)$ if and only if $\nabla_{\mu_V} \phi^{**}$ vanishes, where $\nabla_{\mu_V} \phi^{**}$ is the gradient of ϕ^{**} with regard to its second set of arguments. For such slope coefficients to always consistently estimate the first derivatives of ϕ^*, we must require that $\nabla_{\mu_V} \phi^{**}$ vanish at all parameter points, i.e. that ϕ^{**} can be written without reference to its second argument μ_V. Thus there exists a function ϕ^{***} of M arguments, such that

$$\phi^{**}(\psi, \mu_V) = \phi^{***}(\psi) \qquad (3.11)$$

In view of (3.7),

$$\phi^{***}(\mu_V) = \phi^*(\mu_V) \qquad (3.12)$$

or that ϕ^{***} and ϕ^* are the same function.

This condition is important enough to merit a name (where we return to using t superscripts).

Definition 1: $\bar{V}^t$ is <u>asymptotically sufficient for determining</u> $\bar{x}^t$ if for all $\psi \varepsilon \Phi$ and $\mu_V^t \varepsilon \Phi$

$$\lim_{N^t \to \infty} \tilde{x}(\psi, \mu_V^t, N^t) = \phi^*(\psi) \qquad (3.13)$$

This property is abbreviated as AS in the rest of the exposition.

We can summarize the following discussion as

Theorem 3.4: Assume A1, A2, A3, A4 and A5. Let $\mu_v^t \varepsilon \Phi$, so that $p^*(A|\mu_v^t)$ is the population distribution in period t. OLS slope coefficients from a random sample cross section in period t consistently estimate $\nabla_{\mu_v} \phi^*$ evaluated at μ_v^t, for all $\mu_v^t \varepsilon \Phi$, if and only if AS holds, i.e. $\overline{V}^t$ is asymptotically sufficient for determining $\overline{x}^t$.

In short, AS holds if $\tilde{x}$, viewed as a function of $\overline{V}^t$, has the same functional form in a large sample as the macro relation ϕ^*, viewed as a function of μ_v^t. This condition represents a relatively strong restriction on the forms of f and/or p. However, as indicated in Section 4, AS embodies virtually all types of aggregation assumptions from the economics and statistics literatures.[20] Therefore, AS can be viewed as a generalized aggregation condition.

A small sample counterpart to AS can be defined as

Definition 2: $\overline{V}^t$ is <u>sufficient for determining</u> $\overline{x}^t$ if there exists a function $\tilde{x}^*$ of the M + 1 arguments $\overline{V}^t$ and N^t such that

$$\tilde{x}(\overline{V}^t,\mu_v^t,N^t) = \tilde{x}^*(\overline{V}^t,N^t)$$

The small sample definition requires that $\tilde{x}(\overline{V}^t,\mu_v,N^t)$ can be written without reference to μ_v^t, for all N^t. AS requires this property to hold in the limit as $N^t\to\infty$. Clearly if $\overline{V}^t$ is sufficient for determining $\overline{x}^t$, then AS holds, as well as the conclusion of Theorem 3.4.

The conditions of Definition 2 have appeared previously in the statistics literature in a slightly different context. $\overline{V}^t$ sufficient for determining $\overline{x}^t$ represents the precise condition under which the well-known Rao-Blackwell Theorem holds,[21] which states that $\tilde{x}$ is the best[22] unbiased estimator of $E(\overline{x}^t) = \phi^*(\mu_v^t)$ based on $\overline{V}^t$. AS guarantees that $\tilde{x}$ will converge to such a best estimator as $N^t\to\infty$.

We now turn to two extensions of the basic notation, with the reinterpretation of Theorem 3.4 under each.

3.2 Extensions

There are two extensions of the basic framework which are of interest to empirical uses of the AS property. The first is to allow for the behavioral function f to vary over time. The second is to allow for more distributional parameters than average statistics (L>M). These extensions are discussed formally below and illustrated in the examples of Section 4.

Suppose first that the behavioral function f varies over time, as indicated by a vector of parameters γ^t. Thus, f is rewritten as

$$x_n^t = f(A_n^t, \gamma^t) \tag{3.14}$$

extending the previous notation to include γ^t. From the development of Section 3.1, we see that all functions deriving from expectationsof x_n^t or $\overline{x}^t$ will now depend on γ^t (i.e. $\phi, \phi^*, \nabla_{\mu_v} \phi^*, \tilde{x}, G, G^{**}, H, \phi^{**}, \phi^{***}$ and $\tilde{x}^*$). In particular, the macro function ϕ^* now depends on both μ_v^t and γ^t, with μ_v^t representing distribution parameters and γ^t representing behavioral parameters. The defining condition (3.13) of AS is replaced by

$$\lim_{N^t \to \infty} \tilde{x}(\psi, \mu_v^t, N^t, \gamma^t) = \phi^*(\psi, \gamma^t) \text{ for all } \psi, \mu_v^t \varepsilon \Phi \tag{3.15}$$

where each list of arguments is extended to reflect dependence on γ^t.

Under this additional consideration all of the results given above hold, with the proviso that the γ argument in all functions is held constant at $\gamma^t = \gamma^{t_\circ}$, the behavioral parameters for the period $t_\circ$ of the cross section. Theorem 3.4 is now stated as: given asymptotic sufficiency of $\overline{V}^t$ for $\overline{x}^t$ (using condition (3.15)), the slope coefficients from a cross section random sample at time t_o will consistently estimate $\nabla_{\mu_v} \phi^*$, evaluated at both $\mu_v^{t_o}$ and γ^{t_o}.

This extension is of interest to actual empirical uses of these results because there are often variables common to all micro agents which modify their behavior (e.g. common prices, general economic conditions, etc.). ϕ^* must be modeled with regard to both distributional influences (μ_v^t) and common parameter influences (γ^t). Here OLS slope estimates from a cross section can be used to estimate the derivatives of ϕ^* with regard to distributional variables in a given time period, and thus can be used either to judge restrictive assumptions on the form of ϕ^* or pooled with average time series data for more precise estimation of ϕ^*. In this way, if cross section random samples are available for several time periods, slope estimates from each data base can be used to indicate structural changes in ϕ^*, and thus guide the choice of a model consistent with all available evidence. Similarly, multiple sets of estimates can be pooled in the estimation of such a model. In addition, this extension is important in consideration of exact aggregation models, which are reviewed in Section 4.

For the second extension, assume that f is not changing over time (f is given in (2.1)), but that θ^t is an L vector, L>M, where M is the number of average statistics $\bar{V}_m^t$, m=1,...,M. The inversion $\mu_v^t \leftrightarrow \theta^t$ (Assumption A2) is now performed with regard to M elements of θ^t, conditional on the value of the remaining L - M parameters, denoted θ_o^t.[23] This implies that all functions deriving from expectations using the μ_v^t, θ_o^t parameterization will depend explicitly on θ_o^t (i.e. p^*, ϕ^*, $\tilde{x}$, G, G^{**}, H, ϕ^{**} and $\tilde{x}^*$).

In the same fashion as the first extension, all the results of Section 3.1 hold, with the proviso that θ_o^t is held constant. In particular, if the defining equation (3.13) of AS is replaced by

$$\lim_{N^t \to \infty} \tilde{x}(\psi, \mu_v^t, N^t, \theta_o^t) = \phi^*(\psi, \theta_o^t) \text{ for all } \psi, \mu_v^t \varepsilon \Phi$$

where θ_o^t has been appended to the lists of arguments, then Theorem 3.4 states that OLS slope coefficients from a cross section at time t_o will estimate the partial derivatives of ϕ^* with respect to μ_v, holding θ_o constant at $\theta_o^{t_o}$. This extension is of empirical use when certain distributional characteristics have been observed as constant over time,[24] as the modeling process can embody this constancy.

A word of caution is required for uses of distributional constancy, however, as the choice of θ_o^t (the parameters held constant) is crucial to AS. In other words, a particular choice of L-M parameters θ_o^t may cause a violation of (3.16). This situation is illustrated by example 2 of Section 4. In short, the validity of AS depends on which set of distributional parameters is assumed constant.

As a practical matter, this problem is of small import when $\overline{V}^t$ represents all available distribution data over time. All results must be necessarily prefaced by "holding all unobserved distribution parameters constant." Although not always explicitly stated, this is a requirement of virtually all empirical studies of macro functions.

