AGGREGATION, EFFICIENCY AND CROSS SECTION REGRESSION

by

Thomas M. Stoker

Working Paper #1453-83

June 1983
Revised October 1984

ABSTRACT

In this paper several results are established which provide for the consistent estimation of macroeconomic effects using cross section data, for general assumptions on the movement of the population distribution over time. We show that macroeconomic effects are always consistently estimated by linear instrumental variables coefficients, where the instruments are determined by the form of distribution movement. This leads to a natural way to assess the biases in OLS coefficients as estimators of macroeconomic effects, provides a nonparametric macroeconomic interpretation of linear instrumental variables coefficients when the true microeconomic behavioral model is unknown, and gives a nonparametric interpretation of standard regression decomposition statistics such as R^2 relative to the information costs of nonlinearities in aggregation. All of the results are valid without imposing any testable restrictions on the cross section data.

*Thomas M. Stoker is Associate Professor, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. This research was funded by National Science Foundation Grants No. SES-8308768 and SES-8410030. The author wishes to thank J. Powell for ongoing discussions, and A. Deaton, J. Hausmann, J. Heckman, D. Jorgenson, A. Lewbel, J. Muellbauer, J. Rotemberg, A. Zellner and the referees for helpful comments on this and related work. Special thanks go to Terence Gorman, both for comments and gracious hospitality shown the author during a week at Nuffield College, Oxford, when the ideas of the paper were originally conceived. All errors, etc., remain the responsibility of the author.

AGGREGATION, EFFICIENCY AND CROSS SECTION REGRESSION

1. INTRODUCTION

In Stoker(1982), necessary and sufficient conditions are established for cross section OLS regression coefficients to consistently estimate macroeconomic effects - the effects of varying the means of the regressors on the dependent variable mean.[1] A particularly interesting sufficient condition for this property exists when the distribution of regressor values in the population varies over time via the exponential family form. The striking feature of this condition is that it implies no restrictions which are testable using the cross section data, and therefore establishes a macroeconomic interpretation of linear OLS coefficients when the true behavioral model underlying the data is unknown.

The purpose of this paper is to establish several characterizing results which provide for the consistent estimation of macroeconomic effects using cross section data, for general assumptions on the movement of the population distribution over time. We show that macroeconomic effects are always consistently estimated by linear instrumental variables coefficients, where the instruments are determined by the form of distribution movement. This leads to a natural way to assess the biases in OLS coefficients as estimators of macroeconomic effects, provides a nonparametric macroeconomic interpretation of linear instrumental variables coefficients when the true behavioral model is unknown and gives a nonparametric interpretation of standard regression variance decomposition statistics (such as R^2) relating to the information costs of nonlinearities in aggregation. All of our results are valid without imposing any testable restrictions on the cross section data.

In Section 2 we introduce the basic assumptions required, and discuss both the microeconomic and macroeconomic aspects of the framework. Section 3 provides a brief derivation of the result that macroeconomic effects are always consistently estimated by linear cross section regression coefficients, where the score vectors of distribution movement are used as instruments. Section 4 allows a deeper understanding of the result by noting its intimate connection to the statistical theory of the efficiency of data aggregates, which characterizes valid micro instruments for estimating macroeconomic effects, and shows how the standard OLS variance decomposition terms exactly reflect the Cramer-Rao bounds for dependent variable averages as estimators of their means. Section 5 treats several related topics - exponential family special cases and the characterization of R^2, as well as a result on the use of consistent distribution parameter estimates in the formation of the proper instruments. Section 6 gives a brief numerical example and Section 7 contains some concluding remarks.

It may be helpful to consider at the outset a simple example which we will return to in Section 6. Suppose we are studying the demand y of a commodity over a period of constant prices, with our model stating that demand is determined by total budget expenditure X via a stable Engel curve $E(y|X)=F(X)$ for all families. Suppose that the total expenditure distribution $p(X|\theta^t)$ is lognormal in each time period t, with ln X normal with mean θ^t and constant variance. We can determine the means of y and X in terms of θ^t as $E^t(y)=\phi^*(\theta^t)$ and $\mu^t=E^t(X)=H(\theta^t)$, and the mean of y in terms of μ^t as $E^t(y)=\phi^*(H^{-1}(\theta^t))=\phi(\mu^t)$, which constitutes the aggregate demand function.

Now, suppose that we have cross section data at time $t=t^\circ$ (where $\theta=\theta^\circ$) - a random sample $\{y_k, X_k\}_k$. Since the above lognormal distribution is an exponential family with driving variable ln X, the result of Stoker(1982) implies that if we perform the regression

$$(1.1) \qquad y_k = \hat{a} + \hat{b} \ln X_k + \hat{\varepsilon}_k$$

then $\hat{b}$ consistently estimates $\partial\phi^*/\partial\theta^\circ$ - the effect of changing the log geometric mean of total expenditure θ on mean demand $E(y)$. For most practical purposes, however, we would be more interested in $\partial\phi/\partial\mu^{o}$, the effect of mean total expenditure on mean demand. Here we show that if we estimate the equation

$$(1.2) \qquad y_k = \hat{c} + \hat{d}\, X_k + \hat{u}_k$$

using $\ln X_k$ as the instrumental variable, then $\hat{d}$ consistently estimates $\partial\phi/\partial\mu^{o}$. We characterize how $\ln X_k$ can be regarded as a valid instrument, and we show how the variance decomposition of regressions such as the above relate to the efficiency properties of sample averages - for example the large sample value of R^2 from (1.1) is the "first order efficiency" of the sample average $\bar{y}$. All of these results are valid regardless of the true Engel curve form $E(y|X)=F(X)$.

2. THE BASIC FRAMEWORK

2.1 Notation and Formal Assumptions

This section presents the notation and basic assumptions used. The salient features of the framework are as follows. We model changes over time in the joint distribution of a dependent variable y and a vector of predictor variables X, and formulate the aggregate relationship between the means of y and X over time. When the microeconomic model between y and X is stable over time, the aggregate relationship explains the mean of v as a function of changes in the marginal distribution of X. The results of the paper connect regression statistics from cross section data (observed for a single time period) to the derivatives of the aggregate relationship, or macroeconomic effects.

Formally, we assume that the values of (y,X) across the population at time t represent a random sample from a distribution $\mathcal{P}(y,X|\theta^{t})$, which is absolutely continuous with respect to a σ-finite measure ν, with Radon-Nikodym density $P(y,X|\theta^t) = d\mathcal{P}(y,X|\theta^t)/d\nu$. θ is a vector of parameters which indexes movements over time in the joint distribution of y and X.

At time $t=t^o$, a cross section data set is observed, which consists of K observations y_k, X_k; $k=1,\ldots,K$. These observations are assumed to be a random sample from the distribution with density $P^o(y,X)=P(y,X|\theta^o)$, where $\theta=\theta^o$ for the cross section time period $t=t^o$. We make the following assumptions concerning the local structure of the distribution around $\theta=\theta^o$.

