STATISTICAL AGGREGATION ANALYSIS:
CHARACTERIZING MACRO FUNCTIONS
WITH CROSS SECTION DATA

by

Thomas M. Stoker

WP 1085-79

October 7, 1979
CHARACTERIZATION AND FUNCTIONALIZATION WITH CROSS SECTIONS DATA

By

Loren M. Jacobs

U.S. Patent Office

October 3, 1935
Statistical Aggregation Analysis:

Characterizing Macro Functions with Cross Section Data

1. Introduction

Although econometricians have been estimating time series macro functions over average data for decades, it is rare to find studies which take into account the "averaged" character of the data. 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 is extremely limited, an average consumption-average income regression will not completely describe the true economic structure relating average consumption to the income distribution.\(^1\) In the same sense, any macro function relating average data is likely to ignore many important distributional influences.\(^2\)

To completely characterize the true population structure underlying a macro function, one must specify precisely the micro process that connects the individual component distribution over time. As will be reviewed shortly, work in aggregation theory has produced several micro model-distribution schemes which justify certain macro function formulations. But to just state such a justification is not enough, as in general any violation of the underlying assumptions will significantly alter the true form of the macro function. Moreover, all of the approaches provided by aggregation theory either involve linearity assumptions at the micro level, or very strict distribution assumptions, such as requiring a normal distribution for the population of micro variables.

The import of this discussion is that one really needs to utilize micro data to study any macro function justification which is to be taken seriously. If cross section data, or micro data from a single time period is available, then...
one can in principle completely characterize the micro distribution as well as the interactive structure of the micro variables under various sets of modeling assumptions. But even this process is likely to be very imprecise, leaving large portions of the observed data configuration unexplained. Also, because cross section data is only observed for one time period, the distribution is held constant, and so the impact of time series distributional movements may not be captured by this process.

The purpose of this paper is two-fold: first to discover the conditions under which cross section data is informative about macro functions and second, given these conditions, what information simple statistical analyses, such as least squares micro regressions, can provide. These issues are addressed without any strong micro relation or population distribution assumptions.

The process of using micro data to approximate macro functions, termed "statistical aggregation analysis" in the title, rests on the fact that a randomly sampled cross section data base will accurately mirror the true underlying population distribution. The approach discussed here is the use of simple micro statistics to empirically characterize interactive distributional structure which bears on the macro relation, without recourse to strict modeling assumptions.

Because of the prominence of linear micro relations in aggregation theory, we begin our analysis by studying the conditions under which least squares slope coefficients will accurately describe the macro relation. We find that under asymptotic sufficiency of the average explanatory variables for determining the average dependent variable, micro slope coefficients will consistently estimate the first derivatives of the true macro function.

Asymptotic sufficiency is a property which holds in virtually all of the explicit aggregation schemes which appear in the literature. In particular, one can guarantee it by assuming a linear functional relation at the micro level, with either constant coefficients across population members (exact aggregation)
or coefficients which vary independently of the micro explanatory variables (consistent aggregation). Either of these strategies implies a linear macro relation in the population averages. Alternatively, one can guarantee asymptotic sufficiency by requiring the average explanatory variables to be sufficient for the parameters of the underlying population distribution. Here the macro function is unrestricted in form, and so for this case we present a methodology for estimating and testing the values of all higher order derivatives of the macro function using cross section data.

In section 2 we begin by presenting the notation to be used as well as a brief review of the aggregation literature. Section 3 contains the major results on the use of micro regression analysis to study macro functions. Section 4 then presents elements of statistical aggregation analysis under distributional sufficiency. Finally, section 5 gives a summary and conclusion.

2. Preliminaries

2.1 Notation and Basic Framework

For our discussion we assume 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 \( A^t_n \). For given \( t \), \( A^t_n \) is assumed to capture all differences in individual agents, whether observable or not. Also, for each period \( t \) and each agent \( n \) there is a dependent quantity \( x_n^t \), which we assume is determined by \( A^t_n \) via

\[
x_n^t = f(A^t_n, \gamma^t)
\]

where \( \gamma^t \) is a vector of parameters which given \( t \) are constant over all agents, but may vary over time.

Now, for each \( t \) the set \( \{A^t_n|n=1,\ldots,N^t\} \) may be considered as a random sample from a distribution with density \( p(A|\theta^t) \). \( \theta^t = (\theta^t_1, \ldots, \theta^t_L) \) is an \( L \)-
vector of parameters which account for all changes in the underlying distribution \( p(\theta^t | \theta^t) \) over time \( t \). We denote the parameter space of \( \theta^t \) as \( \Gamma\), where \( \Gamma = \{ \theta^t \in \mathbb{R}^L | p(\Lambda | \theta^t) \text{ is a density} \} \), where \( \mathbb{R}^L \) is \( L \) dimensional Euclidean space.

We begin with this general framework in order to allow for virtually all types of assumptions on the underlying population structure. Assumptions on \( f \) of (2.1) are referred to as functional form assumptions, whereas assumptions on \( p \) are referred to as distributional assumptions.

For each period \( t \), we observe the following average statistics

\[
(2.2) \quad \overline{x^t}_N = \frac{1}{N^t} \sum_{n=1}^{N^t} x^t_n; \quad \overline{v^t}_{mN} = \frac{1}{N^t} \sum_{n=1}^{N^t} v^t_m(A^t_n), \quad m = 1, \ldots, M
\]

where \( v^t_m(A^t_n) \), \( m=1, \ldots, M \) are observable functions of the underlying attributes \( A^t_n \), with \( M \leq L \). Denote the vector \( (v^t_1(A^t_n), \ldots, v^t_M(A^t_n))' \) as \( v(A^t_n) \) and the vector \( (\overline{v^t}_{1N}, \ldots, \overline{v^t}_{MN})' \) as \( \overline{v^t}_N \). Our primary interest here is in the relation between \( \overline{x^t}_N \) and \( \overline{v^t}_N \), referred to as the macro relation.

As an example, suppose that we are studying consumer demand. Here \( x^t_n \) can represent the demand for a particular commodity by family \( n \) in year \( t \), \( v^t_1(A^t_n) \) family income, \( v^t_2(A^t_n) \) family size and \( v^t_3(A^t_n) \) a qualitative variable indicating whether the family has rural residence. Then \( \overline{x^t}_N \) represents average quantity demanded, \( \overline{v^t}_{1N} \) average income, \( \overline{v^t}_{2N} \) average family size and \( \overline{v^t}_{3N} \) the percentage of families in the economy with rural residences. Similarly, in a production context, \( x^t_n \) may represent the output of firm \( n \) in period \( t \), \( v^t_1(A^t_n) \) its capital stock and \( v^t_2(A^t_n) \) its labor input. Then \( \overline{x^t}_N \) is average firm output, \( \overline{v^t}_{1N} \) average capital stock and \( \overline{v^t}_{2N} \) average labor input. In either of these cases elements common to all agents, such as prices, are capturable through the functional form parameters \( \gamma^t \).

We make the following assumption concerning the population structure:

ASSUMPTION A.1: All first and second order moments of \( x^t_n \) and \( v(A^t_n) \) given \( t \) exist

\[
(2.3) \quad \begin{align*}
    i) \quad & E(x^t | \theta^t) = \int f(A^t, \gamma^t) p(A^t | \theta^t) \overline{x^t} = \phi(\theta^t, \gamma^t) \\
    ii) \quad & E(v^t_m(A^t_n) | \theta^t) = \mu^t_m, \quad m = 1, \ldots, M
\end{align*}
\]
and
\[
E\left[ (x_t^t - \phi(\theta^t, \gamma^t))^2 | \theta^t \right] = \sigma_{xx}^t
\]
(2.4)
\[
E\left[ (x_t^t - \phi(\theta^t, \gamma^t))(v_m(A^t) - \mu_m^t) | \theta^t \right] = \sigma_{xm}^t
\]
\[\quad m = 1, \ldots, M\]
\[
E\left[ (v_m(A^t) - \mu_m^t)(v_{m'}(A^t) - \mu_{m'}^t) | \theta^t \right] = \sigma_{mm'}^t
\]
\[\quad m, m' = 1, \ldots, M\]

and the covariance matrix of \(v(A^t_n)\):
\[
\Sigma_v^t = \begin{bmatrix}
\sigma_{11}^t & \cdots & \sigma_{M1}^t \\
\vdots & \ddots & \vdots \\
\sigma_{M1}^t & \cdots & \sigma_{MM}^t
\end{bmatrix}
\]
is nonsingular.

As notation, we collect the other moments in matrices as
\[
\mu_v^t = \begin{bmatrix} \mu_1^t \\ \vdots \\ \mu_M^t \end{bmatrix} ; \quad \Sigma_{xv}^t = \begin{bmatrix} \sigma_{x1}^t \\ \vdots \\ \sigma_{xM}^t \end{bmatrix}
\]
(2.5)

From (2.3) i), we see that the mean of \(x_n^t\) can be written as a function of the distributional parameters \(\theta^t\) and the functional form parameters \(\gamma^t\).