However, this consideration does point out two ways OLS slope regression coefficients from a cross section can fail to describe the macro function. First is the failure of AS, with $H \neq 0$, giving for large N^t that $\tilde{x}$ has a different functional relationship to $\overline{V}^t$ than ϕ^* does to μ_v^t. The second is when there are additional distribution parameters θ_o^t which vary over time, influence the mean of x_n^t, and are not captured by $\overline{V}^t$ movements.[25]

4. EXAMPLES AND THE RELATION TO PREVIOUS AGGREGATION APPROACHES

This section presents examples which illustrate the main theorems and notation, and connects the results here to previous aggregation approaches.

Example 1: Here $f(A_n^t)$ is assumed to be a linear function in $v(A_n^t)$; i.e.

$$x_n^t = a_\circ + a' v(A_n^t) + \varepsilon(A_n^t) \tag{4.1}$$

where $a_\circ$ is a constant, a is a M-vector of constants and $\varepsilon(A_n^t)$ is a residual,[26] with mean 0 and uncorrelated with $v(A_n^t)$.

Under our assumptions we have:

$$\phi^*(\mu_v^t) = E(x_n^t) = a_\circ + a'\mu_v^t$$

$$\nabla_{\mu_v}\phi^* = a$$

$$\tilde{x} = E(\bar{x}^t|\bar{V}^t) = a_\circ + a'\bar{V}^t$$

$$\nabla_{\bar{V}}\tilde{x} = a, \quad G(\mu_v^t) = \operatorname*{plim}_{N^t\to\infty} \nabla_{\bar{V}}\tilde{x} = a$$

$$\phi^{**}(\psi,\mu_v) = a_\circ + a'\psi$$

$$G^{**}(\psi,\mu_v) = \nabla_\psi\phi^{**} = a$$

$$H(\psi,\mu_v) = \nabla_{\mu_v}\phi^{**} = 0$$

Clearly the OLS slope coefficients from a cross section will consistently estimate a, either by usual least squares theory or by our general development. The linear functional form (4.1) eliminates all distribution parameters in ϕ^* other than μ_v^t.

This simple linear functional form has appeared in two extended forms in the economics literature.[27] The first is the exact aggregation format of of Gorman (1953), Muellbauer (1975,1977) and Lau (1980), where the constant coefficients $a_\circ$ and a are allowed to be time varying with respect to a common set of parameters γ^t.[28] Thus, $a_\circ$ and a of (4.1) are replaced by $a_\circ(\gamma^t)$ and $a(\gamma^t)$ to give

$$x_n^t = a_\circ(\gamma^t) + a(\gamma^t)' v(A_n^t) + \varepsilon(A_n^t) \tag{4.2}$$

Our results show that slope coefficients from a cross section at time $t = t_\circ$ will consistently estimate $a(\gamma^{t_\circ})$.

The form (4.2) arises from the existence of aggregate macro functions which are independent of the underlying distribution form. More specifically, Lau(1980) proved the following important and general theorem, summarized in our notation as: Suppose that for all underlying configurations of $\{A_n^t, n=1,\ldots,N^t\}$, $\bar{x}^t$ can be written as

$$\bar{x}^t = F(\gamma^t,\ g_1\ (A_1^t,\ldots,A_{N^t}^t),\ldots,g_M(A_1^t,\ldots,A_{N^t}^t)) \tag{4.3}$$

where g_m, $m=1\ldots,M$ are symmetric functions of $A_1^t,\ldots,A_{N^t}^t$. Then, under some general conditions we must have

$$\begin{aligned} &\text{i)} \quad g_m(A_1^t,\ldots,A_{N^t}^t) = \bar{V}_m^t, \quad m=1,\ldots,M \\ &\text{ii)} \quad x_n^t = a_\circ(\gamma^t) + a(\gamma^t)' v(A_n^t) \\ &\text{iii)} \quad \bar{x}^t = a_\circ(\gamma^t) + a(\gamma^t)' \bar{V}^t \end{aligned} \tag{4.4}$$

With no distributional restrictions, the form (4.3) requires the symmetric functions g_m to be averages, and $x_n^t = f(A_n^t)$ must be a linear function (with constant coefficients given t) in the components of the g_m functions. Thus, a linear function is required for aggregation schemes free of distribution restrictions.

The second extension of the simple linear model (4.1) is the consistent aggregation approach of Theil (1953, 1975), where the fixed coefficients $a_\circ$, a are replaced by coefficients which vary randomly across the population, independently of the predictor variables $v(A_n^t)$, and have constant means over time. Thus, (4.1) is extended as

$$x_n^t = f(A_n^t) = a_\circ(A_n^t) + a(A_n^t)'v(A_n^t) + \varepsilon(A_n^t) \tag{4.5}$$

where $a_o(A_n^t)$ is a scalor random variable and $a(A_n^t)$ is a random M-vector which both vary independently of $v(A_n^t)$.[29] Denoting the (constant) coefficient means as $\alpha_o = E(a_o(A_n^t)|\theta^t)$ and $\alpha = E(a(A_n^t)|\theta^t)$ gives[30]

$$\begin{aligned}
&\phi^*(\mu_v^t) = \alpha_\circ + \alpha'\mu_v^t \\
&\nabla_{\mu_v}\phi^* = \alpha \\
&\tilde{x} = \alpha_\circ + \alpha'\overline{v}^t \qquad \nabla_{\overline{v}}\tilde{x} = \alpha \ , \ G(\mu_v^t) = \alpha \\
&\phi^{**}(\psi,\mu_v) = \alpha_\circ + \alpha'\psi \\
&G^{**}(\psi,\mu_v) = \nabla_\psi\phi^{**} = \alpha \\
&H(\psi,\mu_v) = \nabla_{\mu_v}\phi^{**} = 0
\end{aligned} \tag{4.6}$$

Thus OLS slope coefficients from a cross section will consistently estimate α, the mean of the marginal coefficient distribution. This framework embodies two types of assumptions, the linear functional form assumption of (4.5), and the partial distribution assumption that $a(A_n^t)$ varies independently of $v(A_n^t)$.

Exact and consistent aggregation formats can easily be combined into a general linear model, allowing random coefficients which vary independently of $v(A_n^t)$, and whose means vary over time. This specification is given as

$$x_n^t = a_\circ(A_n^t) + a(A_n^t)'v(A_n^t) + \varepsilon(A_n^t) \tag{4.7}$$

where $\alpha_o(\theta^t) = E(a_o(A_n^t)|\theta^t)$ and $\alpha(\theta^t) = E(a(A_n^t)|\theta^t)$ are time varying. If α_o and α do not vary with μ_v^t,[31] then AS holds, and our previous arguments establish the validity of our previous theorems.

Our next example illustrates where AS holds but does not rely on linearity of f in $v(A_n^t)$.

Example 2: Suppose for simplicity that $x(A_n^t) = A_n^t$, a scalar random variable distributed normally with mean μ^t and variance σ_o^2 at time t, where σ_o^2 is constant over t. Suppose that the true functional relationship is quadratic in A_n^t.

$$x_n^t = f(A_n^t) = a_o + a_1 A_n^t + a_2 (A_n^t)^2 \tag{4.8}$$

where a_o, a_1 and a_2 are constants. Using normality we have (with $\overline{V}^t = \frac{\Sigma A_n^t}{N^t}$)

$$\phi^*(\mu^t) = a_o + a_1\mu^t + a_2(\mu^t)^2 + a_2\sigma_o^2$$

$$\nabla_\mu \phi^* = a_1 + 2a_2\mu^t$$

$$\tilde{x} = E(\overline{x}^t|\overline{V}^t) = a_o + a_1\overline{V}^t + a_2(\overline{V}^t)^2 + a_2 \frac{N^t - 1}{N^t} \sigma_o^2$$

$$\nabla_{\overline{V}}\tilde{x} = a_1 + 2a_2\overline{V}^t$$

$$G(\mu^t) = \operatorname*{plim}_{N^t\to\infty} \nabla_{\overline{V}}\tilde{x} = a_1 + 2a_2\mu^t$$

$$\phi^{**}(\psi,\mu^t) = a_o + a_1\psi + a_2\psi^2 + a_2\sigma_o^2$$

$$G^{**}(\psi,\mu^t) = \nabla_\psi\phi^{**} = a_1 + 2a_2\psi$$

$$H(\psi,\mu^t) = \nabla_\mu\phi^{**} = 0$$

Our results state that regressing $x_k^{t_o}$ on $A_k^{t_o}$ in a cross section gives a slope coefficient which consistently estimates $\nabla_\mu \phi^*$.[32]