ASSUMPTION 1: θ is an M-vector of parameters, and P exists for $\theta\in\Theta$, where Θ is a compact subset of $\mathbb{R}^M$ which contains an open neighborhood of θ^o.

ASSUMPTION 2: $P(y,X|\theta)$ is twice differentiable in θ for all $\theta\in\Theta$.

We define the score vector of P with respect to θ in the usual way as

$$(2.1)\qquad \ell_\theta = \frac{\partial \ln P(y,X|\theta)}{\partial\theta}$$

ASSUMPTION 3: The means of y, X and the score vector ℓ_θ, and the variances and covariances of y, X and ℓ_θ exist and are continuous in θ for all $\theta\in\Theta$. The information matrix $I_\theta=E(\ell_\theta\ell_\theta')$ is positive definite for all $\theta\in\Theta$.

The means of y and X as functions of θ are denoted as follows

$$(2.2)\qquad E(y) = \mu_y = \int yP(y,X|\theta)d\nu = \phi^*(\theta)$$

$$(2.3)\qquad E(X) = \mu = \int XP(y,X|\theta)d\nu = H(\theta)$$

We denote the relevant covariance matrices as follows: $Var(y)=\sigma_{yy}$, $Var(X)=\Sigma_{xx}$, $Cov(X,y)=E((X-\mu)(y-\mu_y))=\Sigma_{xy}$, where the dependence of these parameters on θ is suppressed in the notation, and $Cov(\ell_\theta,y)=E(\ell_\theta y)=\Sigma_{\theta y}$, $Cov(\ell_\theta,X)=E(\ell_\theta X')=\Sigma_{\theta X}$, and $Var(\ell_\theta)=I_\theta$. For the parameter value $\theta=\theta^o$, we denote the score vector ℓ_{θ^o} as ℓ_o and the respective covariance matrices as Σ_{oy}, Σ_{ox} and I_o. The sample moments computed from the cross section data are given as follows: the averages are denoted by $\bar{y}=\Sigma y_k/K$ and $\bar{X}=\Sigma X_k/K$ and the sample covariance matrices are denoted by S with the same subscripts as the population covariances, as in $S_{xx}=\Sigma(X_k-\bar{X})(X_k-\bar{X})'/K$ and $S_{xy}=\Sigma(X_k-\bar{X})(y_k-\bar{y})/K$.

ASSUMPTION 4: ϕ^* and H are differentiable with respect to the components of θ for all $\theta\in\Theta$.

ASSUMPTION 5: H is invertible for all $\theta\in\Theta$.

The overall induced relation between $E(y)=\mu_y$ and $E(X)=\mu$ - the <u>aggregate function</u> - is defined by inverting (2.3) and inserting into (2.2) as[2]

$$\mu^y=\phi^*(H^{-1}(\mu))=\phi(\mu) \tag{2.4}$$

These definitions and assumptions are discussed below. Two additional regularity assumptions are presented in the Appendix.

2.2 Discussion

2.2a Microeconomic Model

The microeconomic model between y and X is represented implicitly in the above framework by the conditional distribution of y given X, which captures all restrictions on the behavioral relationship between y and X. While Assumptions 1-5 suffice for the majority of the development, the most attractive practical advantages of the results exist when the microeconomic model is stable over time. Formally, this occurs when the density P factors as $P(y,X|\theta)=q(y|X)p(X|\theta)$ for all $\theta\in\Theta$, with the conditional density $q(y|X)$ of y given X independent of θ. In this case, the score vector ℓ_θ depends only on the marginal distribution $p(X|\theta)$ of X, with (2.1) simplifying to

$$\ell_\theta = \frac{\partial \ln p(X|\theta)}{\partial\theta} \tag{2.1a}$$

To interpret this framework, consider that a standard (micro) econometric analysis of the relationship between y and X begins with the specification of a behavioral function $y=f_\gamma(X,u)$, where u represents unobserved individual heterogeneity and γ a set of parameters to be estimated, together with the stochastic distribution of u given X, say with density $\tilde{q}(u|X)$. Combining the behavioral function and the heterogeneity distribution gives the conditional density $q_\gamma(y|X)$, which could be used to formulate a likelihood function for estimation. Here, we suppress γ in the notation, just assuming the existence of such a model. Clearly the framework is quite general, easily accomodating standard normal (linear or nonlinear) regression models, logit or probit discrete choice models, etc.

Assumptions 1-5 also allow for the possibility that the conditional distribution of y given X explicitly depends on θ, as in $q_\gamma(y|X,\theta)$. Thus the framework applies to externality models such as those of bandwagon effects in demand, or situations where Lucas critique arguments would suggest that the distribution of X was important for the determination of y by individual agents.

2.2b Distribution Movements and Macroeconomic Effects

The macroeconomic structure of the framework lies in the relation between the means of y and X over time. At time t, there is a value θ^t of the distribution parameters, which determine the values $E^t(y)=\mu_y{}^t=\phi^*(\theta^t)$ and $E^t(X)=\mu=H(\theta^t)$. The induced macroeconomic relation between μ_y and μ over time is given by the aggregate function $\mu_y{}^t=\phi(\mu^t)$ of (2.4). We denote parameter values at the cross section time period $t=t^o$ with 0 superscripts, as $\theta=\theta^o$, $\mu_y{}^o=\phi^*(\theta^o)$ and $\mu^o=H(\theta^o)$.

The distribution parameters θ can represent virtually any shape or location parameters desired - means, variances, skewness, etc. The important specification requirement is Assumption 5, which states that the means μ of the chosen micro predictor variables X completely parameterize distribution movements.[3] If we reparameterize the distribution as $P'(y,X|\mu)=P(y,X|H^{-1}(\mu))$, we can find the score vector and information matrix with respect to μ as

$$(2.5) \qquad \ell_{\mu} = \left[\frac{\partial H}{\partial \theta}\right]^{-1} \ell_{\theta} \quad ; \quad I_{\mu} = \left[\frac{\partial H}{\partial \theta}\right]^{-1} I_{\theta} \left(\left[\frac{\partial H}{\partial \theta}\right]^{-1}\right)'$$

where $\partial H/\partial \theta$ is the differential (Jacobian) matrix of $\mu=H(\theta)$.

The macroeconomic effects of interest are the effects of changes in μ on μ_y, or $\partial\phi/\partial\mu$, which is given explicitly by the chain rule applied to (2.4) as

$$(2.6) \qquad \frac{\partial \phi}{\partial \mu} = \left[\frac{\partial H}{\partial \theta}\right]^{-1} \frac{\partial \phi^*}{\partial \theta}$$

The question of interest to this paper is how these effects (evaluated at $\mu=\mu^\circ$) can be estimated using cross section data at time $t=t^\circ$.