In order to ascertain the large sample relationship between \(x_n^t\) and \(v_N^t\), we reparameterize \(E(x^t|\theta^t) = \phi(\theta^t, \gamma^t)\) in terms of \(\mu_v^t\). Rewriting (2.3) ii) as
\[
\mu_1^t = \int v_1(A^t)p(A^t|\theta^t)\,dA^t = g_1(\theta^t)
\]
(2.6)
\[
\vdots
\]
\[
\mu_M^t = \int v_m(A^t)p(A^t|\theta^t)\,dA^t = g_M(\theta^t)
\]
or in vector format as \(\mu_v^t = g(\theta^t)\), gives \(\mu_v^t\) as a function of \(\theta^t\). We next adopt
ASSUMPTION A.2: $\mu_v^t = g(\theta_v^t)$ is invertible in $\theta_1^t, \ldots, \theta_M^t$, conditional on the value of $\theta_o^t = (\theta_{M+1}^t, \ldots, \theta_L^t)$ if $L > M$, or unconditionally if $L = M$. Moreover, we assume that the range of $g$, i.e. $\{g(\theta) | \theta \in \Gamma \}$ contains an open convex set $\phi$ in $\mathbb{R}^M$, with the realized values $\mu_v^1 = g(\theta_v^1), \ldots, \mu_v^T = g(\theta_v^T)$ interior points of $\phi$.

Performing this inversion, we can reparameterize $p(A^t | \theta^t)$ as $p^*(A^t | \mu_v^t, \theta_v^t)$, so that mean $x^t$ can be written as

$$E(x^t | \mu_v^t, \theta_v^t) = \int f(A^t, Y^t) p^*(A^t | \mu_v^t, \theta_v^t) dA^t$$

(2.7)

Our final background assumption is

ASSUMPTION A.3: $\nabla_{\mu_v^t} \phi^*(\mu_v^t, \theta_v^t, Y^t)$ exists for all $\mu_v^t \in \phi$, where $\nabla$ denotes the gradient operator.

Conditional on $\theta_v^t$, $\phi^*$ represents the correct large sample relationship between $x^t_N$ and $\nu^t_N$, because by the Weak Law of Large Numbers.

(2.8) $\lim_{n \to \infty} x^t_N = \phi^*(\mu_v^t, \theta_v^t, Y^t)$; $\lim_{n \to \infty} \nu^t_N = \mu_v^t$

so that if $N$ is large, we have

(2.9) $x^t_N \sim \phi^*(\nu^t_N, \theta_v^t, Y^t)$

In addition to the macro data (2.2), we also observe a random sample of $K$ agents in a particular period $t^0$; a cross section data base. We index members of this sample by $k = 1, \ldots, K$ and therefore have as data $x_k^{t^0}, v(A_k^{t^0}), k = 1, \ldots, K$. We assume that $K$ is smaller than $N$ but still large enough to employ large sample statistical results. In this paper our major interest is in what can be learned from these micro data about $\phi^*$, the macro function. In particular,
in the next section we study the relation of the derivatives of \( \phi^* \) to the slope coefficients \( \hat{b}_k \) which result from regressing \( x_k^{t_0} \) on \( v_1(A_k^{t_0}), \ldots, v_M(A_k^{t_0}), \) \( k = 1, \ldots, K \) by least squares. \( \hat{b}_k \) is given as the solution of the normal equations:

\[
(2.10) \quad S_{\nu\nu}^{t_0} \hat{b}_k = S_{\nu\nu}^{t_0}
\]

where \( S_{\nu\nu}^{t_0} \) is the MxM matrix with \( i,j \) element

\[
\sum_{k=1}^{K} \frac{(v_i(A_k^{t_0}) - \bar{v}_i^{t_0})(v_j(A_k^{t_0}) - \bar{v}_j^{t_0})}{K}
\]

\( S_{\nu\nu}^{t_0} \) is the Mx1 matrix with \( i^{th} \) element

\[
\sum_{k=1}^{K} \frac{(x_k^{t_0} - \bar{x}_k^{t_0})(v_i(A_k^{t_0}) - \bar{v}_i^{t_0})}{K}
\]

and \( \bar{x}_k^{t_0} = \frac{\sum_{k=1}^{K} x_k^{t_0}}{K}, \bar{v}_i^{t_0} = \frac{\sum_{k=1}^{K} v_i(A_k^{t_0})}{K} \). By standard methods, as \( K \) grows

\[
(2.11) \quad \lim_{K \to \infty} S_{\nu\nu}^{t_0} \hat{b}_k = \Sigma_{\nu\nu}^{t_0} \quad \text{or} \quad \lim_{K \to \infty} \hat{b}_k = \Sigma_{\nu\nu}^{t_0} \Sigma_{\nu\nu}^{t_0}^{-1} x_{\nu\nu}^{t_0}
\]

This concludes the presentation of the basic framework and notation.

2.2 Previous Work

We are now in a position to discuss previous approaches to the aggregation problem. As pioneered by Houthakker (1957), a variety of studies have appeared which specify both the distribution \( p \) and the functional form \( f \) exactly. The macro function is then found by direct integration. A survey of this work can be found in Fisher (1969).\(^{12}\)

Of more recent interest are linear aggregation schemes, which involve relatively weak distributional assumptions. Exact aggregation models first arose out of the work of German (1953), followed later by Muellbauer (1975, 1977).\(^{13}\)
The most important and general result motivating these models is given by
Lau (1977 a - b ), which is stated briefly as: Suppose that for all underlying
configurations of $A^t_n$, $\bar{x}_N^t$ can be written as

\[ \bar{x}_N^t = F(\gamma^t, g, A_1^t, \ldots, A_N^t), \ldots, g_M(A_1^t, \ldots, A_N^t) \]

where $g_m, m = 1, \ldots, M$ is a symmetric function of $A_1^t, \ldots, A_N^t$, then under some
general conditions we must have

\[ i) \quad g_m(A_1^t, \ldots, A_N^t) = \frac{\sum_{n=1}^{N^t} v_m(A^t_n)}{N^t} = \bar{v}_m^t, \quad m = 1, \ldots, M \]

(2.13)

\[ ii) \quad x_n^t = f(A_n^t, \gamma^t) = C(\gamma^t) + \sum_{m=1}^{M} h_m(\gamma^t) v_m(A^t) \]

and so

\[ iii) \quad \bar{x}_N^t = C(\gamma^t) + \sum_{m=1}^{M} h_m(\gamma^t) \bar{v}_m^t \]

With no distributional restrictions, the form (2.12) requires $x_n^t$ to be a linear
function in $v(A_n^t)$, with constant coefficients across the population, giving also
a linear macro relation. An additive residual can clearly be incorporated
in $x_n^t$ without disturbing the aggregation structure, and if uncorrelated with
$v(A_n^t)$, the micro slope coefficients defined in (2.10) will consistently estimate
$h_m(\gamma^t), m = 1, \ldots, M$.

Consistent aggregation approaches assume a linear micro relationship
together with coefficients that vary randomly across the population. If
the coefficient variation is uncorrelated with $v(A_n^t)$, then the macro function
is linear with coefficients equal to the means of micro coefficient distrib-
butions. If in addition, the coefficient variation is independent of $v(A_n^t)$,
then the micro slope estimator (2.10) will consistently estimate the macro
coefficients for period $t_0$. The restrictions on the coefficient variation
amount to partial distributional assumptions which allow the constant micro
coefficient feature of exact aggregation models to be relaxed.
2.3 A Word on Methodology

In consistent and exact aggregation approaches, a linear macro formulation is obtained with either partial or no distributional assumptions. Here, by viewing the agent population in a given year t as a drawing from the distribution with density \( p(A^t|\theta^t) \), we can add more structure to the configuration \( \{A^t_1, \ldots, A^t_N\} \) that exists in the population. Our posture is then to characterize the theoretical macro relation \( \Phi^* \) implicit from \( p \) and \( f \), by using an observed random sample from a cross section survey. This approximating method thus affords a more realistic basis upon which to study distributional influences on macro relations.

The emphasis on large sample techniques is required to guarantee that the observed macro data behaves in accordance with the theoretical macro function, as in (2.9). A sensible approach to aggregation in small samples must either be based on modeling expected values or must contend with all possible configurations of underlying population attributes. Modeling expected values in a small sample context is virtually equivalent to large sample modeling as is done here. As shown by the work of Lau, for an average equation to be invariant with regard to all possible configurations of attribute vectors, the model must be of the exact aggregation type. Thus it appears that discussion of small sample approaches will yield few new results. Moreover, populations of states, countries, etc., are generally quite large, so that large sample approaches are empirically justified. Various authors have argued the need for emphasis on large sample techniques, notably Theil (1954, 1975) and Green (1964).

Another issue in aggregation theory is that of recoverability of precise micro functions from macro function parameters. In an exact aggregation model, values of the macro coefficients \( h_m(\gamma^t), m=1, \ldots, M \) will dictate the values of the
micro coefficients, and thus the micro functions are recoverable from the macro function parameters. In general this property is not available in large sample approaches. Even in consistent aggregation, estimates of the macro coefficients give only the means of the micro coefficient distributions. With arbitrary underlying distribution structures, recoverability requires an exact aggregation format.