This example illustrates the second extension of Section 3.2. Here there are two distributional parameters μ^t and σ_o^2, and AS holds considering σ_o^2 held constant. If σ_o is not constant (denoted σ_o^t) then $\nabla_\mu \phi^*$ only gives the partial derivative of ϕ^* with respect to μ^t only, and thus captures only a part of the change of ϕ^* from distribution movements. Alternatively, if $\tau = \sigma_o^t/\mu^t$ (the coefficient of variation) is the distributional aspect held constant, then reparameterizing the normal distribution in terms of μ^t, τ instead of μ^t, $(\sigma_o^t)^2$ gives

$$\phi^*(\mu^t,\tau) = a_o + a_1\mu^t + a_2(\mu^t)^2 + a_2\tau^2(\mu^t)^2$$

$$\nabla_\mu \phi^* = a_1 + 2a_2\mu^t + 2a_2\tau^2\mu^t$$

$$\tilde{x} = a_o + a_1(\overline{V}^t) + a_2(\overline{V}^t)^2 + a_2\tau^2(\mu^t)^2\,\frac{N^t - 1}{N^t}$$

$$\nabla_{\overline{V}}\,\tilde{x} = a_1 + 2a_2\overline{V}^t$$

$$\phi^{**}(\psi,\mu^t) = a_o + a_1\psi + a_2\psi^2 + a_2\tau^2(\mu^t)^2$$

$$G(\psi,\mu^t) = \nabla_\psi \phi^{**} = a_1 + 2a_2\psi$$

$$H(\psi,\mu^t) = 2a_2\tau^2\mu^t$$

Here AS does not hold, and thus the conclusion of Theorem 3.4 is invalid.[33] Thus, AS depends on exactly which distributional aspects are assumed constant.

The power of the results in example 2 (with σ^2 constant) arises from the normality assumption on A_n^t. In particular, $\overline{V}^t = \frac{\Sigma A_n^t}{N^t}$ is a sufficient statistic for μ^t in the usual statistical sense. More generally, the observed

M vector $\overline{V}^t$ is sufficient for θ^t if the conditional distribution of $A_1^t,\ldots,A_{N^t}^t$ given $\overline{V}^t$ is independent of θ^t: formally if $\overline{P}$ represents the conditional distribution:

$$\overline{P}(A_1^t,\ldots,A_{N^t}^t|\overline{V}^t = \psi) = \frac{\prod_{n=1}^{N^t} p(A_n^t|\theta^t)}{P(\psi,\theta^t)} \; ; \quad \overline{V}^t = \psi$$
$$= 0 \qquad\qquad \overline{V}^t \neq \psi$$

where $P(\overline{V}^t,\theta^t)$ is the marginal distribution of $\overline{V}^t$, then $\overline{V}^t$ is a sufficient statistic for θ^t if $\overline{P}$ does not depend on θ^t.[34] Clearly in this case $\tilde{x} = E(\overline{x}^t|\overline{V}^t)$ depends only on $\overline{V}^t$, and so $\overline{V}^t$ is sufficient for determining $\overline{x}^t$ as in Definition 2 of Section 3. In this case AS holds for an arbitrary micro relation f in accordance with Assumptions A1-A5.

The theory of sufficient statistics is motivated by the question of when a particular set of statistics captures all of the information from a sample relevant to the distributional parameters θ^t. As such, it is a theory of aggregation in the same sense as the linear exact aggregation theory of economics.[35] A major theorem in the statistical literature proven by Koopman (1936), Darmois (1935) and Pitman (1936) states under some regularity conditions that a sufficient statistic $\eta(A_1^t,\ldots,A_{N^t}^t)$ for θ^t of dimension $M < N^t$ exists if and only if

$$\eta(A_1^t,\ldots,A_{N^t}^t) = \sum_{n=1}^{N^t} v(A_n^t) = N^t\overline{V}^t \tag{4.10}$$

i.e. η is a sum of functions of the individual A_n^t and the distribution $p(A|\theta^t)$ has the form[36]

$$p(A|\theta^t) = C(\theta^t)h(A)\exp\left[\sum_{m=1}^{M} \pi_m(\theta^t)v_m(A)\right] \tag{4.11}$$

Distributions of the form (4.11) comprise the exponential family of distributions.[37] Notice the similarity of exact aggregation, requiring a linear f structure, and sufficient statistics, requiring a linear structure for ln p as in (4.11).

In the discussion of examples 1 and 2 above, we have reviewed two sets of aggregation assumptions which embody completely different restrictions on the individual function f and the distribution p. The first is exact aggregation, which requires f to be a linear function of $v(A_n^t)$, with no explicit distribution assumptions. The second uses sufficient statistics, in requiring p to be such that $\overline{V}^t$ is a sufficient statistic for θ^t, with no explicit assumptions on f. Both of these sets of aggregation assumptions imply AS, and therefore asymptotic sufficiency of $\overline{V}^t$ in determining $\overline{x}^t$ can be viewed as a generalized aggregation assumption. In addition, exact aggregation and sufficient statistics represent polar extremes under which AS holds, as shown by the following theorem and corollary:

Theorem 4.1: Under Assumptions A1, A2, A4, A5 and the regularity conditions presented in the Appendix, we have

$$\nabla_{\mu_v} \tilde{x}(\psi,\mu_v^t,N^t) = E(\xi^t\delta^t|\overline{V}^t = \psi)$$

where

$$\delta^t = \overline{x}^t - \tilde{x}(\psi,\mu_v^t,N^t)$$

$$\xi^t = \sum_{n=1}^{N^t} \nabla_{\mu_v} \ln p^*(A_n^t|\mu_v^t) - E(\sum_{n=1}^{N^t} \nabla_{\mu_v} \ln p^*(A_n^t|\mu_v^t)|\overline{V}^t = \psi)$$

Also

$$H(\psi,\mu_v^t) = \lim_{N^t\to\infty} E(\xi^t\delta^t|\overline{V}^t = \psi)$$

Proof: See the Appendix.

The regularity conditions referred to in the statement of Theorem 4.1 just insure that derivative operators may be passed under the integral used in defining $\tilde{x}$. The following corollary is immediate.

Corollary 4.2: Under the conditions of Theorem 4.1, $\overline{V}^t$ is sufficient for determining $\overline{x}^t$ if $\overline{x}^t$ and $\sum_{n=1}^{N^t} \nabla_{\mu_v} \ln p^*(A_n^t|\mu_v^t)$ have zero covariance conditional on $\overline{V}^t$ for all $\mu_v^t \varepsilon \Phi$. AS holds if this covariance converges to zero as $N^t \to \infty$ for all $\mu_v^t \varepsilon \Phi$.

$\delta^t \equiv 0$, or $\overline{x}^t = \tilde{x}(\overline{V}^t, \mu_v^t, N^t)$, holds for an arbitrary distribution form p if and only if $f(A_n^t)$ is a linear function of $v(A_n^t)$; i.e. the conditions for exact aggregation hold. This follows by a straightforward application of Lau's Theorem (Lau (1980)). Similarly $\xi^t \equiv 0$ corresponds to the case where p is of the exponential family form (4.11), with $\overline{V}^t$ a sufficient statistic for θ^t. In this sense exact aggregation and sufficient statistics represent polar extreme sets of assumptions under AS.

Aggregation assumptions making partial functional form and distribution assumptions obey AS if and only if $\overline{V}^t$ effectively determines all interaction between $\overline{x}^t$ and the gradient of the log likelihood function $\sum_{n=1}^{N^t} \nabla_{\mu_v} \ln p^*(A_n^t|\mu_v^t)$. The zero covariance required by Corollary 4.2 thus gives the correct trade-offs between making functional form assumptions and distribution form assumptions under AS. In this way, the consistent aggregation approach of Theil relaxes the constant coefficient feature of exact aggregation models, and appends the assumption of random coefficients which vary independently of the predictor variables $v(A_n^t)$.