3. MACROECONOMIC EFFECTS AND INSTRUMENTAL VARIABLES REGRESSION

In this section we present a quick derivation of the connection between macroeconomic effects and cross section instrumental variables coefficients. We begin by showing

LEMMA 1: Under Assumptions 1, 2, 3, 4 and 6, we have that $\partial\phi^*/\partial\theta = \Sigma_{\theta y}$ and $\partial H/\partial\theta = \Sigma_{\theta X}$.

Proof: Following Cramer(1946, p.67), from (A.1a) of Assumption 6 we can differentiate (2.2) under the integral sign as

$$\frac{\partial \phi^*}{\partial \theta} = \int y \frac{\partial P}{\partial \theta} d\nu = \int y \frac{\partial \ln P}{\partial \theta} P(y,X|\theta)d\nu = \int y \ell_{\theta} P(y,X|\theta)d\nu = \Sigma_{\theta y}$$

Similarly for $\partial H/\partial\theta$ of (2.3). QED

In order to characterize the macroeconomic effects using the cross section data, we define $\ell_{0k}=\partial \ln P(y_k,X_k|\theta^\circ)/\partial\theta$ as the score vector evaluated at $\theta=\theta^\circ$, y_k and X_k, for $k=1,\dots,K$. Now consider the coefficients $\hat{d}$ of the linear cross section equation

$$(3.1) \qquad y_k = \hat{c} + \hat{d}'X_k + \hat{u}_k \qquad\qquad k=1,\dots,K$$

We denote the estimates obtained by instrumenting with ℓ_{ok} by $\hat{d}^*$, defined via[4]

(3.2) $\hat{d}^*=(S_{ox})^{-1}S_{ox}$

where S_{ox}, S_{oy} are the sample covariance matrices between ℓ_{ok} and X_k, y_k respectively. The major result characterizing macroeconomic effects and cross section regression is

THEOREM 2: Under Assumptions 1,2,3,4,5 and 6, we have $d \equiv \lim \hat{d}^* = \partial\phi/\partial\mu^o$ a.s.

Proof: By the Strong Law of Large Numbers (Rao(1973), p.115, SLLN2), Assumption 3 implies that $\lim S_{ox}=\Sigma_{ox}$ a.s. and $\lim S_{oy}=\Sigma_{oy}$ a.s. We therefore have $\lim \hat{d}^*=(\Sigma_{ox})^{-1}\Sigma_{oy}$ a.s., where Σ_{ox} is invertible by Assumptions 4, 5 and Lemma 1. The result follows from Lemma 1 applied to (2.6) evaluated at $\theta=\theta^o$.

QED

Thus, under our general framework, the macroeconomic effects of a change in the mean $\mu=E(X)$ on the mean $\mu_y=E(y)$ at time t^o are consistently estimated by the coefficients of the linear instrumental variables regression of y_k on X_k, using the score vector ℓ_{ok} as the instrumental variable. The interesting feature of this result is that the score vector ℓ_{ok} represents only the differential distribution movement at time t^o, with the actual shape of the base density $P^o(y,X)$ entering only through the random sample feature of the cross section data.

The power of this latter observation is best seen in the case of a stable microeconomic model $q(y|X)$. From (2.1a), the score vectors ℓ_{ok} are determined only by the direction of movement of the marginal distribution of X. Theorem 2 says that the linear coefficients $\hat{d}^*$ consistently estimate the macroeconomic effects regardless of the true form of the microeconomic model $q(y|X)$. Thus Theorem 2 provides a valid large sample macroeconomic interpretation of linear instrumental variables regression coefficients when the true microeconomic model is general and unknown.[5]

A practical problem with Theorem 2 is that the score vectors ℓ_{ok} will in general depend on the true values of the distribution parameters. This problem is treated in Section 5. We now turn to Section 4, where Theorem 2 is integrated with the theory of efficient estimation.

4. AGGREGATION, EFFICIENCY AND CROSS SECTION REGRESSION

To obtain a deeper understanding of the theoretical structure of Theorem 2, we alter the focus of the exposition to consider the seemingly unrelated problem of estimating the population parameter values θ°, μ° and μ_y° using the cross section data observed at time $t=t^\circ$. In Section 4.1, the (Cramer-Rao) theory of first order efficiency of consistent, uniformly asymptotically normal (CUAN) estimators is reviewed, and related to the sample averages $\bar{y}=\Sigma y_k/K$ and $\bar{X}=\Sigma X_k/K$ of the cross section data. In Section 4.2, the efficiency concepts are integrated with Theorem 2 on the estimation of macroeconomic effects with cross section data.

4.1 Efficient Estimation of the Population Parameters

To introduce efficiency concepts, we take y_k, X_k; $k=1,\ldots,K$ to represent a random sample from a distribution with density $P(y,X|\theta)$, and consider the generic problem of estimating population parameter values. The sample averages $\bar{y}=\Sigma y_k/K$ and $\bar{X}=\Sigma X_k/K$ are natural estimators of the means $\mu_y=\phi^*(\theta)$ and $\mu=H(\theta)$ respectively. The asymptotic properties of $\bar{y}$ and $\bar{X}$ are given in Theorem 3, which follows directly from the Strong Law of Large Numbers and the Central Limit Theorem:

THEOREM 3: Under Assumptions 1, 2 and 3, $\bar{y}$ and $\bar{X}$ are strongly consistent estimators of $\mu_y=\phi^*(\theta)$ and $\mu=H(\theta)$ respectively, and the limiting distribution of

$$\sqrt{K}\begin{bmatrix} \bar{y}-\phi^*(\theta) \\ \bar{X}-H(\theta) \end{bmatrix}$$

is multivariate normal with mean 0 and covariance matrix

$$\begin{bmatrix} \sigma_{yy} & \Sigma_{xy}' \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}$$

where convergence is uniform for all $\theta\in\Theta$.

The quality of CUAN estimators (such as $\bar{y}$ and $\bar{X}$) can be judged on the basis of asymptotic efficiency. Formally, we introduce the definition of "first order efficiency" of Rao(1973, p. 348-9) as follows. Denote the probability density of the y_k, X_k sample as $\bar{P}(y_1,X_1,\ldots,y_K,X_K|\theta)=\Pi P(y_k,X_k|\theta)$ and define the normalized score vector as

$$(4.1) \qquad L_\theta = \frac{1}{K}\,\frac{\partial \ln \bar{P}}{\partial \theta} = \frac{\Sigma\, \ell_{\theta k}}{K}$$

where $\ell_{\theta k}$ is the score vector (2.1) evaluated at θ and y_k, X_k. We can now state

DEFINITION: A CUAN estimator T_K of $g(\theta)$ is <u>first order efficient</u> if there exists a matrix $B(\theta)$ such that

$$(4.2) \qquad \text{plim}\ \sqrt{K}(T_K-g(\theta)-B(\theta)L_\theta)=0$$

uniformly in $\theta\in\Theta$.