3. Micro Regressions and Macro Functions

3.1 The Main Result

In this section we show the main result that under certain conditions micro slope regression coefficients as defined in (2.10) will consistently estimate the first derivatives of the macro function \( \phi^* \) with respect to \( \mu_v \).

For the majority of this section we will consider only the time period \( t_o \), that in which the cross section data is observed, and so we omit the time superscripts.

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

\[
(3.1) \quad \tilde{x}_N = E (x_N \mid \bar{V}_N, \mu_v, \theta_o)
\]

In general, \( \tilde{x}_N \) is a function of five arguments; \( \bar{V}_N, \mu_v, \theta_o, \gamma, N \), so that we write

\[
(3.2) \quad \tilde{x}_N = \tilde{x} (\bar{V}_N, \mu_v, \theta_o, \gamma, N)
\]

We require that \( \tilde{x} \) obey some regularity properties, as summarized in

**ASSUMPTION A.4:** \( \tilde{x} \) exists and is continuous and differentiable in \( \bar{V}_N \) for all \( N \), and \( \frac{\partial}{\partial \bar{V}_N} \tilde{x}_N^{15} \) approaches a finite limit \( C (\mu_v, \theta_o, \gamma) \neq 0 \) as \( N \) approaches infinity and \( \bar{V}_N \) approaches \( \mu_v \).
We can obtain the following result concerning the large sample behavior of \( \bar{x}_N, \bar{V}_N \) and \( \bar{x} \):

**LEMMA 3.1:**

a) Under Assumptions A.1 and A.2, we have that as \( N \) increases

\[
\lim_{N \to \infty} \bar{x}_N = \phi^*(\mu_v, \theta_o, \gamma); \lim_{N \to \infty} \bar{V}_N = \mu_v
\]

and that the asymptotic distribution of

\[
\frac{1}{\sqrt{N}} \begin{pmatrix} \bar{x}_N - \phi^*(\mu_v, \theta_o, \gamma) \\ \bar{V}_N - \mu_v \end{pmatrix}
\]

is multivariate normal with mean zero and variance covariance matrix

\[
\begin{pmatrix} \sigma_{xx} & \Sigma_{xv} \\ \Sigma_{vx} & \Sigma_{vv} \end{pmatrix}
\]

b) Under Assumptions A.1, A.2 and A.4, as \( N \to \infty \)

\[
\sqrt{N} \left[ \bar{x} \left( \bar{V}_N, \mu_v, \theta_o, \gamma, N \right) - \phi^*(\mu_v, \theta_o, \gamma) \right]
\]

converges in distribution to

\[
\sqrt{N} \left( \bar{V}_N - \mu_v \right)' G \left( \mu_v, \theta_o, \gamma \right)
\]

Proof: Part a) is a standard application of the Weak Law of Large Numbers and the Central Limit Theorem.\(^{16}\) Part b) is shown in the Appendix. QED.

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

**THEOREM 3.2:** Consider the micro slope regression coefficients \( \hat{b}_K \) as defined in (2.10), obtained by regressing \( x_K \) on \( v (A_K) \) (and a constant) in a micro random sample. Under Assumptions A.1, A.2 and A.4, we have \( \lim_{K \to \infty} \hat{b}_K = G(\mu_v, \theta_o, \gamma) \).
Proof: Multiply \( \sqrt{N} (\bar{x}_N - \phi^* (\mu, \theta, \gamma)) \) by \( \sqrt{N} (\bar{v}_N - \mu_v) \) and take the expectation, giving
\[
\Sigma_{vx} = E \left( N \left( \bar{x}_N - \phi^* (\mu_v, \theta_v, \gamma) \right) (\bar{v}_N - \mu_v) \right)
\]
which expands as
\[
\Sigma_{vx} = E \left( N \left( \bar{x}_N - \tilde{x}_N - \phi^* (\mu_v, \theta_v, \gamma) \right) (\bar{v}_N - \mu_v) \right) + E \left( N \left( \bar{x}_N - \tilde{x}_N \right) \left( \bar{v}_N - \mu_v \right) \right)
\]
\[
= E \left( N \left( \bar{x}_N - \phi^* (\mu_v, \theta_v, \gamma) \right) (\bar{v}_N - \mu_v) \right)
\]
where the second term vanishes by first conditioning on \( \bar{v}_N \) and then taking the overall expectation. We also clearly have that
\[
E \left( N \left( \bar{v}_N - \mu_v \right) (\bar{v}_N - \mu_v)' \right) = \Sigma_{vv}
\]
Applying Lemma 3.1 b), we obtain the equality
\[
\lim_{N \to \infty} E \left( N \left( \bar{x}_N - \phi^* (\mu_v, \theta_v, \gamma) \right) (\bar{v}_N - \mu_v) \right) = \lim_{N \to \infty} E \left( N \left( \bar{v}_N - \mu_v \right) (\bar{v}_N - \mu_v)' \right) G (\mu_v, \theta_v, \gamma)
\]
or, in view of the above developments
\[
\Sigma_{vx} = \Sigma_{vv} G (\mu_v, \theta_v, \gamma)
\]
Now, as presented in (2.11), the normal equations defining \( \hat{b}_k \) yield that
\[
\Sigma_{vx} = \Sigma_{vv} \lim_{K \to \infty} \hat{b}_k \quad \text{which implies}
\]
\[
\lim_{K \to \infty} \hat{b}_k = G (\mu_v, \theta_v, \gamma)
\]
QED.

From applying results of the Central Limit Theorem, we have just shown that least squares slope coefficients from a randomly sampled cross section will consistently estimate the large sample partial derivatives of \( \bar{x}_N \) with respect to \( \bar{v}_N \). In order to relate this result to the derivatives of the macro function \( \phi^* \), we must be very specific about the role of \( \bar{v}_N \) in \( \bar{x}_N \). In
order to clarify the roles of $V_N$ and $\mu_v$ in $\bar{x}$, we introduce an $M$-vector of dummy arguments $\psi$ and rewrite $\bar{x}$ as

$$\tilde{x}_N = \bar{x}(\psi, \mu_v, \theta_o, \gamma, N) |_{\psi = V_N}$$

Similarly the gradient of $\tilde{x}$ with respect to $V_N$ is rewritten as

$$\nabla_{V_N} \tilde{x} = \nabla_{\psi} \tilde{x}(\psi, \mu_v, \theta_o, \gamma, N) |_{\psi = V_N}$$

As above, we have pointwise convergence of these two functions as in

$$\lim_{N \to \infty} \tilde{x}(\mu_v, \mu_v, \theta_o, \gamma, N) = \phi^*(\mu_v, \theta_o, \gamma)$$

$$\lim_{N \to \infty} \nabla_{\psi} \tilde{x}(\mu_v, \mu_v, \theta_o, \gamma, N) = G(\mu_v, \theta_o, \gamma)$$

Now, in order to remove some pathological cases from our analysis, we adopt the following assumption on the gradient vectors $\nabla_{\psi} \tilde{x}$ and $\nabla_{\mu_v} \tilde{x}$:

**ASSUMPTION A.5:** $\nabla_{\psi} \tilde{x}$ converges uniformly$^{18}$ to a vector function $G^{**}(\psi, \mu_v, \theta_o, \gamma)$ as $N \to \infty$. Also $\nabla_{\mu_v} \tilde{x}$ exists and converges uniformly to a vector function $H(\psi, \mu_v, \theta_o, \gamma)$ as $N \to \infty$.

A.5 implies that $\tilde{x}$ converges to a function $\phi^{**}(\psi, \mu_v, \theta_o, \gamma)$ as $N \to \infty$.

From (3.5) we have

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

$$\phi^{**}(\mu_v, \mu_v, \theta_o, \gamma) = \phi(\mu_v, \theta_o, \gamma)$$

and by the uniform convergence assumption$^{19}$

$$\nabla_{\psi}(\mu_v, \mu_v, \theta_o, \gamma) = G^{**}(\psi, \mu_v, \theta_o, \gamma)$$

(3.7) so

$$\nabla_{\psi}(\mu_v, \mu_v, \theta_o, \gamma) = G(\mu_v, \theta_o, \gamma)$$

and

$$\nabla_{\mu_v}(\mu_v, \mu_v, \theta_o, \gamma) = H(\psi, \mu_v, \theta_o, \gamma)$$
We can now decompose the gradient of the macro function $\phi^*$ with respect to $\mu_v$ as

$$
\nabla_{\mu_v} \phi^* (\mu_v, \theta_0, \gamma) = \nabla_{\psi} \phi^{**} (\mu_v, \mu_v, \theta_0, \gamma) + \nabla_{\mu_v} \phi^{**} (\mu_v, \mu_v, \theta_0, \gamma)
$$

$$
= G (\mu_v, \theta_0, \gamma) + H (\mu_v, \mu_v, \theta_0, \gamma)
$$

In view of this discussion, we have shown

**THEOREM 3.3:** Under Assumptions A.1, A.2, A.3, A.4 and A.5

$$
\text{plim} \sum_{k=1}^K \frac{1}{K} \nabla_{\mu_v} \phi^* (\mu_v, \theta_0, \gamma)
$$

if and only if

$$
H (\mu_v, \mu_v, \theta_0, \gamma) = 0.
$$

Thus, at a given point in time, the micro regression coefficients will consistently estimate the first derivatives of the macro function $\phi^*$ if and only if $\nabla_{\mu_v} \phi^{**}$ vanishes at that point. For such regression coefficients to always consistently estimate these first derivatives, we must require that $\nabla_{\mu_v} \phi^{**}$ vanish at all possible parameter points. In this case we can omit reference to $\mu_v$ in $\phi^{**}$, giving

$$
\phi^{**} (\psi, \mu_v, \theta_0^t, \gamma^t) = \phi^{**} (\psi, \theta_0^t, \gamma^t)
$$

(3.10)

$$
= \phi^* (\psi, \theta_0^t, \gamma^t)
$$

where we have returned to our convention of denoting the time period by a $t$ subscript.