In order to further illustrate Theorem 4.1, consider the following example motivated by the standard errors-in-variables model:

Example 3: Suppose that

$$x_n^t = \beta u(A_n^t) + s(A_n^t)$$

and $$v(A_n^t) = u(A_n^t) + r(A_n^t)$$

where $u(A_n^t)$, $s(A_n^t)$ and $r(A_n^t)$ have independent normal distributions with $E(u(A_n^t)) = \mu^t$, $E(s(A_n^t)) = E(r(A_n^t)) = 0$, $Var(u(A_n^t)) = \sigma_u^2$, $Var(s(A_n^t)) = \sigma_s^2$, $Var(r(A_n^t)) = \sigma_r^2$, and σ_u^2, σ_r^2 and σ_s^2 are assumed constant over time. Our aim is to study $E(x_n^t|\mu^t) = \phi^*(\mu^t) = \beta\mu^t$ as a function of $E(v(A_n^t)|\mu^t) = \mu^t$. We have that

$$x_n^t = \beta v(A_n^t) - \beta r(A_n^t) + s(A_n^t)$$

and so (using normality)

$$\tilde{x}(\overline{V}^t,\mu^t,N^t) = (\beta - \beta\lambda)\overline{V}^t + \beta\lambda\mu^t$$

$$\nabla_{\overline{V}}\,\tilde{x} = \beta - \beta\lambda = G(\mu^t)$$

$$\phi^{**}(\psi,\mu^t) = (\beta - \beta\lambda)\psi + \beta\lambda\mu^t$$

$$G^{**}(\psi,\mu^t) = \nabla_\psi\phi^{**} = \beta - \beta\lambda$$

$$H(\psi,\mu^t) = \nabla_{\mu_v}\phi^{**} = \beta\lambda$$

where $\lambda = \dfrac{\sigma_r^2}{\sigma_u^2 + \sigma_r^2}$. Unless $\lambda = 0$ ($\beta = 0$ is ruled out by Assumption A4), AS does not hold. Corresponding to this is the familiar result that plim $\hat{b}_K = \beta(1 - \lambda) \neq \beta$, where $\hat{b}_K$ is the OLS slope coefficient obtained by regressing x_k^t on $v(A_k^t)$ in a cross section. In accordance with Theorem 4.1, we have

$$\delta^t = \beta\lambda(\overline{V}^t - \mu^t) - \beta\overline{r}^t + \overline{s}^t$$

and

$$\xi^t = \frac{N^t}{\sigma_u^2}(\lambda(\bar{v}^t - \mu^t) - \bar{r}^t)$$

with $\bar{r}^t$, $\bar{s}^t$ defined as the appropriate averages, and we can easily calculate

$$E(\delta^t\xi^t|\bar{v}^t = \psi) = \beta\lambda = H(\psi,\mu^t)$$

This illustrates the result of Theorem 4.1.

Consistent and exact aggregation schemes directly imply a linear macro function ϕ^* in μ_v^t. Aggregation schemes using sufficient statistics rely wholly on assumption on p, and can be consistent with both linear and nonlinear ϕ^* formulations. In the next section we show how additional cross section moments can be used to estimate the derivatives of $\phi^*(\mu_v^t)$ of all orders, when the distribution p is of the exponential family form (4.11). Through this development a general test of linearity (consistent or exact aggregation) emerges, which relies only on our basic population assumptions.

5. SUFFICIENT STATISTICS AND MACRO FUNCTIONS

In this section a methodology is presented for estimating second-order derivatives of the macro function with respect to μ_v from cross section moments, when p is a member of the exponential family (4.11). This methodology amounts to repeated application of derivatives, and extends to derivatives of ϕ^* of all orders.

We begin by adopting:

ASSUMPTION A6: p is a member of the exponential family in its natural parameterization

$$p(A|\pi^t) = C(\pi^t)h(A) \exp\left(\sum_{m=1}^{M} \pi_m^t v_m(A)\right) \tag{5.1}$$

where

$$C(\pi^t) = \left(\int h(A) \exp\left(\sum_{m=1}^{M} \pi_m^t v_m(A)\right)\right)^{-1}$$

and where θ^t has been reparameterized by $\pi^t = (\pi_1^t,\ldots,\pi_M^t)'$.

θ^t of (4.11) has been replaced in (5.1) by the coefficients $\pi_m(\theta^t)$, m=1,...,M; here considered as independent parameters. (5.1) holds without loss of generality from (4.11) if the mapping $\theta^t \to (\pi_1(\theta^t),\ldots,\pi_M(\theta^t))$ is of full rank M. Thus, Assumption A6 just eliminates constraints across $\pi_m(\theta^t)$, m=1,...,M, which, from an empirical point of view, are unnecessary at the outset.

Two useful textbook facts about the form (5.1) are:

Lemma 5.1: Under Assumption A6, the natural parameter space

$\Gamma = \{\pi^t | p(A|\pi^t)$ is a density$\}$ is convex.

Lemma 5.2: If $\psi(A_1^t,\ldots,A_{N^t}^t)$ is a function for which the integral

$$\int\cdots\int \psi(A_1^t,\ldots,A_{N^t}^t) \prod_{n=1}^{N^t} h(A_n^t) \exp\left[\sum_{m=1}^{M} \pi_m N^t \bar{v}^t_{mN}\right] \partial A_1^t,\ldots,\partial A_{N^t}^t$$

exists for all $\pi\epsilon\Gamma$, then this integral is an analytic function of π at all interior points of Γ, and derivatives of all orders with respect to $\pi\epsilon\Gamma$ may be passed beneath the integral sign (for discrete exponential families this integral is replaced by a sum.)

Proofs of these lemmae can be found in Lehmann (1959). They allow a computational method for taking derivatives of various expectations.

Recall, as in earlier sections, that we denote

$$\phi(\pi^t) = E(x^t|\pi^t)$$

and that

$$C(\pi^t) = \left[\int h(A)\exp \sum \pi_m v_m(A)\, \partial A\right]^{-1}$$

$C(\pi^t)$ appears in (5.1) as just a normalizing factor to make $p(A|\pi^t)$ a density. Both ϕ and C have some remarkable properties, however, as shown in the following lemma:

Lemma 5.3: Under Assumption A1 and A6, all derivatives of ϕ and ln C with respect to π are expressible as functions of moments of the $x^t, v(A^t)$ distribution. In particular, we have for C that

$$-\frac{\partial \ln C}{\partial \pi_m} = E(v_m(A)|\pi^t) = \mu_m^t, \quad m=1,\dots,M$$

$$-\frac{\partial^2 \ln C}{\partial \pi_m \partial \pi_{m'}} = E((v_m(A) - \mu_m^t)(v_{m'}(A) - \mu_{m'}^t)|\pi^t = \sigma_{mm'}^t, \quad m'=1,\dots,M$$

and

$$-\frac{\partial^2 \ln C}{\partial \pi_m \partial \pi_{m'} \partial \pi_\ell} = E((v_m(A) - \mu_m^t)(v_{m'}(A) - \mu_{m'}^t)(v_\ell(A) - \mu_\ell^t)|\pi^t)$$

$$= \sigma_{mm'\ell}^t, \quad m,m',\ell=1,\dots,M$$

For ϕ we have

$$\frac{\partial \phi}{\partial \pi_m} = E((x - \phi(\pi^t)) \quad (v_m(A) - \mu_m^t)|\pi^t) = \sigma_{xm}^t$$

and

$$\frac{\partial^2 \phi}{\partial \pi_m \partial \pi_{m'}} = E((x - \phi(\pi^t)) \quad (v_m(A) - \mu_m^t)(v_{m'}(A) - \mu_{m'}^t)|\pi^t)$$

$$= \sigma_{xmm'}^t, \quad m,m'=1,\dots,M$$

Proof: The first statement follows from Proposition 5.2. The formulae are obtainable by direct computation.[38]

QED.

We are primarily interested in the behavior of $\phi(\pi^t)$ with respect to changes in μ_v^t. We proceed as before to reparameterize via the mapping.

$$\mu_v^t = E(v(A)|\pi^t) = g(\pi^t) \tag{5.3}$$

In view of Lemma 5.3, this mapping is expressible as

$$\mu_v^t = -\nabla_{\pi^t} \ln C(\pi^t) = g(\pi^t) \tag{5.4}$$

We can reparameterize the distribution (5.1) in terms of μ_v^t if the mapping g is invertible; i.e. if the differential (Jacobean) matrix dg^t is non-singular. This matrix, again from Lemma 5.3, can be written as

$$dg^t = \left(- \frac{\partial^2 \ln C}{\partial\pi_m \partial\pi_{m'}} \bigg|_{\pi^t} \right) = \Sigma_{vv}^t \tag{5.5}$$

the covariance matrix of $v(A_n^t)$. Thus, under Assumptions A1 and A6, Assumption A2 is guaranteed. We therefore form

$$\pi^t = g^{-1}(\mu_v^t)$$

$$p^*(A|\mu_v^t) = p(A|g^{-1}(\mu_v^t)) \tag{5.6}$$

and $\quad \phi^*(\mu_v^t) = \phi(g^{-1}(\mu_v^t))$

Under the additional assumption A6, we can show the main result of Section 3.1 by direct computation.

Theorem 5.4: Under Assumptions A1 and A6 the gradient of ϕ^* with respect to μ_v is

$$\nabla_{\mu_v} \phi^* (\mu_v^t) = (\Sigma_{vv}^t)^{-1} \Sigma_{xv}^t$$

and so is consistently estimated by micro slope regression coefficients from a single period random sample cross section.