Notice that condition (4.2) implies that the limiting distribution of $\sqrt{K}(T_K-g(\theta))$ is normal with mean 0 and covariance matrix $B(\theta)I_\theta B(\theta)'$, which is the matrix representation of the asymptotic Cramer-Rao lower bound. Formally, it is well known that if T_K^* is a CUAN estimator of $g(\theta)$ with asymptotic variance $V(\theta)$, then $V(\theta)-B(\theta)I_\theta B(\theta)'$ is a positive semi-definite matrix for all θ. In particular, if $g(\theta)=\theta$, then $B(\theta)=I_\theta^{-1}$, so that T_K has asymptotic variance I_θ^{-1}, or if $g(\theta)=\mu$, then T_k has asymptotic variance I_μ^{-1}. Finally, we note that if $g(\theta)$ is a scalar parameter, then (4.2) implies that the asymptotic squared (canonical) correlation between T_k and L_θ is unity, and that we can define the "first order efficiency" of any CUAN estimator as its asymptotic squared (canonical) correlation with L_θ.[6]

Recalling that our interest is in estimation using the cross section data at time $t=t^\circ$, we utilize these concepts first to define the locally best estimator of θ near $\theta=\theta^\circ$, and second to indicate the inefficiency inherent in the sample averages $\bar{X}$ and $\bar{y}$ as estimators of their means. Following Efron(1975) and Pitman(1979), we define the best locally unbiased estimator of θ near θ° as

$$(4.3) \qquad U_{\theta^\circ} = (I_o)^{-1} \frac{\Sigma \ell_{ok}}{K} + \theta^\circ$$

Clearly U_{θ° is locally unbiased and attains the asymptotic Cramer-Rao lower bound. For our purposes, we note that U_{θ° is a sample average: if we define $Z_k=(I_o)^{-1}\ell_{ok}+\theta^\circ$, then $U_{\theta^\circ}=\Sigma Z_k/K=\bar{Z}$.

Finally, in view of Theorem 3, we denote the asymptotic inefficiency of $\bar{X}$ as an estimator of $\mu=\mu^\circ$ as the difference between its asymptotic variance and the asymptotic Cramer-Rao bound as

$$(4.4) \qquad IN(\bar{X}) = \Sigma_{xx} - (I_{\mu^\circ})^{-1} = \Sigma_{xx} - \left[\frac{\partial H}{\partial \theta^\circ}\right]' (I_o)^{-1}\left[\frac{\partial H}{\partial \theta^\circ}\right]$$

where the latter equality follows from (2.5). The asymptotic inefficiency of $\bar{y}$ is similarly denoted as

$$(4.5) \qquad IN(\bar{y}) = \sigma_{yy} - \left[\frac{\partial \phi^*}{\partial \theta^\circ}\right]'(I_o)^{-1}\left[\frac{\partial \phi^*}{\partial \theta^\circ}\right]$$

4.2 Efficiency and the Cross Section Estimation of Macroeconomic Effects

The above development on efficiency can be usefully integrated with Theorem 2 on the cross section estimation of macroeconomic effects by considering the use of the locally efficient estimator components Z_k as instruments for the cross section regression (3.1):

$$(3.1) \qquad y_k = \hat{c} + \hat{d}'X_k + \hat{u}_k \qquad\qquad k=1,\ldots,K$$

We begin by noting the covariance structure between Z_k, y_k and X_k as

$$(4.6a) \qquad \Sigma_{zy} = E^o((Z-\theta^o)y) = (I_o)^{-1}\Sigma_{oy}$$

$$(4.6b) \qquad \Sigma_{zx} = E^o((Z-\theta^o)X') = (I_o)^{-1}\Sigma_{ox}$$

$$(4.6c) \qquad \Sigma_{zz} = E^o((Z-\theta^o)(Z-\theta^o)') = (I_o)^{-1}$$

If we denote the instrumental variables slope coefficient estimator of (3.1) as $\hat{d}=(S_{zx})^{-1}S_{zy}$, where S_{zx}, S_{zy} are the sample covariance matrices between Z_k and y_k, X_k respectively, then we can state

COROLLARY 4: Under Assumptions 1,2,3,4,5 and 6, we have $d = \lim \hat{d} = \partial\phi/\partial\mu^o$ a.s.

Proof: Obvious from Theorem 2, noting that $S_{zx}=(I_o)^{-1}S_{ox}$ and $S_{zy}=(I_o)^{-1}S_{oy}$.

QED

Consequently, the Z_k vectors provide valid instruments for estimating macroeconomic effects.

The theoretical attractiveness of the Z_k instruments over the score vectors ℓ_{ok} derives from the following two corollaries, which pertain to the reduced form regressions underlying (3.1):

$$(4.7) \qquad y_k = \hat{a} + \hat{b}'Z_k + \hat{\varepsilon}_k$$

$$(4.8) \qquad X_k = \hat{A} + \hat{B}'Z_k + \hat{v}_k$$

If we denote the respective OLS slope coefficient estimators as $\hat{b}=(S_{zz})^{-1}S_{zy}$ and $\hat{B}=(S_{zz})^{-1}S_{zx}$, where S_{zz} is the sample covariance matrix of the Z_k vectors, then we have

COROLLARY 5: Under Assumptions 1,2,3,4 and 6, we have

(4.9a) $\quad b \equiv \lim \hat{b} = \dfrac{\partial \phi^*}{\partial \theta^\circ}$ a.s.

(4.9b) $\quad B \equiv \lim \hat{B} = \dfrac{\partial H}{\partial \theta^\circ}$ a.s.

Proof: Obvious from Lemma 1 and the Strong Law of Large Numbers applied to S_{zz}, S_{zy} and S_{zx}. QED

Consequently the OLS slope coefficients from the reduced form equations consistently estimate the macroeconomic effects on $E(y)=\mu_y$ and $E(X)=\mu$ of changes in the original parameters θ evaluated at $\theta=\theta^\circ$.[7] The "indirect least squares" method of calculating $\hat{d}$ of (3.1) is reflected in $\hat{B}^{-1}\hat{b}=\hat{d}$, which is consistent with the chain rule formula (2.6).

The final corollary concerns the large sample analysis of variance decomposition of the reduced form equations. If we define $s_{\varepsilon\varepsilon}=\Sigma\hat{\varepsilon}_k^2/K$ and $S_{vv}=\Sigma(\hat{v}_k\hat{v}_k')/K$, then we have:

COROLLARY 6: Under Assumptions 1,2,3,4 and 6 we have that

(4.10a) $\quad \sigma_{\varepsilon\varepsilon} \equiv \lim s_{\varepsilon\varepsilon} = IN(\bar{y})$ a.s

(4.10b) $\quad \Sigma_{vv} \equiv \lim S_{vv} = IN(\bar{X})$ a.s.