This condition is important enough to merit a name:

**DEFINITION:** $\bar{V}_N^t$ is asymptotically sufficient for determining $\bar{x}_N^t$ conditional on $\theta_0^t$ if

$$
\lim_{N \to \infty} \bar{x} (\psi, \mu_v^t, \theta_0^t, \gamma^t, N^t) = \phi^{**} (\psi, \theta_0^t, \gamma^t)
$$

$\bar{V}_N^t$ is asymptotically sufficient for determining $\bar{x}_N$ (unconditionally)
if
\[ \lim_{N \to \infty} x(\psi, \mu_t, \theta_o, \gamma_t, N^t) = \phi^{**}(\psi, \gamma_t) \]

We can summarize the results of this section as

THEOREM 3.4: Under Assumptions A.1, A.2, A.3, A.4 and A.5 micro slope regression coefficients will always consistently estimate the first derivatives of the macro function \( \phi^* \) (holding \( \theta_o^t \) constant) if and only if \( \overline{V}_N^t \) is asymptotically sufficient for determining \( \overline{x}_N^t \) (conditional on \( \theta_o^t \)).

Similarly, small sample counterparts to asymptotic sufficiency can be defined.

DEFINITION: \( \overline{V}_N^t \) is sufficient for determining \( \overline{x}_N^t \) conditional on \( \theta_o^t \) if \( \tilde{x} \) can be written as
\[ \tilde{x}(\overline{V}_N^t, \mu_v^t, \theta_o^t, \gamma_t, N^t) = \tilde{x}(\overline{V}_N^t, \theta_o^t, \gamma_t, N^t) \]

Obviously, if \( \overline{V}_N^t \) is sufficient for determining \( \overline{x}_N^t \) (conditional on \( \theta_o^t \)) then \( \overline{V}_N^t \) is asymptotically sufficient for determining \( \overline{x}_N^t \) (conditional on \( \theta_o^t \)), and so all of our results hold if (small sample) sufficiency obtains.

We close this section by presenting a theorem and corollary which can be used under certain conditions to characterize asymptotic sufficiency as defined above.

THEOREM 3.5: Assume A.1, A.2, A.4, A.5 and let
\[ p^{**}(A_1^t, \ldots, A_N^t | V_0, \mu_v^t, \theta_o^t) = \begin{cases} 0 & \text{if } \overline{V}_N^t \neq V_0 \\ \prod p^*(A_n^t | \mu_v^t, \theta_o^t) & \text{if } \overline{V}_N^t = V_0 \\ \frac{p(V | u^t, \theta^t)}{p(V | u^t, \theta^t)} & \text{if } \overline{V}_N^t = V_0 \end{cases} \]
denote the density of $A_1^t, \ldots, A_N^t$ conditional on $\overline{V}_N^t = V_o$ where $P (\overline{V}_N^t \mid \mu_v^t, \theta_o^t)$ is the marginal density of $\overline{V}_N^t$, and let $\epsilon_i$ be the $M$-vector with $i$th component 1 and all other components zero, $i = 1, \ldots, M$. If $p^*$ and $P$ are differentiable with respect to each component of $\mu_v^t$, and the difference quotients
\[ i) \frac{1}{h} \left( p^{**} (\cdot \mid V_o, \mu_v^t + e_i h, \theta_o^t) - p^{**} (\cdot \mid V_o, \mu_v^t, \theta_o^t) \right) \]
\[ ii) \frac{x_N^t}{h} \left( p^{**} (\cdot \mid V_o, \mu_v^t + e_i h, \theta_o^t) - p^{**} (\cdot \mid V_o, \mu_v^t, \theta_o^t) \right) \]

$i = 1, \ldots, M$

are all bounded by integrable functions of $A_1^t, \ldots, A_N^t$, $V_o$ for $0 < |h| < h_o$ then

\[
\nabla_{\mu_v^t} \overline{x} (V_o, \mu_v^t, \theta_o^t, \gamma^t, N^t) = E \left( \xi^t_N \delta^t_N \mid \overline{V}_N^t = V_o \right)
\]

where

\[
\delta_N^t = \overline{x}_N^t - \overline{x}_N^t
\]

\[
\xi^t_N = \sum_{n=1}^{N} \nabla_{\mu_v^t} \ln p^* (A_n^t \mid \mu_v^t, \theta_o^t)
\]

\[- E \left\{ \sum_{n=1}^{N} \nabla_{\mu_v^t} \ln p^* (A_n^t \mid \mu_v^t, \theta_o^t) \bigg| \overline{V}_N^t = V_o \right\} \]

Proof: See the Appendix.

We immediately obtain the obvious corollary:

**Corollary 3.6:** Under the conditions of Theorem 3.5,

$\overline{x}_N^t$ is sufficient for determining $\overline{\theta}_o^t$ conditional on $\overline{V}_N^t$ if

$\overline{x}_N^t$ and $\sum_n \nabla_{\mu_v^t} \ln p (A_n^t \mid \mu_v^t, \theta_o^t)$ have zero covariance conditional on $\overline{V}_N^t$, for all parameter points. If this covariance converges in probability to zero as $N^t \to \infty$, $\overline{V}_N^t$ is asymptotically sufficient for determining $\overline{x}_N^t$ conditional on $\theta_o^t$. 

We see heuristically that micro slope coefficients are useful in characterizing the macro function \( \phi^* \) if \( V_N^t \) effectively determines all interaction between \( x_n^t \) and the gradient of the log-likelihood function \( \sum \nabla \ln p (A_n^t | \mu_v^t, \theta_o^t) \) in a large sample. As discussed shortly, this can occur by \( V_N^t \) completely characterizing either \( x_n^t \) or \( \sum \nabla \mu_v^t \ln p (A_n^t | \mu_v^t, \theta_o^t) \) individually, or in a joint way in the sense of Corollary 3.6.

3.2 Discussion and Examples

The concept of asymptotic sufficiency is closely related to the notion of correctly specifying the predictor variables in the micro relation. To see this, consider the following simple errors-in-variables model.

Suppose that

\[
x_n^t = \beta u_n^t + s_n^t
\]

and

\[
y_n^t = u_n^t + r_n^t
\]

where \( u_n^t, r_n^t, s_n^t \) are independent normal variables (as functions of the underlying \( A_n^t \)) with

\[
E(u_n^t) = \mu, \quad E(r_n^t) = E(s_n^t) = 0, \quad \text{VAR}(u_n^t) = \sigma_u^2,
\]

\[
\text{VAR}(r_n^t) = \sigma_r^2 \quad \text{and} \quad \text{VAR}(s_n^t) = \sigma_s^2.
\]

Our aim is to study \( x_N^t \) as a function of \( y_N^t \), or equivalently \( \phi^* (\mu, \sigma_u^2, \sigma_r^2, \sigma_s^2) = \beta \mu \) as a function of \( \mu = E(y_n^t) \). We have that

\[
x_N^t = E(x_N^t | y_N^t, \mu, \sigma_n^2, \sigma_r^2, \sigma_s^2) = \beta \lambda \mu + (\beta - \beta \lambda) y_N^t
\]

where \( \lambda = \sigma_r^2 / (\sigma_r^2 + \sigma_u^2) \). Also, in accordance with Theorem 3.5, we have

\[
\delta_N^t = \beta \lambda (y_N^t - \mu) - \beta \bar{r}_N^t + \bar{s}_N^t
\]

and

\[
\xi_N^t = \frac{N^t}{\sigma_u^2} (\lambda (y_N^t - \mu) - \bar{r}_N^t)
\]

Now, we can easily calculate

\[
E(\delta_N^t \xi_N^t | y_N^t, \mu, \sigma_u^2, \sigma_r^2, \phi^2) = \beta \lambda
\]
giving both \( \frac{\partial \hat{x}_N}{\partial \mu}^t \) and the value of the asymptotic (downward) bias in the least squares estimator of \( \beta \). Obviously, sufficiency holds if \( \lambda = 0 \) (no error in \( y_n^t \)) of \( \beta = 0 \) (No structural relation).

From an empirical point of view, the concept of unconditional asymptotic sufficiency is the most important for micro analysis of macro functions. In a sense, all micro-macro statistical modeling is conditional on certain unobserved distributional movements, for if average statistics were observed that depicted these movements, they would be incorporated directly. Moreover, all of the results conditional on \( \theta_o^t \) depend crucially on the way \( \theta_o^t \) is chosen in the \( \theta^t + \mu_y^t \) reparameterization. A simple example makes this point concretely. For a given period \( t_o \), take \( A_k \) to be a scalar random variable distributed normally with mean \( \mu \) and variance \( \sigma^2 \).