Proof: By the chain rule

$$\nabla_{\mu_v} \phi^* = (dg^t)^{-1} \nabla_{\pi} \phi$$

Now $(dg^t)^{-1} = (\Sigma_{vv}^t)^{-1}$ and by Lemma 5.3, $\nabla_{\pi}\phi = \Sigma_{xv}^t$.

QED.

We can similarly calculate all higher order derivatives of ϕ^* with respect to μ_v as functions of moments of the x_n^t, $v(A_n^t)$ distribution. Because these calculations increase greatly in complexity as the order of the derivatives increase, we present only the second derivative calculation. We first require some new notation to facilitate the formulae:

$\Omega_{\ell\pi\pi}^t$ denotes the M X M matrix with m, m′ element $\sigma_{\ell mm'}^t$, $\ell=1,\ldots,M$

$\Omega_{\pi\pi}^t$ denotes the M X M^2 matrix $\Omega_{\pi\pi}^t = [\Omega_{1\pi\pi}^t, \ldots, \Omega_{M\pi\pi}^t]$

Σ_{xvv}^t denotes the M X M matrix with m, m′ element $\sigma_{xmm'}^t$

and

$$D^t = (\Sigma_{vv}^t)^{-1} \Sigma_{xvv}^t (\Sigma_{vv}^t)^{-1} - (\Sigma_{vv}^t)^{-1} \, \Omega_{\pi\pi} [(\Sigma_{vv}^t)^{-1} \, \Sigma_{xv}^t \otimes (\Sigma_{vv}^t)^{-1}] \qquad (5.7)$$

We can now show

Theorem 5.5: Under Assumptions A1 and A6, the matrix of second order partial derivatives of ϕ^* with respect to μ_v evaluated at period t is given as

$$\nabla^2_{\mu_v}\phi^* = D^t$$

$$\left(\nabla^2_{\mu_v}\phi^* \text{ is the } M \times M \text{ matrix with } m, m' \text{ element } \left.\frac{\partial^2\phi^*}{\partial\mu_m\partial\mu_{m'}}\right|_{\mu_v^t}\right)$$

The proof is by direct computation, with a sketch of it presented in the Appendix.

The formula (5.7) for the second derivatives of ϕ^* is sufficiently complex to warrant illustration by a simple example. Suppose that $M = 1$, or that A_n^t is distributed according to

$$p(A|\pi^t) = C(\pi^t)h(A) \exp (\pi^t v_1(A))$$

where π^t is a scalar parameter. Here no assumption is made on the micro functional form $x_n^t = f(A_n^t)$, other than its expectation exists. We have therefore

$$E(x|\pi^t) = \phi(\pi^t) = \phi^*(\mu_1^t)$$

In accordance with Theorem 5.4, we find that

$$\frac{\partial\phi^*}{\partial\mu_1} = \frac{\partial\phi}{\partial\pi}\frac{\partial\pi}{\partial\mu_1} = \sigma_{x1}^t(\sigma_{11}^t)^{-1} = \frac{\sigma_{x1}^t}{\sigma_{11}^t} = \operatorname*{plim}_{K\to\infty} \hat{b}_K$$

where $\hat{b}_K$ is the estimated coefficient from the cross section regression $x_k^t \cong a + bv_1(A_k^t)$. Now

$$\frac{\partial^2\phi^*}{\partial\mu_1{}^2} = \frac{\partial^2\phi}{\partial\pi^2}\left(\frac{\partial\pi}{\partial\mu_1}\right)^2 + \frac{\partial\phi}{\partial\pi}\frac{\partial^2\pi}{\partial\mu_1{}^2} \tag{5.8}$$

By Lemma 5.3, we have

$$\frac{\partial^2\phi}{\partial\pi^2} = \sigma_{x11}^t \;;\quad \frac{\partial\pi}{\partial\mu_1} = \frac{1}{\sigma_{11}{}^t};\quad \frac{\partial\phi}{\partial\pi} = \sigma_{x1}^t$$

and so we must find $\frac{\partial^2\pi}{\partial\mu_1{}^2}$. Since

$$\frac{\partial \mu_1}{\partial \pi} \; \frac{\partial \pi}{\partial \mu_1} = 1$$

by differentiation with respect to μ_1 we get

$$0 = \frac{\partial^2 \mu_1}{\partial \pi^2} \frac{\partial \pi}{\partial \mu_1} + \frac{\partial^2 \pi}{\partial \mu_1{}^2} \frac{\partial \mu_1}{\partial \pi}$$

or

$$\frac{\partial^2 \pi}{\partial \mu_1{}^2} = - \frac{\frac{\partial^2 \mu_1}{\partial \pi^2} \left(\frac{\partial \pi}{\partial \mu_1} \right)^2}{\frac{\partial \mu_1}{\partial \pi}} = - \frac{\sigma_{111}{}^t}{(\sigma_{11}{}^t)^3}$$

Inserting these values into (5.8) gives

$$\frac{\partial^2 \phi^*}{\partial \mu_1{}^2} = \frac{\sigma_{x11}{}^t}{(\sigma_{11}{}^t)^2} - \frac{\sigma_{x1}{}^t \sigma_{111}{}^t}{(\sigma_{11}{}^t)^3} \tag{5.9}$$

which agrees with D^t of (5.7) for M = 1.

As we have shown, we can express the second order derivatives ϕ^* in terms of moments of the underlying exponential family population density. This holds for arbitrary micro functional forms $x_n^t = f(A_n^t)$ obeying Assumption A1. Estimating these moments by their sample counterparts in a cross section data base allows consistent estimation of $\nabla^2_{\mu_v} \phi^* = D^t$ for that time period.[39] Asymptotic inferences using these estimates are possible by standard methods.[40] Thus in particular, we can test whether ϕ^* is a linear function of μ_v^t.

The testing of linearity on the basis of D^t extends beyond the case of sufficient statistics, as shown in

Theorem 5.6: Assume that the moments defining D^t exist, and that Assumption

A1 holds. If $x_n^t = f(A_n^t)$ is of the generalized linear form (4.7), then $D^t = 0$.

Proof: See the Appendix.

Thus, asymptotic inferences on the estimate of D^t can be used to test whether a generalized linear form aggregation model is consistent with a cross section data base.[41] In particular, if $D^t = 0$ is rejected, then the generalized linear form is rejected as inconsistent with the micro data. Notice that this property relies on extremely weak underlying assumptions, namely the existence of the moments required by Assumption A1, the construction of D^t, and the application of the Central Limit Theorem to the sample moments used in estimating D^t.

6. CONCLUSION

The first major result of this paper is that micro slope regression coefficients will consistently estimate the first derivatives of the true macro relation if and only if AS holds. The AS property is seen as a generalized aggregation condition, embodying both linear aggregation assumptions and assumptions for sufficient statistics, as well as providing the relevant structure for partial functional form and distribution form assumptions.

In addition, we have shown that if the predictor averages are sufficient for the underlying population parameters, then in principle (when the population density is a member of the exponential family) one can empirically characterize macro function derivatives of all orders using cross section data, making possible a test of a linear, quadratic or some higher order nonlinear macro function. These techniques extend to provide a general test of linear aggregation schemes, such as the consistent and exact aggregation models.

The main appeal of these results is that they make possible an empirical characterization of macro functions using micro data, without restrictive modelling assumptions (besides AS). In addition even if the true macro function is linear, the independent effects of the average variables over time may be difficult to identify because of trending behavior or other data problems (referred to as multicollinearity). In this spirit, a first order approximation of the true macro relation using average and cross section data is provided by an exact aggregation model, as the estimates obtained from each data source will coincide in large samples, and allow the analyst to take advantage of the increased data input by increasing the precision of the final estimate values. Moreover, the exact aggregation scheme can easily incorporate structural change as indicated by additional cross section data sources.

The techniques given here can provide additional insight into the distributional influences on macroeconomic relations. Hopefully they will help end the practice of neglecting such issues, a practice which is now so prevalent.

