# A Study On Consistent Estimation of Scaled Coefficients

Tokumaru KIMURA

## Introduction

This paper considers the generic econometric modeling situation in which a dependent variable $y$ is modeled as a function of a vector of explanatory variables $X$ and stochastic terms, where the conditional expectation of $y$ given $X$ can be written in the single index form $E(y|X) = F(\alpha + X'\beta)$. This situation exists for many standard models of discrete choice, censoring, and selection, but is clearly not limited to such models. The question of interest is what can be learned about the coefficients $\beta$ without specific assumptions on the distribution of unobserved stochastic terms or other functional form aspects; in other words, when the true form of the function $F$ is misspecified or unknown.

And this paper proposes an approach for studying $\beta$ based on estimation of average behavioral derivatives, and shows how information on the marginal distribution of $X$ can be used to estimate average derivatives. In particular, a direct link between average derivatives and covariance estimators is established, which shows how $\beta$ can be estimated up to a scalar multiple by the sample covariance between $y$ and appropriately defined score vectors of the marginal distribution of $X$. $\beta$ is also consistently estimated up to scale by the slope coefficients of the linear equation of $y$ regressed on $X$ using the score vectors as instrumental variables.

## Method

Let us consider the situation where data are observed on a dependent variable $y_i$ and an M-vector of explanatory variables $X_i$ for $i=1, \ldots\ldots, N$, where $M>2$. $(y_i, X_i)$, $i=1, \ldots\ldots, N$, represent random drawing from a distribution D which is absolutely continuous with respect to a $\sigma$-finite measure $v$, with Radon-Nikodym density $P(y, X) = \partial D/\partial v$. $P(y, X)$ factors as $P(y, X) = q(y/X)\,p(X)$, where $p(X)$ is the density of the marginal distribution of $X$. The conditional density $q(y|X)$ represents the true behavioral econometric model, which we assume permits the conditional expectation $E(y|X)$ to be written in the form

( I ) $E(y|X) = F(\alpha + X'\beta) = F(Z)$

for some function $F$, where $\alpha$ is a constant, $\beta = (\beta_1, \ldots\ldots, \beta_M)'$ is an $M$-vector of constants, and $Z$ is defined as $Z = \alpha + X'\beta$. I can refer to $Z$ as an index variable, with ( I ) a single index model. This framework is very general, subsuming many limited dependent variables models, but is not restricted to such models. Before proceeding to specific examples, it is useful to note a generic special case of ( I ). Suppose that $Z^*$ is a general index variable such that $e = Z^* - Z$ is independent of $X$; then if $E(y|X, e) = F^*(Z^*)$ for some function $F^*$, ( I ) is implied. This includes many models that employ a latent variable $Z^* = \alpha + X'\beta + e$, where e is independent of $X$. Note also that this implies that behavioral variables can be omitted from $X$ without affecting the results, provided that the omitted variables are independent of the included ones[1). I now turn to some specific examples:

EXAMPLE 1 ——:

BINARY DISCRETE CHOICE: Suppose that $y$ represents a dichotomous random variable modeled as $y=1$ if $e > -(\alpha + X'\beta)$, and $y=0$ otherwise. Here, $E(y|X) = F(\alpha + X'\beta)$ is the probability of $y=1$ given the value of $X$, with the true function $F$ determined by the distribution of $e$. If $e$ is distributed normally with mean 0 and variance $\sigma^2$, then the familiar probit model results, with $F(\alpha + X'\beta) = C((\alpha + X'\beta)/\sigma)$, where C is the cumulative normal distribution function. Logit models, etc., can easily be included.

EXAMPLE 2 ——:

TOBIT MODELS: Suppose that $y$ is equal to an index $Z^*$ only if $Z^*$ is positive, as in the censored Tobit specification $y = \alpha + X'\beta + e$, if $e > -(\alpha + X'\beta)$, and $y$=o otherwise. Alternatively, if $y$ and $Y$ are observed only when $e > -(\alpha + X'\beta)$, we have the truncated Tobit specification.

EXAMPLE 3 ——:

DEPENDENT VARIABLE TRANSFORMATION: Suppose there exists a function $g(y)$ such that the true model is of the form $g(y) = \alpha + X'\beta + e$, where $g(y)$ is invertible everywhere except for a set of measure 0. A specific example is the familiar Box-Cox transformation where $y^{(\lambda)} = \alpha + X'\beta + e$ with $y^{(\lambda)} = [(y^{\lambda} - 1)/\lambda]$ for $\lambda \neq 0$, $y^{(\lambda)} = \ln(y)$ for $\lambda = 0$.