Suppose that \( x_k \) is functionally determined by \( A_k \) as

\[
x_k = \gamma_o + \gamma_1 A_k + \gamma_2 A_k^2
\]

This is an exact aggregation model if \( x_k \) is explained by \( A_k \) and \( A_k^2 \).

However, suppose that we have only \( \bar{V}_N = \sum A_n / N = \bar{A}_N \) as data along with \( \bar{x}_N \) for the full population. We therefore regress \( x_k \) on \( A_k \) as in \( x_k = a + b A_k \) and denote the least squares estimate of \( b \) as \( \hat{b}_K \). Given the true model, simple specification analysis techniques give that

\[
\lim_{K \to \infty} \hat{b}_K = \gamma_1 + 2\gamma_2 \mu
\]

We can form \( \phi^* \) and \( \bar{x}_N \) for this problem as

\[
\phi^* (\mu, \sigma^2, \gamma) = \gamma_0 + \gamma_1 \mu + \gamma_2 (\mu^2 + \sigma^2)
\]

\[
\bar{x} = E (\bar{x}_N | \bar{V}_N, \mu, \sigma^2) = \gamma_0 + \gamma_1 \bar{A}_N + \gamma_2 \bar{A}_N^2 + \gamma_2 \frac{N-1}{N} \sigma^2
\]

Thus \( \bar{A}_N \) is (asymptotically) sufficient for determining \( \bar{x}_N \) conditional on \( \sigma^2 \), and we find that
\[
\lim_{k \to \infty} \hat{b}_k = \gamma_1 + 2\gamma_2 \mu = \left. \frac{\partial \phi^*}{\partial \mu} \right|_{\mu, \sigma^2}
\]
in accordance with our results. Alternatively, suppose that we parameterize the normal distribution of \( A_k \) in terms of the mean \( \mu \) and the coefficient of variation \( \omega = \frac{\sigma}{\mu} \). In this case
\[
\phi^* (\mu, \omega, \gamma) = \gamma_0 + \gamma_1 \mu + \gamma_2 (\mu^2 + \omega^2 \mu^2)
\]
and
\[
\bar{x}_N = \gamma_0 + \gamma_1 \bar{A}_N + \gamma_2 \bar{A}_N^2 + \gamma_2 \frac{N-1}{N} \omega^2 \mu^2
\]
Now \( \bar{A}_N \) is not asymptotically sufficient for determining \( \bar{x}_N \) and we find that
\[
\lim_{k \to \infty} b_k = \gamma_1 + 2\gamma_2 \mu \neq \gamma_1 + 2\gamma_2 \mu (1 + \omega^2) = \left. \frac{\partial \phi^*}{\partial \mu} \right|_{\mu, \omega}
\]
Thus all the conditional results weigh heavily on exactly which parameters are held constant. Because of this property, we consider mainly results of the unconditional type in the rest of the exposition.

The concept of asymptotic sufficiency as presented above provides the correct foundation for using single period micro slope estimates to study the macro function. If asymptotic sufficiency does not hold, then all the slope coefficients provide is information on sampling corrections to be applied to \( \bar{x}_N \) when deviations of \( \bar{V}_N \) are observed (\( \bar{V}_N \phi^{**} \) in our notation), instead of the full structural effect (\( \bar{V}_N \phi^* \)). In this sense, absence of this assumption severely limits the usefulness of micro data analysis.

For example, suppose we are studying the relation of consumption to income, and a positive coefficient is found in a micro regression of consumption on income. To infer that mean consumption will increase with increases in mean income implicitly places bounds on \( \bar{V}_N \phi^{**} \), the unobserved component of \( \bar{V}_N \phi^* \). To quantify the macro effect with the micro coefficient requires full asymptotic sufficiency of average income in determining average consumption.

At this point, it is useful to discuss two sets of modeling assumptions which appear as polar extremes under which sufficiency of \( \bar{V}_N \) in
determining $\tilde{x}_N^t$ holds. The first is termed functional sufficiency, where

$$\delta_N^t \equiv 0 \; \text{or} \; \tilde{x}_N^t \equiv \frac{\int f (A_n^t, \gamma^t)}{N^t} \equiv \tilde{x}_N^t (V_N^t, \gamma^t)$$

for all underlying distribution forms $p^*$. This is just a restatement of the exact aggregation form, and so under the general conditions of Lau's Theorem (2.12) and (2.13), $f (A_n^t, \gamma^t)$ must be linear in $v (A_n^t)$, as well as $\tilde{x}_N^t$ in $V_N^t$.

The second set of assumptions is termed distributional sufficiency. Suppose that $f (A_n^t, \gamma^t)$ is unrestricted, but that $V_N^t$ is a sufficient statistic for $\theta^t$ in the distribution of $A_1^t, ..., A_N^t$. Then $V_N^t$ is obviously sufficient for determining $\tilde{x}_N^t$, as $\tilde{x}_N^t$ is always a function of $V_N^t, \gamma^t$ and $N^t$ only. In addition, $\phi^*$ can be a nonlinear function of $\mu_v^t$. Finally, if $\xi_N^t \equiv 0$, and Lau's Theorem is applicable to

$$\sum_{n=1}^{N^t} \frac{\ln p^* (A_n^t, \mu_v^t, \theta_o^t)}{\mu_v^t} = E \left( \sum_{n=1}^{N^t} \ln p^* (A_n^t | \mu_v^t, \theta_o^t) | V_N^t \right)$$

we obtain that $p^*$ must be a member of the exponential family of distributions. The topic of the next section is the characterization of $\phi^*$ when $p^*$ is a member of this family.

4. Distributional Sufficiency and Macro Functions

In this section we present a methodology for estimating second order derivatives of $\phi^*$ with respect to $\mu_v$ from cross section moments, under the assumption that $p$ is a member of the exponential family of distributions. This methodology in principle extends to derivatives of $\phi^*$ of all orders; however due to its complexity we consider only second order derivatives here.

The question of how to estimate higher order derivatives of $\phi^*$ in the general case (that is, with asymptotic sufficiency and without distributional
sufficiency) is still open. However, the formula presented below will allow tests of exact and consistent aggregation models, and thus extend somewhat beyond the distributional sufficiency case.

\( p(A|\theta^t) \) is a member of the exponential family of distributions if it can be written in the form

\[
(4.1) \quad p(A|\theta^t) = C(\theta^t) h(A) \exp \left( \sum_{m=1}^{M} \pi_m (\theta^t) \nu_m(A) \right)
\]

where

\[
C(\theta^t) = \left( \int h(A) \exp \left( \sum_{m=1}^{M} \pi_m (\theta^t) \nu_m(A) \right) dA \right)^{-1}
\]

The joint density of the population \( A_1^t, \ldots, A_N^t \) can then be written as

\[
(4.2) \quad \prod_{n=1}^{N} p(A_n^t|\theta^t) = C(\theta^t)^N \prod_{n=1}^{N} h(A_n^t) \exp \left( \sum_{m=1}^{M} \pi_m (\theta^t) N^t \nu_m^t \right)
\]

By the factorization criterion, \( V_N^t \) (and \( N^t \)) are sufficient for \( \theta^t \).

Beyond the motivation for (4.1) given in section 3.2 based on Lau's Theorem, the appeal of the form (4.1) is that under relatively general conditions, if a sufficient statistic \( \tau \) of dimension \( M < N^t \) exists, then the density of the distribution must have the form (4.1). Very briefly, if the range of variation of \( A_n^t \) does not depend on \( \theta^t \), a continuously differentiable sufficient statistic \( \tau \) for \( \theta^t \) exists and \( p(A|\theta^t) \) is continuously differentiable in \( A \) and \( \theta^t \) (as well as some other mild regularity conditions), then \( p(A|\theta^t) \) must locally have the form (4.1). If \( p(A|\theta^t) \) is further assumed to be analytic, then (4.1) is the global form of the density. This theorem is quite complex, and to adequately discuss it here would take us too far afield. Therefore, the interested reader is referred to references in the statistical literature.

We begin our study by adopting
ASSUMPTION A.6: \( p \) can be written as a member of the exponential family in its natural parameterization:

\[
(4.3) \quad p(A|\pi^t) = C(\pi^t)h(A)\exp\left(\sum_{m=1}^{M} \pi_m^t v_m(A)\right)
\]

where

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

where \( \theta^t \) has been reparameterized by \( \pi^t = (\pi_1^t, \ldots, \pi_M^t) \).

(4.3) holds from (4.1) without loss of generality if the mapping \( \theta^t \rightarrow (\pi_1^t(\theta^t), \ldots, \pi_M^t(\theta^t)) \) - \( \pi^t \) is of rank \( M \). Therefore, Assumption A.6 just eliminates constraints across \( \pi_m^t(\theta^t), m = 1, \ldots, M \) which, from an empirical point of view, are unnecessary at the outset.

Two useful textbook facts about the form (4.3) are

**LEMMA 4.1:** Under Assumption A.6, the natural parameter space

\[
\Gamma = \{\pi^t | p(A|\pi^t) \text{ is a density}\}
\]

is convex.