Appendix: Omitted Proofs

Proof of Lemma 3.1 b)

Lemma 3.1 b) is shown as the result of combining Lemma 3.1 a) with two other propositions, the first is shown in Rao (1973) Section 6.2 a;

Lemma AP.1

Let τ_N be an M dimensional statistic $(\tau_{1N},\ldots,\tau_{MN})'$ such that the asymptotic distribution of $\sqrt{N}\ (\tau_{1N} - \gamma_1),\ldots,\sqrt{N}\ (\tau_{MN} - \gamma_M)$ is M-variate normal with mean zero and variance covariance matrix Σ_τ. Further, let $g(\tau_{1N},\ldots,\tau_{MN}, N)$ be a function which is totally differentiable in $\tau_{1N},\ldots,\tau_{MN}$, and that $\nabla_{\tau_N} g \to G \neq 0$ as both $N\to\infty$ and $\tau_N \to (\gamma_1,\ldots,\gamma_M)' = \gamma$. Then the asymptotic distribution of

$$\sqrt{N}(g(\tau_{1N},\ldots,\tau_{MN},N) - g(\gamma_1,\ldots,\gamma_M,N))$$

is the same as that of

$$\sqrt{N}(\tau_N - \gamma)'G$$

that is, normal with mean zero and variance

$$G'\Sigma_\tau G$$

Moreover, $g(\gamma_1,\ldots,\gamma_M,N)$ may be replaced in the above by $g^*(\gamma_1,\ldots,\gamma_M)$ if

$$\lim_{N\to\infty} \sqrt{N}[g(\gamma_1,\ldots,\gamma_M,N) - g^*(\gamma_1,\ldots,\gamma_M)] = 0$$

Lemma AP.2

$$\lim_{N\to\infty}\sqrt{N}\ (\tilde{x}\ (\mu_v,\mu_v,N) - \phi^*(\mu_v)) = 0$$

Proof: Fix N and consider

$$E[\sqrt{N}\ (\bar{x} - \phi^*(\mu_v)) + \sqrt{N}\ (\tilde{x}(\bar{V}_N,\mu_v,N) - \tilde{x}(\mu_v,\mu_v,N))]$$

$$= E(\sqrt{N}(\bar{x} - \tilde{x}(\bar{V},\mu_v,\bar{V},\mu_v,N))) + \sqrt{N}(\tilde{x}(\mu_v,\mu_v,N) - \phi^*(\mu_v))$$

$$= 0 + \sqrt{N}(\tilde{x}(\mu_v,\mu_v,N) - \phi^*(\mu_v))$$

Now as $N\to\infty$, the first expectation approaches zero by virtue of Lemma 3.1 a) and Lemma AP.1 applied to $\tilde{x}$. Thus

$$\lim_{N\to\infty} \sqrt{N}\ (\tilde{x}(\mu_v,\mu_v,N) - \phi^*(\mu_v)) = 0$$

QED

Applying Lemma AP.1 to $\tilde{x}$ in view of Lemma 3.1 a) and AP.2 gives Lemma 3.1 b).

Additional conditions for Theorem 4.1:

Let
$$\bar{P}(A_1^t,\ldots,A_{N^t}^t|\psi,\mu_v^t) = \frac{\Pi p^*(A_n^t|\mu_v^t)}{P(\psi,\mu_v^t)}\ ; \quad \bar{V}^t = \psi$$
$$= 0 \qquad \bar{V}^t \neq \psi$$

be the distribution of $A_1^t,\ldots,A_{N^t}^t$ conditional on $\bar{V}^t = \psi$. Let e_i be the M-vector with i^{th} component 1 and all other components 0, i=1,...,M. Assume p^* and P are differentiable with respect to each component of μ_v^t, and that the difference quotients.

i) $\frac{1}{h}(\bar{p}(\cdot|\psi,\mu_v^t + e_i h) - \bar{p}(\cdot|\psi,\mu_v^t))$

ii) $\frac{\bar{x}^t}{h}\ (\bar{p}(\cdot|\psi,\mu_v^t + e_i h) - \bar{p}(\cdot|\psi,\mu_v^t))$

are all bounded by integrable functions of $A_1^t,\ldots,A_{N^t}^t$, for $0 < |\ h\ | < h_o$.

Proof of Theorem 4.1:

The above conditions ii) allow differentiation of

$$\tilde{x}(\bar{V}^t,\mu_v^t,N^t) = E(\bar{x}^t|\bar{V}^t)$$

$$= \int \bar{x}^t \, \bar{P}(A_1^t,\ldots,A_{N^t}^t|\bar{V}^t,\mu_v^t) \; \partial A_1^t,\ldots,\partial A_{N^t}^t$$

under the integral sign, which gives

$$\nabla_{\mu_v} \tilde{x} = E(\bar{x}^t \Psi^t | V^t,\mu_v^t)$$

where

$$\Psi^t = \sum_{n=1}^{N^t} \nabla_{\mu_v} \ln p^*(A_n^t|\mu_v^t) - \nabla_{\mu_v} \ln P(\bar{V}^t|\mu_v^t)$$

Theorem 4.1 is shown if

$$E(\Psi^t|\bar{V}^t,\mu_v^t) = 0 \qquad \text{(AP.1)}$$

By condition i) above, we can differentiate

$$1 = E(1|\bar{V}^t,\mu_v^t)$$

under the integral sign, which gives (AP.1) above

QED

Proof Sketch for Theorem 5.5

Denote the components of $\pi^t = g^{-1}(\mu_v^t)$ by $g^{-1}(\mu_v^t) = (g_1^{-1}(\mu_v^t),\ldots,g_M^{-1}(\mu_v^t))$.

As in Theorem 5.4

$$\nabla_{\mu_v} \phi^* = \begin{pmatrix} \frac{\partial \phi^*}{\partial \mu_1} \\ \vdots \\ \frac{\partial \phi^*}{\partial \mu_M} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial \mu_1} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_1} \\ \vdots & & \vdots \\ \frac{\partial g_1^{-1}}{\partial \mu_M} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_M} \end{pmatrix} \begin{pmatrix} \frac{\partial \phi}{\partial \pi_1} \\ \vdots \\ \frac{\partial \phi}{\partial \pi_M} \end{pmatrix}$$

Therefore

$$\frac{\partial^2 \phi^*}{\partial \mu_i \, \partial \mu_j} = \sum_{m=1}^{M} \frac{\partial \phi}{\partial \pi_m} \; \frac{\partial^2 g_m^{-1}}{\partial \mu_i \, \partial \mu_j}$$

$$+ \sum_{m=1}^{M} \sum_{m'=1}^{M} \frac{\partial^2 \phi}{\partial \pi_m \partial \pi_{m'}} \left(\frac{\partial g_m^{-1}}{\partial \mu_i} \right) \left(\frac{\partial g_{m'}^{-1}}{\partial \mu_j} \right) \tag{AP.2}$$

The second term of the above (the double sum) is expressible in full matrix format as

$$\begin{bmatrix} \frac{\partial g_1^{-1}}{\partial \mu_1} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1^{-1}}{\partial \mu_M} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_M} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 \phi}{\partial \pi_1^2} & \cdots & \frac{\partial^2 \phi}{\partial \pi_M \partial \pi_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \phi}{\partial \pi_1 \partial \pi_M} & \cdots & \frac{\partial^2 \phi}{\partial \pi_M^2} \end{bmatrix} \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial \mu_1} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1^{-1}}{\partial \mu_M} & \cdots & \frac{\partial g_M^{-1}}{\partial \mu_M} \end{bmatrix}$$

which by Lemma 5.3 equals

$$\left(\Sigma_{vv}^{t}\right)^{-1} \Sigma_{xvv}^{t} \left(\Sigma_{vv}^{t}\right)^{-1} \qquad \text{(AP.3)}$$

giving the second term in the statement of the theorem. Now if $B = \left[b_{ij}(y)\right]$ is an MxM matrix of functions of y then we denote by D_y B the matrix

$$D_y B = \left[\frac{\partial b_{ij}(y)}{\partial y}\right]$$

The first term of (AP.2) is expressible in matrix format as

$$\left[D_{\mu_1}(dg^{-1})\ \nabla_\pi \phi, \cdot \cdot \cdot, D_{\mu_M}(dg^{-1})\ \nabla_\pi \phi\right] \qquad \text{(AP.4)}$$

Now, in order to evaluate $D_{\mu_m}(dg^{-1})$, m = 1,•••, M, we use the relation