Proof: Clearly we have that $\sigma_{\varepsilon\varepsilon}=\lim(s_{yy}-\hat{b}'S_{zz}\hat{b})=\sigma_{yy}-b'\Sigma_{zz}b$ a.s., so that the result follows from Corollary 5, (4.6c) and (4.5). Similarly for Σ_{vv}. QED

Consequently, the large sample analysis of variance decompositions of each of the reduced form equations can be characterized as follows: the overall average sum of squares corresponds to the asymptotic variance of the average of the dependent variable, the average fitted sum of squares corresponds to the relevant Cramer-Rao lower bound and the average residual sum of squares corresponds to the relevant first order inefficiency.

Before proceeding, it is useful to note the role of the θ

parameterization above. For both the instrumental variables regression (3.1) and the variance decompositions of Corollary 6, the particular θ parameterization used to index distribution changes is irrelevant. If $\tilde{\theta}=G(\theta)$ represented an equivalent parameterization, then the entire development in terms of $\tilde{\theta}$ yields the same results, provided that Z_k is correctly defined as the component of the locally efficient estimator $U_{\tilde{\theta}^\circ}$. The only difference is that the reduced form coefficients $\hat{b}$ and $\hat{B}$ would then consistently estimate the appropriate derivatives with respect to $\tilde{\theta}$. In particular, we could have begun with $\tilde{\theta}=\mu=E(X)$, defined $Z_k=(I_{\mu^\circ})^{-1}\ell_{\mu^\circ}+\mu^\circ$ and obtained the results directly, with B an NxN identity matrix and $b=d=\partial\phi/\partial\mu^\circ$. This emphasizes that distribution movement is the important feature of the framework, not the specific θ parameterization used to index it, and justifies choosing the most convenient parameterization in the examples that follow.

5. THE COMPUTATION OF ESTIMATES AND RELATED DISCUSSION

In this section we consider the important special case of exponential family distribution movement, where for particular parameterizations the required instruments Z_k can be written as independent of the population parameter value $\theta=\theta^\circ$. We then show a result which indicates how to construct Z_k in general (nonexponential family) circumstances, and close with some related remarks.

5.1 Distribution Movement of the Exponential Family

5.1a The Exponential Family with Driving Variable X

The first natural question concerns when X_k is an appropriate instrument for estimating macroeconomic effects, with $\hat{d}$ of (3.1) (and $\hat{b}$ of (4.7)) the OLS slope coefficient of y_k on X_k. By the development, this case occurs when $\bar{X}$ is an efficient estimator of $\mu=\mu^\circ$, which by the Fisher-Darmois-Koopman-Pitman Theorem[8] occurs if and only if $F(y,X|\theta)$ is (locally) an exponential family with driving variable X; given here as

(5.1) $P(y,X|\theta)=C(\theta)h(y,X)\exp[\pi(\theta)'X]$

where the range of $\pi(\theta)$ contains a neighborhood of $\pi(\theta^o)$ and $C(\theta)=[\int h(y,X)\exp[\pi(\theta)'X]d\nu]^{-1}$ is a normalizing constant.[9] Without loss of generality, π can be defined so that $\pi(\theta^o)=0$, in which case $h(y,X)=P(y,X|\theta^o)$. It is well known that if Σ_{xx} is nonsingular for all $\theta\in\Theta$, then (5.1) with Assumptions 1, 2 and 3 imply Assumptions 4, 5 and 6, and if the conditional distribution of y given X exists for $\theta=\theta^o$, a stable microeconomic model is implied as well.

If we take $\pi=\pi(\theta)$ as the appropriate distribution parameters, then $\ell_{\pi}=X-\mu$ and $I_{\pi}=\Sigma_{xx}$; $\mu=H^*(\pi)$ and $\mu_y=\phi^{**}(\pi)$ have derivatives $\partial H^*/\partial\pi=\Sigma_{xx}$ and $\partial\phi^{**}/\partial\pi=\Sigma_{xy}$ so that $\partial\phi/\partial\mu=(\Sigma_{xx})^{-1}\Sigma_{xy}$, which for $t=t^o$ is the a.s. limit of $\hat{d}=\hat{b}=(S_{xx})^{-1}S_{xy}$. For the $\mu=H^*(\pi)$ parameterization, $\ell_{\mu}=(\Sigma_{xx})^{-1}(X-\mu)$ and $I_{\mu}=(\Sigma_{xx})^{-1}$, so by (4.3) we have $U_{\mu^o}=\bar{X}$ and $Z_k=X_k$.

As originally pointed out in Stoker(1982), the interesting feature of this special case is that it justifies a nonparametric interpretation of cross section OLS slope coefficients when the true microeconomic model is general and unknown, namely as the macroeconomic effects of changes in $E(X)=\mu$ on $E(y)=\mu_y$ evaluated at $t=t^o$. Thus a Taylor series interpretation of OLS coefficients is always valid, where the relevant Taylor series is that of the aggregate function $\mu_y=\phi(\mu)$ around $\mu=\mu^o$, and distribution movements are in form (5.1).[10]

5.1b <u>The Information Interpretation of R^2</u>

Applying Corollary 6 to the case of distribution movements of the form (5.1) provides an interesting interpretation of the goodness of fit statistic R^2 from cross section OLS regression. Since $\bar{X}$ is an efficient estimator, Σ_{VV} is a matrix of zeros. $\sigma_{\varepsilon\varepsilon}$, representing the inefficiency in $\bar{y}$, will be nonzero unless y_k is an exact linear function of X_k a.s.

If we denote the large sample value of the squared multiple correlation coefficient as $R^2=1-(\sigma_{\varepsilon\varepsilon}/\sigma_{yy})$, Corollary 6 implies that the appropriate Cramer-Rao lower bound is the exact fraction R^2 of the asymptotic variance of $\bar{y}$. Clearly, R^2 represents the first order efficiency of $\bar{y}$ as defined in Section 4. Equivalently, since the Cramer-Rao lower bound can be regarded as the reciprocal of the relevant amount of statistical information, the use of $\bar{y}$ to estimate $\mu_y=\mu_y^{\circ}$ incorporates R^2 of the available information, and $1-R^2$ is lost.

Thus, when distribution movements are of the form (5.1), the costs of nonlinearities in aggregation are measured by $1-R^2$; the exact percentage of information loss in $\bar{y}$, regardless of the true form of the microeconomic model. Note that this implies that the source of departures from linearity is irrelevant; R^2 is the exact fraction of the available information utilized regardless of whether the true microeconomic model is nonlinear, or linear with an additive stochastic disturbance.

5.1c The Macroeconomic Interpretation of Cross Section Instrumental Variables Estimators

The final exponential family case of interest is when Z_k represents observable data, but $Z_k \neq X_k$. Z_k may contain some of the components of X_k or nonlinear functions of X_k, as in $Z_k=D(X_k)$, or represent entirely different observed variables from the cross section data set. To accomodate this case, consider the expanded population density $P(y,X,Z|\theta)$, where we denote the implied mean of Z as $E(Z)=\mu_Z$.