**LEMMA 4.2:** If \( \psi(A_{1}^t, \ldots, A_{N}^t) \) is a function for which the integral

\[
\int \cdots \int \psi(A_{1}^t, \ldots, A_{N}^t) \prod_{n=1}^{N} h(A_{n}^t)\exp\left[\sum_{m=1}^{M} \pi_{m}^t v_{m}^{t} A_{m}^t\right] dA_{1}^t, \ldots, dA_{N}^t
\]

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

Proofs of these lemmatae 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

\[ \phi(\pi^t, \gamma^t) = \mathbb{E}(x^t|\pi^t) \]

and that

\[ C(\pi^t) = \left( \int h(A) \exp\left( \sum_{m=1}^{M} \pi_m \gamma_m(A) \right) dA \right)^{-1} \]

\( C(\pi^t) \) appears in (4.3) as just a normalizing factor to make \( p(A|\pi^t) \) a density. Both \( \phi \) and \( C \) have some remarkable properties, however, as shown in the following lemma:

**LEMMA 4.3:** Under Assumptions A.1 and A.6, all derivatives of \( \phi \) and \( \ln C \) with respect to \( \pi \) are expressible as moments of the \( x^t, \gamma(A^t) \) distribution. In particular, we have for \( C \) that

\[
- \frac{\partial \ln C}{\partial \pi_m} = \mathbb{E}(\gamma_m(A)|\pi^t) = \mu_{m^t}, \quad m = 1, \ldots, M
\]

\[
- \frac{\partial^2 \ln C}{\partial \pi_m \partial \pi_{m'}} = \mathbb{E}((\gamma_m(A) - \mu_{m^t})(\gamma_{m'}(A) - \mu_{m'^t})|\pi^t) = \sigma_{mm'^t}, \quad m, m' = 1, \ldots, M
\]

and

\[
- \frac{\partial^2 \ln C}{\partial \pi_m \partial \pi_{m'} \partial \pi_x} = \mathbb{E}((\gamma_m(A) - \mu_{m^t})(\gamma_{m'}(A) - \mu_{m'^t})(\gamma_x(A) - \mu_{x^t})|\pi^t)
\]

\[ = \sigma_{mm'^tx}, \quad m, m', x = 1, \ldots, M \]

For \( \phi \) we have

\[
\frac{\partial \phi}{\partial \pi_m} = \mathbb{E}((x - \phi(\pi^t, \gamma^t))(\gamma_m(A) - \mu_{m^t})|\pi^t) = \sigma_{xm}
\]

and

\[
\frac{\partial^2 \phi}{\partial \pi_m \partial \pi_{m'}} = \mathbb{E}((x - \phi(\pi^t, \gamma^t))(\gamma_m(A) - \mu_{m^t})(\gamma_{m'}(A) - \mu_{m'^t})|\pi^t)
\]

\[ = \sigma_{xmm'^t}, \quad m, m' = 1, \ldots, M \]

Proof: The first statement follows from Proposition 4.2. The formulae are obtainable by direct computation.
We are primarily interested in the behavior of $\phi(\pi^t, y^t)$ with respect to changes in $\mu^t_v$. We proceed as before to reparameterize via the mapping

$$
\mu^t_1 = E(v_1(A)|\pi^t) = g_1(\pi^t)
$$

$$
\vdots
$$

$$
\mu^t_M = E(v_M(A)|\pi^t) = g_M(\pi^t)
$$

In view of Lemma 4.4, this mapping is expressible as

$$
\mu^t_v = -\pi^t \ln C(\pi^t) = g(\pi^t)
$$

We can reparameterize the distribution (4.3) in terms of $\mu^t_v$ if the mapping $g$ is invertible $M$; i.e. if the differential matrix $dg^t$ is nonsingular. This matrix, again from Lemma 4.4, can be written as

$$
dg^t = \left(\frac{\partial g_m}{\partial \pi_m} \right)_{\pi^t} = \left(\frac{\partial^2 \ln C}{\partial \pi_m \partial \pi_m} \right)_{\pi^t} = \Sigma^t_{vv}
$$

the covariance matrix of $v(A^t)$: Thus, Assumption A.2 is now (under A.6) implied by Assumption A.1. We therefore form

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

$$
\mathbf{p}^*(A|\mu^t_v) = \mathbf{p}(A|g^{-1}(\mu^t_v))
$$

and

$$
\phi^*(\mu^t_v, y^t) = \phi(g^{-1}(\mu^t_v), y^t)
$$

We are now in a position to show the main result of section 3.1 by direct computation.

**Theorem 4.4:** Under Assumptions A.1 and A.6 the gradient of $\phi^*$ with respect to $\mu^t_v$ is

$$
\nabla_{\mu^t_v} \phi^*(\mu^t_v, y^t) = (\Sigma^t_{vv})^{-1} \Sigma^t_{xy}
$$

and so is consistently estimated by the micro slope regression coefficients from a single period cross section data source.
Proof: By the chain rule,

\[ \nabla_{\mu_\nu} \phi = (c \phi^{-1}) \nabla_{\Pi} \phi \]

Now \[ d \phi^{-1} = (d \phi)^{-1} = (\Sigma_{\nu \nu}^t)^{-1} \] and by Lemma 4.4, \[ \nabla_{\Pi} \phi = L_{x v}^t \]

QED.

We can similarly calculate all higher order derivatives of \( \phi^* \) with respect to \( \mu_\nu \) as functions of moments of the \( x^t, v(A^t) \) distribution. Because these calculations increase greatly in complexity as the order of the derivatives increase, we present only the second derivative calculation, which will allow statistical investigation of the case of a macro function linear in \( \mu_\nu \). We first require some new notation to facilitate the formulae:

\( \Omega_{\pi \pi}^t \) denotes the \( M \times M \) matrix with \( m, m' \) element \( \sigma_{x x m m}^t \) (see Lemma 4.4)

\( E = 1, \ldots , M \)

\( \Omega_{\pi \pi}^t \) denotes the \( M \times M^2 \) matrix \nolimits

\[ \Omega_{\pi \pi}^t = [\Omega_{1 \pi \pi}^t, \ldots , \Omega_{M \pi \pi}^t] \]

\( \Sigma_{x v v}^t \) denotes the \( M \times M \) matrix with \( m, m' \) element \( \sigma_{x x m m}^t \)

\( I_M \) is the identity matrix of order \( M \) and finally, \( P_M \) denotes the \( M^2 \times M^2 \) matrix

\[ P_M = \begin{bmatrix} P_{11} & \cdots & P_{1 M} \\ \vdots & \ddots & \vdots \\ P_{M 1} & \cdots & P_{M M} \end{bmatrix} \]

\( P_{m m}^t \) is the \( M \times M \) matrix with \( m', m \) element 1 and remaining entries 0.

We can now show
THEOREM 4.5: The matrix of second order derivatives of $\phi^*$ with respect to $\mu_v$ evaluated at period $t$, written as $V^2_{\mu_v} \phi^*$, can be computed as

\[
V^2_{\mu_v} \phi^* = -(\Sigma_{VV}^t)^{-1} \Omega_{MM} \left( I_M \otimes \Sigma_{WW}^t \right) P_M \left( I_M \otimes \Sigma_{WW}^t \right) \left( \Sigma_{WW}^t \right)^{-1} + (\Sigma_{WW}^t)^{-1} \Sigma_{WW}^t \left( \Sigma_{WW}^t \right)^{-1}
\]

\[
\left( V^2_{\mu_v} \phi^* \right) \text{ is the MxM matrix with } m,m' \text{ element } \frac{\partial^2 \phi^*}{\partial \mu_m \partial \mu_{m'}} \bigg|_{t,Y_t}
\]

The proof is by direct computation, but is relatively complex so we present a sketch of it in the Appendix.

The formula (4.9) for second derivatives of $\phi^*$ is sufficiently complex to warrant illustration by a simple example. Suppose $M = 1$, or that $A_t$ is distributed according to

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

with $\pi^t$ a scalar parameter, and

\[
E(x|\pi^t) = \phi(\pi^t,^t) = \phi^*(\mu^t,^t)
\]

where $\mu^t$ is the (scalar) mean of $v_1(A)$. Now, in accordance with Theorem 4.4 we find that

\[
\frac{\partial \phi^*}{\partial \mu_1} = \frac{\partial \phi}{\partial \pi} \frac{\partial \pi}{\partial \mu_1} = \frac{\sigma_{x1}^t}{\sigma_{11}^t} (\sigma_{11}^t)^{-1} = \frac{\sigma_{x1}^t}{\sigma_{11}^t} \text{ plim } \frac{\hat{b}}{K}^K
\]

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

\[
\frac{\partial^2 \phi^*}{\partial \mu_1^2} = \frac{\partial \phi}{\partial \pi} \left( \frac{\partial \pi}{\partial \mu_1} \right)^2 + \frac{\partial \phi}{\partial \pi} \frac{\sigma_{x1}^t}{\sigma_{11}^t}
\]

By Lemma 4.4, we have

\[
\frac{\partial^2 \phi}{\partial \pi^2} = \sigma_{x11}^t; \quad \frac{\partial \phi}{\partial \mu_1} = \frac{1}{\sigma_{11}^t}; \quad \frac{\partial \phi}{\partial \pi} = \sigma_{x1}^t
\]
and so we must find \( \frac{\partial^2 \pi}{\partial \mu_1^2} \). Since
\[
\frac{\partial \mu_1}{\partial \pi} \cdot \frac{\partial \pi}{\partial \mu_1} = 1
\]
by differentiation with respect to \( \mu_1 \) we get
\[
0 = \frac{\partial^2 \mu_1}{\partial \mu_1^2} \frac{\partial \pi}{\partial \mu_1} + \frac{\partial^2 \pi}{\partial \mu_1^2} \frac{\partial \mu_1}{\partial \pi}
\]
or
\[
\frac{\partial^2 \pi}{\partial \mu_1^2} = -\frac{\partial^2 \mu_1}{\partial \mu_1^2} \left( \frac{\partial \pi}{\partial \mu_1} \right)^2 = -\frac{\sigma_{111}^t}{(\sigma_{11}^t)^3}
\]

Inserting these values into (4.10) gives

\[
(4.11) \quad \frac{\partial^2 \phi^*}{\partial \mu_1^2} = \frac{\sigma_{x_{111}^t}}{(\sigma_{11}^t)^2} - \frac{\sigma_{x_{11}^t \sigma_{111}^t}}{(\sigma_{11}^t)^3}
\]

which agrees with (4.9) for \( M = 1 \).