$(dg^{-1})\ (dg) = I_M$ so that (if O_M is an MxM matrix of zeros)

$$D_{\mu_m}(dg^{-1})\ dg + dg^{-1}\left(D_{\mu_m}(dg)\right) = O_M \qquad m = 1,\ldots,M$$

or

$$D_{\mu_m}(dg^{-1}) = -dg^{-1}\left(D_{\mu_m}(dg)\right) dg^{-1}$$

Now, if $\tilde{g}_{m\pi\pi}$ denotes the MxM matrix with i, j element $\frac{\partial^2 g_m}{\partial \pi_i\ \partial \pi_j}$, we express $D_{\mu_m}(dg)$ as

$$\left[\tilde{g}_{1\pi\pi} D_{\mu_m} g^{-1}, \cdot \cdot \cdot, g_{M\pi\pi}\ D_{\mu_m} g^{-1}\right]$$

so that

$$D_{\mu_m}(dg^{-1}) = -dg^{-1}\left[\tilde{g}_{1\pi\pi}\, D_{\mu_m} g^{-1}, \cdot \cdot \cdot, g_{M\pi\pi}\, D_{\mu_m} g^{-1}\right] dg^{-1} \qquad \text{(AP.5)}$$

The proof is completed by inserting (AP.5) into (AP.4), making the associations

$$dg^{-1} = (\Sigma^t_{vv})^{-1} \quad ; \quad \nabla_\pi \phi = \Sigma^t_{xv}$$

$$\tilde{g}_{m\pi\pi} = \Omega^t_{m\pi\pi} \quad ; \quad m = 1, \cdot \cdot \cdot, M$$

$$\left[D_{\mu_1} g^{-1}, \cdot \cdot \cdot, D_{\mu_M} g^{-1}\right] = dg^{-1} = \left[\Sigma^t_{vv}\right]^{-1}$$

and rewriting the whole expression in terms of $\Omega^t_{\pi\pi}$

QED

Proof of Theorem 5.6:

For the generalized linear model (4.7), we have

$$\Sigma^t_{xv} = \Sigma^t_{vv}\,\alpha(\theta^t);$$

$$\Sigma^t_{xvv} = [\Omega_{1\pi\pi}\alpha(\theta^t), \ldots, \Omega_{M\pi\pi}\alpha(\theta^t)]$$

so

$$D^t = (\Sigma^t_{vv})^{-1}[\Omega_{1\pi\pi}\alpha(\theta^t), \ldots, \Omega_{M\pi\pi}\alpha(\theta^t)](\Sigma^t_{vv})^{-1}$$

$$- (\Sigma^t_{vv})^{-1}(\Omega^t_{\pi\pi})(\alpha(\theta^t) \otimes (\Sigma^t_{vv})^{-1})$$

since $(\Sigma^t_{vv})^{-1}\Sigma^t_{xv} = \alpha(\theta^t)$. Now, by symmetries in the construction of $\Omega^t_{\pi\pi}$, we have

$$[\Omega_{1\pi\pi}\alpha(\theta^t), \ldots, \Omega_{M\pi\pi}\alpha(\theta^t)](\Sigma^t_{vv})^{-1}$$

$$= \Omega_{\pi\pi}(\alpha(\theta^{t}) \otimes (\Sigma_{vv}^{t})^{-1})$$

by direct computation, which gives $D^{t} = 0$.

QED

FOOTNOTES

1. One of the reasons Friedman's book The Theory of the Consumption Function is so masterful is that the distributional foundation is clearly stated and investigated empirically with both macro and micro data, although not using pooled methods as advocated here. Other early works in demand analysis which estimated income elasticities from cross section data and applied them to time series analysis were Wold (1953) and various work of Stone, although these authors did not use aggregating models specifically. A recent demand application of an exact aggregation model is Jorgenson, Lau and Stoker (1979).
2. This critique applies equally well to studies of aggregate variables such as national income, total personal consumption expenditures, etc.
3. This becomes a major empirical problem when there are several predictor variables, as then the full (multivariate) distribution of underlying attributes must be characterized. Moreover, if the cross section data is available for only one time period, the underlying distribution is held constant, and so distribution movements over time cannot be captured by this process.
4. See Theil (1954, 1975), Green (1964), Gorman (1953), Muellbauer (1975, 1977) and Lau (1980).
5. The theory of sufficient statistics is presented in most standard textbooks on mathematical statistics; c.f. Lehman (1959), Ferguson (1967) or Rao (1973).
6. p may just be taken as the density of the sample distribution in the population. However, with N^t sufficiently large, p may be taken as a continuous approximation to this density. We utilize this framework in order to allow structure to be given to the population configuration $\{A_n^t | n=1,...,N^t\}$ via $p(A|\theta^t)$.

7. See Rao (1973), section 2c.3 for a statement of the Weak Law of Large Numbers.

8. Each index k of random sample has a counterpart n index in the population ($n=1,\ldots N^t$) numbering. We utilize the k indices only when discussing statistics of the cross section.

9. Typical numbers for a study of U.S. family demand behavior are $N^t = 70$ million for 1972, with a budget study of size $K = 10{,}000$.

10. See Section 4.

11. For instance, Jorgenson, Lau and Stoker (1979) differentiate individual families on the basis of 17 income and demographic variables.

12. For example, if $\overline{V}^t$ is a sufficient statistic for the distributional parameters θ^t - c.f. Section 4.

13. Variables common to all families, such as prices, can be entered as parameters of f, as in Section 3.2. If prices vary over families, they should be considered as components of $v(A_n^t)$.

14. $\nabla_{\overline{V}}\tilde{x}$ represents the gradient of $\tilde{x}$ with respect of $\overline{V}$, i.e. the M vector with i^{th} component

$$\left.\frac{\partial\tilde{x}}{\partial\overline{V}_i}\right|_{\overline{V},\mu_v,N}$$

15. Rao (1973) section 2c is an excellent reference for these theorems; also, see section 6a for some useful corollaries.

16. It is useful to point out that our underlying population assumptions give $\hat{b}_K$ a slightly different asymptotic distribution than in the standard linear model. In particular, $\sqrt{K}\,(\hat{b}_K - G(\mu_v))$ approaches a normal vector as $K\to\infty$ with mean zero and variance coveriance matrix

$$\Sigma_b = (\Sigma^{t_\circ}_{vv})^{-1}(\Sigma^{t_\circ}_{(xv)(xv)})(\Sigma^{t_\circ}_{vv})^{-1}$$

where $\Sigma^{t_\circ}_{(xv)(xv)}$ is the matrix with mm´ element

$$E[((x^{t_\circ} - \phi^*)(v_m(A^{t_\circ}) - \mu_m^{t_\circ}) - \sigma_{xm}^{t_\circ}) \cdot$$

$$((x^{t_\circ} - \phi^*)(v_{m'}(A^{t_\circ}) - \mu_{m'}^{t_\circ}) - \sigma_{xm'}^{t_\circ})]$$

Σ_b will correspond to the usual expression (i.e. $\sigma^2(\Sigma^t_{vv})^{-1}$, σ^2 is residual variance if there is a zero correlation between $\hat{u}^2$ and $(v_i(A_k^{t_\circ}) - \mu_i^{t_\circ})$ $(v_j(A_k^{t_\circ}) - \mu_j^{t_\circ})$ for i,j,...,l,...,M, where $\hat{u}_k = x_k^{t_\circ} - \bar{x}_K^{t_\circ} - (v(A_k^{t_\circ}) - \bar{V}_K^{t_\circ})'\hat{b}_K$. Use of the standard estimators may provide an adequate approximation to Σ_b if the sample counterparts to these correlations are small.

17. Suppose that x_n^t is functionally related to $v(A_n^t)$, i.e. there exists f^* such that $x_n^t = f(A_n^t) = f^*(v(A_n^t))$. A related but different question than that asked here is under what conditions will plim $\hat{b}_K = \nabla_v f^*(\mu_v^{t_\circ})$? This problem is addressed by White (1978), where relatively restrictive conditions on f^* and p are found.

18. The definition of uniform convergence can be found in Apostol (1967), p. 424 and Buck (1965), p. 180-2.

19. This standard result of analysis is available in most books on advanced calculus, c.f. Buck (1965), section 4.2 (Theorem 21 in particular).

20. The only exception known to this author is Friedman's permanent income - permanent consumption model. See example 3 of Section 4 (errors in variables) for illustration of this fact.

21. See Rao (1973), Section 5a.2 for the usual statement of the Rao-Blackwell Theorem. Note 2 (p. 321-322) verifies the property referred to here, as pointed out by Arnold and Katti (1972).

22. With respect to any convex loss function - e.g. minimum variance. See Rao (1973), p.322.

23. i.e. Assumption A2 is replaced by the full invertibility of the function

$$\mu_v^t = g(\theta^t)$$

$$\theta_o^t = \theta_o^t$$

where $\theta = (\theta_1^t, \theta_o^t)'$. Inverting gives

$$\theta_1^t = g^{-1}(\mu_v^t, \theta_o^t)$$

$$\theta_o^t = \theta_o^t$$

and so θ_1^t can be replaced by $g^{-1}(\mu_v^t, \theta_o^t)$ in forming ϕ^*.