Z_k represents a correct instrument if and only if $\bar{Z}$ is an efficient estimator of $\mu_Z=\mu_Z^{\circ}$, which as above occurs if and only if $P(y,X,Z|\theta)$ can be written (locally) in exponential family form with driving variables Z;

$$(5.2) \qquad P(y,X,Z|\theta)=C(\theta)h(y,X,Z)\exp[\pi(\theta)'Z]$$

where the range of $\pi(\theta)$ contains a neighborhood of $\pi(\theta^{\circ})$ and $C(\theta)$ is the appropriate normalizing constant. In this case it is easy to verify that $\partial\phi/\partial\mu=(\Sigma_{zx})^{-1}\Sigma_{zy}$, which for $t=t^{o}$ is the a.s. limit of the linear instrumental variables coefficients $\hat{d}=(S_{zx})^{-1}S_{zy}$.

This expanded case is of interest because form (5.2) accomodates situations where there is a stable microeconomic model of the form $q(y,X|Z)$, which clearly allows for general (simultaneous equation) behavioral models where components of X are endogenous variables, with Z appropriately ancillary to the assumed stochastic structure. In this case $\hat{d}$ consistently estimates the induced effects of changing $\mu=E(X)$ on $\mu_y=E(y)$ evaluated at $t=t^{o}$, which provides a macroeconomic interpretation of $\hat{d}$ when the true microeconomic model is general or unknown.

5.2 Computation in the General Case

Unless distribution movements are locally in exponential family form (5.1) or (5.2), the proper instruments ℓ_{ok} or Z_k will necessarily depend on the true population parameter value $\theta=\theta^{o}$. In this section we consider this general case where ℓ_θ depends not only on the movement parameters θ, but also on additional shape parameters Λ, as in

$$(5.3)\qquad \ell = \frac{\partial \ln P(y,X|\theta,\Lambda)}{\partial\theta} = \frac{\partial \ln P(y,X|\Gamma)}{\partial\theta}$$

In this extension, $\Gamma=(\theta,\Lambda)$ represents a finite set of parameters required for the complete characterization of the score vectors, although our interest is only in estimating the macroeconomic effects of varying θ. When there is a stable microeconomic model, Λ can represent additional parameters describing the shape of the marginal distribution of X (see Section 6 for an example, and Section 5.3 for the case of a varying microeconomic model, each with Λ nonempty). The true value of these parameters at time $t=t^{o}$ is $\Gamma^{o}=(\theta^{o},\Lambda^{o})$, and we assume that there is a strongly consistent estimator $\hat{\Gamma}=(\hat{\theta},\hat{\Lambda})$ of $\Gamma=\Gamma^{o}$ available. Under the additional structure of Assumption 7, we show that all of our previous results are valid when $\hat{\Gamma}$ is used in place of Γ^o in the construction of ℓ_{ok} or Z_k.

We begin by defining $\hat{\ell}_{ok}$ as the score vector evaluated at y_k, X_k and $\hat{\Gamma}$ as

$$(5.4) \qquad \hat{\ell}_{ok} = \frac{\partial \ln P(y_k, X_k | \hat{\Gamma})}{\partial \theta}$$

Denote the sample covariance matrices formed using $\hat{\ell}_{ok}$ as $\hat{S}_{oy}$, $\hat{S}_{ox}$ and $\hat{S}_{oo}$. The major result can now be stated as

Theorem 7: Under Assumptions 1, 2, 3 and 7, if $\hat{\Gamma}$ is such that $\lim \hat{\Gamma}=\Gamma^o$ a.s., then $\lim \hat{S}_{oy}=\Sigma_{oy}$ a.s., $\lim \hat{S}_{ox}=\Sigma_{ox}$ a.s. and $\lim \hat{S}_{oo}=I_o$ a.s.

Proof: Define $\ell_k(\Gamma)=\partial \ln P(y_k,X_k|\Gamma)/\partial\theta$, so that $\ell_k(\Gamma^o)=\ell_{ok}$ and $\ell_k(\hat{\Gamma}^o)=\hat{\ell}_{ok}$. To prove the result for $\hat{S}_{oy}$, define the matrix $S_y(\Gamma)=\Sigma\ell_k(\Gamma)y_k/K$, so that $S_y(\Gamma^o)=S_{oy}$ and $S_y(\hat{\Gamma})=\hat{S}_{oy}$. In view of (A.2a) of Assumption 7, Lemma 2.3 of White(1980b) implies that $S_y(\Gamma)$ converges uniformly in Γ to $E^o(\ell(\Gamma)y)$. Since $\lim \hat{\Gamma}=\Gamma^o$ a.s., by Lemma 4 of Amemiya(1973), $\lim \hat{S}_{oy}=\lim S_y(\hat{\Gamma})=E^o(\ell(\Gamma^o)y)=\Sigma_{oy}$, giving the result. Similar arguments establish the remainder of the Theorem.

QED

Theorem 7 allows for the consistent estimation of macroeconomic effects in nonexponential family circumstances. If $\hat{Z}_k$ is similarly defined using $\hat{\Gamma}$ instead of Γ^o, then corresponding versions of Theorem 2 and Corollaries 4, 5 and 6 using $\hat{\ell}_{ok}$ and $\hat{Z}_k$ can easily be shown to be valid, as long as Assumption 7 is appended to the list of antecedents.[11]

5.3 Further Remarks on Biases in OLS

In this Section, we have shown how macroeconomic effects can be computed under alternative assumptions on distribution movement. The central special case is given by (5.1) - exponential family distribution movement with driving variables X - which is associated with cross section OLS as a consistent method of estimating macroeconomic effects. If $\Sigma_{VV} \neq 0$ in (4.10b), then X fails to be a proper instrument, with the OLS slope coefficients of y on X potentially biased estimators of the macroeconomic effects. Consequently, distribution movements not of the form (5.1) can be regarded as inducing an "errors-in-variables" problem into the cross section OLS regression of y on X, with Σ_{VV} a precise measure of the departure of distribution movements from exponential family with driving variables X.[12] A natural method of assessing the size of the biases is to compute the proper consistent estimates as above, and compare them to the OLS estimates.

Finally, as indicated in Section 2, the most attractive practical advantages of our results exist for the case of a stable microeconomic model. In this case the relevant instruments depend only on the movement of the marginal distribution of X, and provide macroeconomic interpretations of cross section regression estimators regardless of the true form of behavior connecting y and X. When the microeconomic model varies with θ, our results are still valid, however the proper instruments will depend on precisely how θ affects individual behavior. In other words, in this case there is a behavioral component of the score vector - $\partial \ln q_{\gamma}(y|X,\theta)/\partial\theta$ - which may depend on the exact specification of q and/or the values of the behavioral parameters Υ. The additional parameter vector Λ would then contain Υ, which must be estimated prior to the construction of the score vector $\hat{\ell}_{ok}$.