As we have shown, we can express the second order derivatives \( \phi^* \) in terms of moments of the underlying exponential family population density. Estimating these moments by their sample counterparts in a cross section data base will allow consistent estimation of the second derivatives in that time period. Asymptotic inferences using these estimates are possible by standard methods.\(^{25}\) Thus, in particular, we can test whether \( \phi^* \) is a linear function of \( \mu_1 \).\(^{26}\)

In addition, one can easily show that if \( x_{n}^t = f(A_{n}^t, Y_t) \) has either the exact aggregation form (linear in \( v(A_{n}^t) \) with constant coefficients) or the consistent aggregation form (linear with coefficients that vary independently of \( v(A_{n}^t) \)) that the expression in (4.9) is identically zero without distributional sufficiency. Thus, (4.9) can be used to test whether either of these model
forms holds, although it no longer has the second derivative interpretation. In this sense the usefulness of (4.9) extends beyond the distributional sufficiency case. 27

5. Conclusion

The major result of this paper is that under asymptotic sufficiency, slope regression coefficients will consistently estimate the first derivatives of the true macro function. Asymptotic sufficiency is a general concept which appears to underlie all quantitative inferences made from cross section data about macro functions. A more general study can only be made with additional information, such as micro data from several time periods, as in panel or reinterview studies.

In addition to the regression result, we have shown that if the average explanatory variables are sufficient for the underlying population distribution parameters, then in principle (in the case of a population density of the exponential family) one can empirically characterize macro function derivatives of all orders using cross section data. Under this additional structure, one can actually test whether the macro function is linear, quadratic or some higher order nonlinear function. These techniques extend to tests of linear aggregation schemes such as exact and consistent aggregation models.

The main appeal of these techniques is that they allow an empirical characterization of macro functions using micro data. In addition, even if the true macro function is linear, the independent effects of the average explanatory variables may be difficult to ascertain because of trending behavior or other data problems (referred to as multicollinearity). As justified here, micro data can be used to estimate these effects, where they are generally more visible. In this spirit, if one want to approximate a macro function
to the first order using average and micro data, then an exact aggregation scheme
should be assumed, as the estimates obtained from each data source will coincide
in large samples, thus allowing one to take advantage of the increased data
input in terms of increased precision of the final estimate values. Moreover,
the macro function coefficients can be allowed to vary if a structural change
is indicated from an additional cross section data source.

Future research in this area can be directed to methods for distinguishing
whether a particular data configuration is consistent with either functional
or distributional sufficiency. Also, there is a need for techniques to
characterize nonlinearities in $\phi$ when neither of the above types of sufficiency
hold. Finally, and most important, are tests of asymptotic sufficiency,
from which the relevance of any of these techniques can be considered.

The results of this research constitute a first step into the area
of empirically characterizing macro function with micro data. Hopefully
they will help end the practice of neglecting distributional issues in the
study of macro data, a practice which is now so prevalent.
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 shown in Rao (1973) Section 6.2 a;

Lemma AP.1

Let $\tau_N$ 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 $\Sigma_t$. Further, let $g(\tau_{1N}, \ldots, \tau_{MN}, N)$ be a function which is totally differentiable in $\tau_{1N}, \ldots, \tau_{MN}$, and that $\nabla_{\tau_{MN}} g + g \neq 0$ as both $N \to \infty$ and $\tau_{MN}^+(\gamma_1, \ldots, \gamma_M)' = \gamma$. 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

$$G' \Sigma_t G$$

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

$$\lim_{N \to \infty} \sqrt{N}(\bar{x}(\mu, \mu, \theta, \gamma, N) - \phi^*(\mu, \theta, \gamma)) = 0$$

Proof: Fix $N$ and consider

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

$$E(\sqrt{N}(\bar{x}_N - \bar{x}(\bar{V}_N, \mu, \theta, \gamma, N))) + \sqrt{N}(\bar{x}(\mu, \mu, \theta, \gamma, N) - \phi^*(\mu, \theta, \gamma))$$

$$= 0 + \sqrt{N}(\bar{x}(\mu, \mu, \theta, \gamma, N) - \phi^*(\mu, \theta, \gamma))$$

Now as $N \to \infty$, the first expectation approaches zero by virtue of Lemma 3.1 a) and Lemma AP.1 applied to $\bar{x}$. Thus
\[
\lim_{N \to \infty} \sqrt{N} (\bar{x}(\mu, \nu, \theta, \gamma, N) - \phi^*(\mu, \theta, \gamma)) = 0 \quad \text{QED}
\]

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

Proof of Theorem 3.5

Conditions (3.11) ii) allow differentiation of
\[
\bar{x}_N^t = \int x_N^t p(A_1^t, \ldots, A_N^t | V_0, \mu, \theta^o) \quad \forall A_1^t, \ldots, A_N^t
\]
under the integral sign, which gives
\[
\nabla_{\mu} \bar{x} = E(\nabla_{\mu} x_N^t | V_N^t = V_0, \mu, \theta^o) \nabla_{\mu} x_N^t
\]
where
\[
\psi^t_N = \sum_{n=1}^{N^t} \nabla_{\mu} \ln p(A_n^t | \mu, \theta^o) - \nabla_{\mu} \ln p(V_N^t | \mu, \theta^o)
\]

Theorem 3.5 is shown if

\[(AP.1) \quad E(\psi^t_N | V_N^t = V_0, \mu, \theta^o) = 0
\]

(or \( E(\nabla_{\mu} \ln p(A_n^t | \mu, \theta^o) | V_N^t = V_0, \mu, \theta^o) = \nabla_{\mu} \ln p(V_0, \mu, \theta^o) \)).

By conditions (3.11) i) we can differentiate
\[
1 = E(1 | V_N^t = V_0, \mu, \theta^o) \quad \text{under the integral sign, which gives}
\]

\[(AP.1) \quad \text{above} \quad \text{QED}
\]

Proof Sketch for Theorem 4.5:

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

As in Theorem 4.4
\[
\nabla_{\mu} \phi^* = \begin{pmatrix}
\frac{\partial \phi}{\partial \mu_1} \\
\vdots \\
\frac{\partial \phi}{\partial \mu_M}
\end{pmatrix}
= \begin{pmatrix}
\frac{\partial g_1}{\partial \mu_1} & \cdots & \frac{\partial g_M}{\partial \mu_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_1}{\partial \mu_M} & \cdots & \frac{\partial g_M}{\partial \mu_M}
\end{pmatrix}
\begin{pmatrix}
\frac{\partial \phi}{\partial \pi_1} \\
\vdots \\
\frac{\partial \phi}{\partial \pi_M}
\end{pmatrix}
\]
Therefore

\[
\frac{\partial^2 \phi^*}{\partial \mu_i \partial \mu_j} = \sum_{m=1}^{M} \frac{\partial \phi}{\partial \mu_i} \frac{\partial^2 \phi}{\partial \mu_j} + \sum_{m=1}^{M} \sum_{m'=1}^{M} \frac{\partial^2 \phi}{\partial \mu_i \partial \mu_{j'}} \left( \frac{\partial g_{m}}{\partial \mu_{j'}} \right) \left( \frac{\partial g_{m}}{\partial \mu_{j}} \right) \]