24. For example, the stylized fact that the coefficient of variation of the U.S. log - income distribution is roughly constant.

25. If (3.16) is replaced by

$$\lim_{N^t \to \infty} \tilde{x}(\psi, \mu_v^t, N^t, \theta_o^t) = \phi^*(\psi) \quad \text{for all } \psi, \mu_v^t \varepsilon \Phi \text{ and all } \theta_o^t$$

this second problem is avoided. However, this structure is more restrictive than (3.16) in the text, and depends on the precise role of θ_o^t in p and the $(\mu_v^t, \theta_o^t) \leftrightarrow (\theta_1^t, \theta_o^t)$ reparameterization.

26. Recall that A_n^t is just used to signify dependence on the underlying distribution $p(A|\theta^t)$.

27. The basic form (4.1) represents the "perfect" aggregation conditions of Theil (1953) and Green (1964).

28. This reflects the first extension discussed in Section 3.3.

29. Theil (1953, 1975) assumes $a(A_n^t)$ uncorrelated with $v(A_n^t)$, which gives a linear macro function. AS requires that squared terms involving the components of $v(A_n^t)$ are uncorrelated with $a(A_n^t)$, and so we assume independence,

although weaker conditions may suffice.

30. The independence assumption allows the (derived) distribution of $a_o(A_n^t)$, $a(A_n^t)$ and $v(A_n^t)$ to be written as the product of the marginal distribution of $a_o(A_n^t)$, $a(A_n^t)$ and the marginal distribution of $v(A_n^t)$. We assume that the marginal distribution of $a_o(A_n^t)$ and $a(A_n^t)$ has a constant mean over time t.

31. That is, the means of the marginal coefficient distribution referred to in footnote 30 are determined by distributional parameters other than $\theta_1^t = g^{-1}(\mu_v^t, \theta_o^t)$ of footnote 23.

32. Simple specification analysis techniques verify this formally, if one estimates $x_k^{t_o} = b_o + b_1 A_k^t$ with (4.8) as the true model, then plim $\hat{b}_1 = a_1 + 2a_2\mu^t$.

33. In the notation of footnote 32, we have plim $\hat{b}_1 = a_1 + 2a_2\mu^t \neq \nabla_{\mu_v}\phi^* = a_1 + 2a_2\mu^t + 2a_2\tau^2\mu^t$.

34. For definitions and further discussions of sufficient statistics, see Lehmann (1959), Rao (1973) and Ferguson (1967).

35. Lau (1980) mentions sufficient statistics in some concluding remarks. However, his framework is not general enough to precisely describe the role of sufficient statistics in aggregation, as is done here. Actually, the sufficient statistic structure underlies the model in Houthakker (1956). This type of model, arrived at by direct integration of a behavioral function over a specific distribution, has appeared in several works, as surveyed by Fisher (1969), with a recent example MacDonald and Lawrence (1978).

36. Briefly, the regularity conditions required are that the range of variation of A_n^t does not depend on θ^t, a continuously differentiable sufficient statistic η for θ^t exists and $p(A|\theta^t)$ is continuously diffentiable in A and θ^t, plus some conditions on the dimension of possible variation in A_n^t. Under these conditions $p(A|\theta^t)$ must have the form (4.11) locally. If $p(A|\theta^t)$ is further assumed to be analytic, (4.11) is the global form of the density. For an excellent paper that proves this theorem in more generality than that needed here, see Barankin and Maitra (1963).

37. The exponential family form (4.11) is quite general. Examples of univariate distributions expressible in this form are the normal (μ,σ^2), Poisson (γ), negative binomial (r,θ), the gamma distributions and the beta distributions. Examples of multivariate distributions expressible in this form include the normal with mean μ and variance covariance matrix Σ. Distributions which are not of the form (4.1) include the uniform and Cauchy distributions. See Ferguson (1967) for more details.

38. Actually the formulae involving the first and second order derivatives of -lnC appear as an exercise in Lehmann (1959), p. 58, problem 14.

39. Here we are referring to using the method of moments for estimating D^t. A potential empirical problem with this approach is that the sample variances of high order moments can be quite large. See Kendall and Stuart (1963) p. 234 for a discussion of this problem. While D^t incorporates only third order moments, extensions of our methodology to higher order derivatives of ϕ^* will involve fourth and higher order moments, and thus the sampling variability problem of the method of moments may be more critical.

40. This is because the formulae (5.7) is a continuous and differentiable function of the moments comprising it. "Standard methods" refer to applications of theorems such as Lemma AP.1.

41. Although D^t of (5.7) is directly estimable from cross section moments, it would be useful if D^t could be related to simpler statistics, such as regression coefficients. In this sense it is easily shown that if $M = 1$, performing the micro regression

$$x_k^{t_\circ} \cong C_\circ + C_1 V_1 (A_K^{t_\circ}) + C_2 (v_1 (A_K^{t_\circ}))^2$$

gives

$$\underset{K \to \infty}{\text{plim}}\ \hat{C}_2 = \frac{\sigma_{x11}^{t_\circ} \sigma_{11}^{t_\circ} - \sigma_{111}^{t_\circ} \sigma_{11}^{t_\circ}}{\sigma_{11}^{t_\circ} \sigma_{1111}^{t_\circ} - (\sigma_{111}^{t_\circ})^2 - (\sigma_{11}^{t_\circ})^3}$$

which is proportional to (5.9), and thus provides an easily computable test of (5.9) equaling zero (although bear in mind stochastic structure differences, as in fn 16). The natural conjecture is that including all squared and cross product terms in a micro regression produces coefficients which consistently estimate D^t up to a proportion matrix. Unfortunately, proving or disproving this result is a computational mightmare, and to date the author has not solved this problem.

REFERENCES

Arnold, J.C., and Katti, S.K. (1972), "An Application of Rao-Blackwell Theorem in Preliminary Test Estimators", Journal of Multivariate Analysis, 2, 236-238.

Apostol, T.M. (1967), Calculus, Vol.1, 2nd ed., Waltham, Mass.: Blaisdell.

Barankin, E.W., and Maitra, A.P. (1963), "Generalization of the Fisher-Darmois-Koopman-Pitman Theorem on Sufficient Statistics", Sankhya, A 25, 217.

Buck, R.J. (1965), Advanced Calculus, 2nd ed., New York: McGraw-Hill.

Darmois, G. (1935), "Sur les lois de Probabilite a Estimation Exhaustive", C.R. Acad. Sci., Paris, 200, 1265.

Ferguson, T.S. (1967), Mathematical Statistics, A Decision Theoretic Approach, New York: Academic Press.

Fisher, F.M. (1969), "The Existence of Aggregate Production Functions", Econometrica, 37, 553-77.

Friedman, M. (1957), A Theory of the Consumption Function, NBER no. 63, general series, Princeton, New Jersey: Princeton University Press.

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

Green, A.J. (1964), Aggregation in Economic Analysis, Princeton, New Jersey: Princeton University Press.

Houthakker, H.S. (1956), "The Pareto Distribution and the Cobb-Douglas Function in Activity Analysis", Review of Economic Studies, 23, 27-31.

Jorgenson, D.W., Lau, L.J., and Stoker, T.M. (1979), "An Exact Aggregation Model of Demand", forthcoming in Advances in Econometrics.

Kendall, M.G., and Stuart, A. (1963), The Advanced Theory of Statistics, Vol.1, 2nd ed., Hafner Publishing Company.

Koopman, B.O., (1936), "On Distributions Admitting a Sufficient Statistic", Trans. Amer. Math Soc., 39, 399.

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

Lehmann, E.L. (1959), Testing Statistical Hypotheses, New York: Wiley.

MacDonald, R.J., and Lawrence, A.G. (March 1978), "Aggregate CES Input Demand with Polytomous Micro Demand", Econometrica, 46, 365-78.

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

Pitman, E.J.G. (1936), "Sufficient Statistics and Intrinsic Accuracy", Trans. Amer. Math Soc., 39, 399.

Rao, C.R. (1973), Linear Statistical Inference and Its Applications, 2nd ed., New York: Wiley.

Theil, H. (1954), Linear Aggregation of Economic Relations, Amsterdam: North Holland.

Theil, H. (1975), Theory and Measurement of Consumer Demand, Vol. 1, Amsterdam: North Holland.

White, H. (1978), "Using Least Squares to Approximate Unknown Regression Functions", forthcoming in International Economic Review.

Wold, H., and Jureen, L. (1953), Demand Analysis, New York: Wiley.