6. A COMPUTED EXAMPLE

In this section we present some illustrative calculations based on the lognormal distribution movement example of the Introduction. Suppose that demand y for a commodity depends on total expenditure X via a true microeconomic model q(y|X), where in each time period the total expenditure distribution is lognormal

$$(6.1)\quad p(X|\theta^t)=(1/2\pi\sigma(\theta^t)X)\exp[-(\ln X-\theta^t)^2/2\sigma(\theta^t)^2)$$

where ln X has mean θ^t and variance $\sigma(\theta^t)^2$. For illustration, we consider two forms of distribution movement:

Case A (Fixed log-variance): $\sigma(\theta^t)=\sigma^o$, constant over time.

Case B (Fixed log-coefficient of variation): $\sigma(\theta^t)=\lambda\theta^t$, where λ is a constant.

Mean total expenditure under each scenario is given as

$$(6.2) \qquad \mu=E(X)=\exp(\theta+(1/2)\sigma(\theta)^2)$$

and mean demand is formally given as $\mu_y=E(y)=\phi^*(\theta)\doteq\phi(\mu)$.

Now, suppose we observe cross section observations on y and X at time $t=t^o$, where $\theta=\theta^o$. Our results say that the linear coefficient of y regressed on X obtained by instrumenting with either the score ℓ_θ or the component Z of the efficient estimator of θ will consistently estimate the macroeconomic effect $\partial\phi/\partial\mu^o$. For Case A, we have $\ell_\theta=(\ln X-\theta)/(\sigma^o)^2$, and $Z=\ln X$, which is consistent with the fact that under Case A, (6.1) is a exponential family with driving variable $Z=\ln X$. For Case B, we have $\ell_\theta=[(\ln X)^2-\theta\ln X-(\lambda\theta)^2]/\lambda^2\theta^3$ and $Z=(I_o)^{-1}\ell_o+\theta^o$, where $I_\theta=[1-2\lambda^2]/\lambda^2\theta^2$. Here we would use the results of Section 5 to construct either ℓ_o or Z using a strongly consistent estimator of $\Gamma^o=(\theta^o,\lambda)$, where λ $(=\Lambda^o$ in (5.3)) is an additional shape parameter.

For illustration, we present the computed values of the macroeconomic effects assuming a precise Engel curve function of the form

$$(6.3) \qquad y = \Upsilon_o X + \Upsilon_1 X\ln X$$

which is Muellbauer's(1975) PIGLOG form (see also Deaton and Muellbauer(1980) and Jorgenson, Lau and Stoker(1982)). The aggregate functions for cases A and B can be seen to be

$$(6.4A) \quad \mu_y=E(y)=\phi(\mu)=\mu[\Upsilon_o+\Upsilon_1(\ln\mu+(1/2)(\sigma^o)^2)]$$

$$(6.4B) \quad \mu_y=E(y)=\phi(\mu)=\mu[\Upsilon_o+\Upsilon_1(\ln\mu+[-1+\sqrt{1+2\lambda^2\ln\mu}]^2)]$$

All of our results regarding limits of regression estimators can be formally verified here, which we leave to the reader. In Table 1 we present the values of the various effects for typical demand coefficient values $\Upsilon_0=.2$ and $\Upsilon_1=-.01$, and population parameter values $\theta^o=9.119$, $\sigma^o=.578$ (and $\lambda=\sigma^o/\theta^o=.0634$), which represent the values of these parameters for the 1974 distribution of family total expenditures in the United States given in Stoker(1979). Also, for comparison, we present the large sample value of the OLS slope coefficient of y on X, namely $\beta=(\Sigma_{XX})^{-1}\Sigma_{Xy}$ as well as the asymptotic bias β-d.

From Table 1 we see that there is virtually no difference in the true macroeconomic effects between the two scenarios, while in each the true macroeconomic effect d differs substantially from the leading coefficient of demand $\Upsilon_0=.2$. The asymptotic bias in least squares is quite small in both cases. X is more highly correlated with the density score ℓ_0 under Case B than Case A, although the efficiencies of $\bar{y}$ and $\bar{X}$ are reasonably high under each scenario. Finally, we should note that while the demand formulation (6.3) does not include an additive random disturbance, including one would only lower the R_ε^2 values of Table 1, and narrow the difference between them.

7. CONCLUDING REMARKS

In this paper we have shown how macroeconomic effects can always be consistently estimated using cross section linear instrumental variables regression coefficients, by developing a general framework connecting cross section regression and the efficiency properties of data aggregates. The striking feature of our results is that they involve no testable restrictions on the cross section data, and thus provide interpretations of standard regression statistics when the true microeconomic model is misspecified or unknown. The sensitive feature of the framework is the precise specification of distribution movement, which determines the form of the proper instruments

TABLE 1: Comparison of Cases A and B

$\Theta^0 = 9.119$, $\sigma^0 = .578$, $\gamma_0 = .2$, $\gamma_1 = -.01$

	Case A	Case B
$B = \frac{\partial E(X)}{\partial \theta_1}$	10,786.40	11,181.19
$b = \frac{\partial E(y)}{\partial \theta_1}$	1,029.78	1,063.44
$d = \frac{\partial E(y)}{\partial E(X)}$	.0955	.0951
β (OLS)	.0937	.0937
$\beta - d$	-.0018	-.0014
Efficiency of		
$\bar{X} - R_v^2$	.8423	.9127
$\bar{y} - R_\varepsilon^2$	.8729	.9354

for cross section regression. Consequently, the broad context of the results on estimating macroeconomic effects may be viewed as an exchange of specific functional form assumptions for specific assumptions on distribution movement.

These features naturally suggest a number of related theoretical research areas. For instance, a general interpretation of the standard omitted variable bias formula is given in Stoker(1983a), and the macroeconomic effects of behavioral parameter changes is analyzed in Stoker(1983b). Topics for future research include the extension of the micro linearity tests of Stoker(1982) to general distribution movements (and how such tests would compare to traditional specification tests), as well as how one would test (say using panel data) whether a particular parameterization of distribution movements was appropriate.

Appendix: Further Regularity Assumptions

For the purpose of differentiating under integral signs, define the difference quotients as

$$D_{yj}(y,X,h) = \frac{y[P(y,X|\theta^o+he_j)-P(y,X|\theta^o)]}{h}$$

$$D_{ij}(y,X,h) = \frac{X_i[P(y,X|\theta^o+he_j)-P(y,X|\theta^o)]}{h}$$

where X_i is the i^{th} component of X, e_j is the unit vector with j^{th} component 1 and h is a scalar. We now assume

ASSUMPTION 6: There exist ν-integrable functions $g_{yj}(y,X)$, $g_{ij}(y,X)$; $i,j=1,\ldots,M$ and $h_o>0$, such that for all h where $0 \leq |h| \leq h_o$,

(A.1a) $\quad |D_{yj}(y,X,h)| \leq g_{yj}(y,X)$

(A.2b) $\quad |D_{ij}(y,X,h)| \leq g_{ij}(y,X)$

for all $i,j=1,\ldots,M$.