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

\[
\begin{bmatrix}
\frac{\partial g_1}{\partial \mu_1} & \cdots & \frac{\partial g_M}{\partial \mu_1} \\
\frac{\partial g_1}{\partial \mu_2} & \cdots & \frac{\partial g_M}{\partial \mu_2} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_1}{\partial \mu_M} & \cdots & \frac{\partial g_M}{\partial \mu_M}
\end{bmatrix}
\begin{bmatrix}
\frac{\partial^2 \phi}{\partial \mu_1} \\
\frac{\partial^2 \phi}{\partial \mu_2} \\
\vdots \\
\frac{\partial^2 \phi}{\partial \mu_M}
\end{bmatrix}
\begin{bmatrix}
\frac{\partial g_1}{\partial \mu_1} & \cdots & \frac{\partial g_M}{\partial \mu_1} \\
\frac{\partial g_1}{\partial \mu_2} & \cdots & \frac{\partial g_M}{\partial \mu_2} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_1}{\partial \mu_M} & \cdots & \frac{\partial g_M}{\partial \mu_M}
\end{bmatrix}
\]

which by Lemma 4.3 equals

\[(\Sigma_{vv})^{-1} \Sigma_{xvv} \left( \Sigma_{vv}^{-1} \right)^{-1}\]

giving the second term in the statement of the theorem. Now, if \( \mathbf{B} = [b_{ij}(y)] \)
is an \( M \times M \) matrix of functions of \( y \) then we denote by \( \frac{\partial}{\partial y} \mathbf{B} \) the matrix

\[
\frac{\partial}{\partial y} \mathbf{B} = \begin{bmatrix}
\frac{\partial b_{ij}(y)}{\partial y} \\
\vdots \\
\frac{\partial b_{ij}(y)}{\partial y}
\end{bmatrix}
\]
The first term of (AP.2) is expressible in matrix format as

\[(\text{AP.4}) \quad \left[ d_{\mu_1} (d^{-1}) \nabla_{\f_1}, \cdots, d_{\mu_M} (d^{-1}) \nabla_{\f_M} \right] \]

Now, in order to evaluate \(d_{\mu_m} (d^{-1})\), \(m = 1, \cdots, M\), we use the relation

\((d^{-1}) (d) = I_M\) so that (if \(O_M\) is an \(M\times M\) matrix of zeros)

\[d_{\mu_m} (d^{-1}) dg + d^{-1} \left( d_{\mu_m} (d) \right) = O_M \quad m = 1, \ldots, M\]

or

\[d_{\mu_m} (d^{-1}) = -d^{-1} \left( d_{\mu_m} (d) \right) dg^{-1}\]

Now, if \(g_{m\pi\pi}\) denotes the \(M\times M\) matrix with \(i, j\) element \(\frac{\partial^2 g_m}{\partial \pi_i \partial \pi_j}\), we express \(d_{\mu_m} (dg)\) as

\[d_{\mu_m} (dg) = \left[ g_{1\pi\pi} d_{\mu_m} g^{-1}, \cdots, g_{M\pi\pi} d_{\mu_m} g^{-1} \right] \]

so that

\[(\text{AP.5}) \quad d_{\mu_m} (d^{-1}) = -d^{-1} \left[ g_{1\pi\pi} d_{\mu_m} g^{-1}, \cdots, g_{M\pi\pi} d_{\mu_m} g^{-1} \right] dg^{-1}\]

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

\[dg^{-1} = (\Sigma^t)^{-1}_{\f \f}; \quad \nabla_{\f} = \Sigma^t_{\f \f} \]

\[g_{m\pi\pi} = \Omega^t_{m\pi\pi}; \quad m = 1, \cdots, M \]

\[\left[ d_{\mu_1} g^{-1}, \cdots, d_{\mu_M} g^{-1} \right] = dg^{-1} = \left[ \Sigma^t_{\f \f} \right]^{-1} \]

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

QED
NOTES

* I wish to thank Jerry Green for helpful discussion during the early stages of this work. Also helpful were D. Schmalensee, F. Fisher and D. McFadden. All errors are, of course, the responsibility of the author.

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. Ideally, one requires micro data from each period of the time series data in order to study the influence of all distributional movements. Unfortunately, this much data is generally not available. Some exceptions are the longitudinal, or panel data sets now available on wage rates, hours worked and demographic characteristics across families and over time.

4. The only exception known to this author is Friedman's permanent income - permanent consumption model. See the first example of Section 3.2 (errors in variables) for illustration of this fact.

5. $p$ 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 \( \{A_n^t | n=1, \ldots, N^t \} \) via \( p(A|\theta^t) \).

6. An alternative formulation of this set-up is to take \( x_n^t, A_n^t \) as having a joint distribution in the population. Then we would study

\[
\bar{x}_n^t = E(x_n^t | A_n^t) = f(A_n^t, \gamma^t)
\]

where \( \gamma^t \) contains parameters of this joint distribution. All analysis in the text is then valid for \( \bar{x}_n^t \) in place of \( x_n^t \), although a bit must be said about the relation of \( \bar{x}_n^t \) to \( \bar{x}_N^t \).

7. Of course, if common prices affect the underlying population distribution (e.g. the distribution of earnings) they may enter as elements of \( \theta^t \). Also if prices vary across the population, they are expressible as components of \( v(A) \).

8. If this is not true, we can always invert for a subset of \( \mu_1^t, \ldots, \mu_M^t \). The full invertibility assumption does not affect the character of our results, but just simplifies the notation. Similarly, we could allow for \( M > L \).


10. Each index \( k \) of the 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 random sample.

11. Typical numbers for a study of U.S. family demand behavior are \( N^t = 70 \) million for 1972, with a budget study of size \( K = 10,000 \).

12. See also MacDonald and Lawrence (1978) for a recent contribution.

13. In addition, see Green (1964).

14. This approach was introduced by Theil (1954); see also Theil (1975), chapter 5 and Green (1964).

15. \( \nabla_{\bar{V}} \bar{x} \) represents the gradient of \( \bar{x} \) with respect to \( \bar{V}_N \); i.e. the \( M \)-vector with

\[
\frac{\partial \bar{x}}{\partial \bar{V}}_{|N} | \bar{V}_N, \mu, \theta, \gamma, N
\]
16. Rao (1973), section 2c is an excellent reference for these theorems; also, see section 6a for some useful corollaries.

17. 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, \( \sqrt{K}(\hat{b}_K - G(\mu_v, \theta_0, \gamma)) \) approaches a normal vector as \( K \to \infty \) with mean zero and variance covariance matrix

\[
\Sigma_b = (\Sigma_v^{-1})^T (\Sigma_v^{-1}) (\Sigma_v^{-1})^T
\]

where \( \Sigma_v \) is the matrix with mm' element

\[
E[(x^T - \phi^*)(v_m(A^T - \mu_m) - \sigma_{xm})]
\]

\[
((x^T - \phi^*)(v_m(A^T - \mu_m) - \sigma_{xm})]
\]

\( \Sigma_b \) will correspond to the usual expression (i.e. \( \sigma^2 x^T v_v^{-1} \)), \( \sigma^2 \) is residual variance if there is a zero correlation between \( u_i^2 \) and \( (v_i(A^T - \mu_m) - \sigma_{xm}) \), for \( i,j,1,\ldots,M \), where \( \hat{u}_k = x_k^T - \bar{x}_k^T - (v(A_k^T - \bar{v}_k^T) \hat{b}_k \). Use of the standard estimators may provide an adequate approximation to \( \Sigma_b \) if the sample counterparts to these correlations are small.


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. For instance, see the second example of section 3.2 where \( \sigma^2 \) is set to a fixed constant. In this case \( \Lambda_N \) is sufficient for \( \mu \) and \( \phi^* \) is a quadratic function of \( \mu \).


22. The exponential family form (4.1) is quite general. Examples of univariate distributions expressible in this form are the normal \((\mu, \sigma^2)\), Poisson \((\gamma)\), negative binomial \((r, \theta)\), the gamma distributions and the beta distributions.
Examples of multivariate distributions expressible in this form include the normal with mean $\mu$ and variance covariance matrix $\Sigma$. Distributions which are not of the form (4.1) include the uniform and Cauchy distributions. See Ferguson (1967) for more details.

23. This theorem is mentioned in Ferguson (1967), p. 129 and Lehmann (1959), p. 51 without proof. The original references to its proof are Koopman (1936), Darmois (1935) and Pitman (1936), prompting some authors to refer to the exponential form as the Koopman-Pitman-Darmois form. An excellent paper that proves the theorem in more generality than is needed here (with the original result as corollary) is Barankin and Maitra (1963).


25. This is because the formula (4.9) is a continuous and differentiable function of the moments comprising it. "Standard methods" refer to applications of theorems such as Lemma A.1.

26. Of course $\nabla^2 \phi^* = 0$ is only a necessary condition for linearity, but still its failure is reason to reject linearity.

27. Although formula (4.9) is directly estimable from cross section moments, it would be useful if (4.9) 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

$$x_t^* = C_0 + C_1 u_1 (A_1^t) + C_2 (u_1 (A_1^t))^2$$

gives

$$\lim_{K^{\infty}} \hat{C}_2 = \frac{\sigma_{x1}^2 \sigma_{11}^2 - \sigma_{11}^2 \sigma_{11}^2}{\sigma_{11}^2 \sigma_{11}^2 - (\sigma_{11}^2)^2 - (\sigma_{11}^2)^3}$$

which is clearly proportional to (4.11), and thus provides an easily computable test of (4.11) equalling zero (although bear in mind stochastic
structure differences, as in fn. 17). The natural conjecture is that including all squared and cross-product terms in a micro regression produces a matrix proportional expression to (4.9). Unfortunately, proving or disproving this result is a computational nightmare, and to date the author has not completely resolved this problem.
REFERENCES