Clearly Assumption 6 will be guaranteed by Assumption 3 if the carrier set over which ν is defined is compact. Note that Assumption 6 could be replaced by a less restrictive but much more technical assumption, along the lines of Roussas (1972).

For using consistent estimates of the population parameters in Section 5, we assume that $\Gamma=(\theta,\Lambda)$ is an M+L vector of parameters, where $\Gamma\epsilon\Theta^*$, where Θ^* is a compact subset of $\mathbf{R}^{M+L}$ containing an open neighborhood of Γ^o. We denote the j^{th} component of ℓ_θ from (5.3) as $\ell_{\theta j}$. We then require

ASSUMPTION 7: There exists measurable functions $h_{oj}(y,X)$, $h_{ij}(y,X)$ and $h_{ij}^*(y,X)$; $i,j=1,\ldots,M$ such that

(A.2a) $\quad |y\ell_{\theta j}| \leq h_{oj}(y,X) \qquad j=1,\ldots,M$

(A.2b) $\quad |X_i\ell_{\theta j}| \leq h_{ij}(y,X) \qquad i,j=1,\ldots,M$

(A.2c) $\quad |\ell_{\theta i}\ell_{\theta j}| \leq h_{ij}^*(y,X) \qquad i,j=1,\ldots,M$

for all $\Gamma\in\Theta^*$, where $E^o|h_{oj}|^{1+\delta}$, $E^o|h_{ij}|^{1+\delta}$ and $E^o|h_{ij}^*|^{1+\delta}$ are bounded for some $\delta>0$, $i=1,\ldots,M$, $j=1,\ldots,M$.

NOTES

1. The terms "predictor variable" and "regressor" are used interchangeably to describe X_k in the equation $y_k=\hat{a}+\hat{b}'X_k+\hat{\varepsilon}_k$, with "dependent variable" refering to y_k.

2. See Stoker(1984) for several examples of computed aggregate functions.

3. X could include squares, cross products, etc., of a smaller set of variables.

4. By convention, we always assume that a constant is appended to the set of instrumental variables for calculation of estimates.

5. It should be noted that if the function $\partial\phi(\mu)/\partial\mu$ were known exactly, then better estimates of its value at $\mu=\mu^o$ are available than the slope coefficients. In particular, if $\hat{\mu}$ is a consistent, first order efficient estimator of $\mu=\mu^o$, then $\partial\phi(\hat{\mu})/\partial\mu$ is consistent and first order efficient for $\partial\phi/\partial\mu^o$.

6. See Rao(1973) for a thorough development of these ideas. It should be noted that under some further regularity conditions, maximum likelihood estimators are first order efficient - see Rao(1973) or Roussas(1972), the latter containing a very general development.

7. The aggregation - sufficient statistic connections noted in Stoker(1982) are also at work here, since U_{θ^o} is a locally sufficient statistic for θ near $\theta=\theta^o$ - see Barankin and Maitra(1963).

8. See Barankin and Maitra(1963) for a generalized version of this theorem, and Stoker(1982) for the original references.

9. Textbook treatment of the exponential family can be found in Lehmann(1959) and Ferguson(1967). For modern treatments see Barndorf-Neilson(1978), Efron(1975,1978) and Efron and Hinkley(1978).

10. White(1980a) shows that OLS coefficients consistently estimate the derivatives of the true microeconomic model evaluated at the mean of X only under very restrictive conditions.

11. While the ability to find strongly consistent estimators depends on the exact form of the density $P(y_k,X_k|\theta,\Lambda)$, it should be noted that if (μ,Λ) is used to parameterize the density, then $\bar{X}$ is a strongly consistent estimator of $\mu=\mu^o$ by Theorem 3.

12. One should be careful with this interpretation, because the standard errors-in-variables bias formulae are not generally applicable here. Also, it should be noted that $\Sigma_{yv}=0$ is a stronger condition than zero statistical curvature (Efron(1975)), since Σ_{yv} depends on the particular variables X.

REFERENCES

Amemiya, T.(1973),"Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41, pp.997-1016.

Barndorff-Neilson, O.(1978), Information and Exponential Families in Statistical Theory, Wiley, New York.

Barankin, E.W. and A.P. Maitra(1963),"Generalization of the Fisher-Darmois-Koopman-Pitman Theorem on Sufficient Statistics," Sankhya, A25,pp.217-227.

Cramer, H.(1946) Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J.

Deaton, A. and J. Muellbauer(1980), "An Almost Ideal Demand System," American Economic Review, 70, pp.312-326.

Efron, B.(1975), "Defining the Curvature of a Statistical Problem (with Application to Second Order Efficiency)," Annals of Statistics, 3, pp.1189-1242.

Efron, B.(1978), "The Geometry of Exponential Families," Annals of Statistics, 6, pp.362-376.

Efron, B. and E. V. Hinkley(1978), "Assessing the Accuracy of the Maximum Likelihood Estimator," Biometrika, 65, pp.457-482.

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

Jorgenson, D.W., L.J. Lau and T.M. Stoker(1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in R. Basmann and G. Rhodes, eds., Advances in Econometrics, Vol. 1, JAI Press.

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

Muellbauer, J.(1975), "Aggregation, Income Distribution and Consumer Demand," Review of Economic Studies, 42, pp.525-543.

Pitman , E.J.G.(1979), Some Basic Theory for Statistical Inference, Chapman and Hall, London.

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

Roussas, G.G.(1972), Contiguity of Probability Measures, Cambridge University Press, Cambridge, England.

Stoker, T.M.(1979), Aggregation Over Individuals and Demand Analysis, unpublished doctoral dissertation, Harvard University.

Stoker, T.M.(1982), "The Use of Cross Section Data to Characterize Macro Functions," Journal of the American Statistical Association, 77, pp.369-380.

Stoker, T.M.(1983a), "Omitted Variable Bias and Cross Section Regression," MIT Sloan School of Management Working Paper 1460-83, August.

Stoker, T.M.(1983b), "Aggregation, Structural Change and Cross Section Estimation," MIT Sloan School of Management Working Paper 1485-83, September, revised October 1984.

Stoker, T.M.(1984), "Completeness, Distribution Restrictions and the Form of Aggregate Functions," Econometrica, 52, pp. 887-908.

White, H.(1980a), "Using Least Squares to Approximate Unknown Regression Functions," International Economic Review, 21, pp.149-170.

White, H.(1980b), "Nonlinear Regression on Cross Section Data," Econometrica, 48, pp.721-746.