These examples serve to illustrate the wide-spectrum of models covered by the single index form ( I ) with general function $F$, and many examples can be found. Multiple index models will be considered later.

Now, we turn to the other required assumption. $X$ is assumed to be continuously distributed, having support $\Omega$ of the following form.[2]

(ASSUMPTION 1; $\Omega$ is a convex subset of $R^M$ with nonempty interior. The underlying measure $v$ can be written in product form as $v=v_y * v_X$, where $v_X$ is Lebegue measure on $R^M$.)

Therefore, no component of $X$ is functionally determined by other components of $X$, and no two components of $X$ are perfectly correlated. Denote $l(X)$ as the score vector[3] of the marginal density $p(X)$ as:

$$\text{(II)} \quad l(X) = -\frac{\partial \ln p(X)}{\partial X}.$$

The main regularity conditions on the marginal density $p(X)$ are given in the following assumptions.

(ASSUMPTION 2: $p(X)$ is continuously differentiable in the components of $X$ for all $X$ in the interior of $\Omega$. $E(l)$ and $E(ll')$ exist.)

(ASSUMPTION 3: For $X \epsilon d\Omega$, where $d\Omega$ is the boundary of $\Omega$, we have $p(X)=0$.)

ASSUMPTION 3 allows for unbounded $X$'s, where $\Omega = R^M$ and $d\Omega = 0$. While the majority of the results employ ASSUMPTIONS 2 and 3, the incorporation of discrete (qualitative) explanatory variables will be discussed later.

Many of the results will apply to a general random variable $\tilde{y}$ and its conditional expectation $E(\tilde{y}|X) \equiv G(X)$. For expositional simplicity, I refer to derivatives of conditional expectations, such as $\partial G/\partial X$, as behavioral derivatives. The regularity condition required for a general random variable $\tilde{y}$ holds if $(\tilde{y}, G(X))$ satisfies CONDITION A:

(CONDITION A: $G(X)$ is continuously differentiable for all $X \epsilon \bar{\Omega}$, where $\bar{\Omega}$ differs from $\Omega$ by a set of measure 0. $E(\tilde{y})$, $E(\partial G/\partial X)$, and $E(l\tilde{y})$ exist.)

The main regularity condition on the behavioral model (II) is contained in ASSUMPTION 4.

(ASSUMPTION 4: (a) $(y, F(\alpha + X'\beta))$ satisfies CONDITION A. $E(dF/dZ)$ is non-zero. (b) $(X_j, X_j)$ satisfies CONDITION A for each $j = 1, \ldots\ldots, M$.)

This completes the list of main assumptions. While somewhat formidable technically, these assumptions are collectively very weak.

The main thrust of this paper concerns how information on the marginal density $p(X)$ can be used to estimate $\beta$ up to scale. Consequently, the majority of the exposition assumes that the value of $l(X)$ at each $X_i$ is known, and denoted $l_i = l(X_i)$, $i = 1, \ldots\ldots, N$. Use of empirical characterizations of $p(X)$ will be discussed later. Finally, sample averages are denoted via overbars as in $\bar{y} = \sum y_i / N$, with the means of $y$ and $X$ denoted as $\mu_y = E(y)$ and $\mu_X = E(X)$. Sample covariances are denoted using $S$ as in $S_{ly} = \sum (l_i - \bar{l})(y_i - \bar{y})/N$, with population counterparts denoted using $\sum$ as in $\sum_{ly} = \mathrm{Cov}(l, y)$.

**Verifications and Polemical Points**

(1) Behavioral Derivatives and Covariance Estimators ……

Here, I consider a fundamental connection between derivatives and covariance estimators that is the basis of the consistency results of the following section. This connection is given in the THEOREM 1 which is interpreted after the proof.

THEOREM 1: Given ASSUMPTIONS 1-3, if $(\tilde{y}, G(X))$ satisfies CONDITION A, then

$$\text{(III)} \quad E\left[\frac{\partial G}{\partial X}\right] = E(l(X)\tilde{y}) = \textstyle\sum_{l\tilde{y}}.$$

* Proof of THEOREM 1 ——:

Let $X_1$ denote the first component of $X$, and $X_0$ the other components, so that $X = (X_1, X_0')'$. For a given value of $X_0$, denote the range of $X_1$ as $\mathrm{w}(X_0) = \{X_1 \mid (X_1, X_0')' \varepsilon\, \Omega\}$. Now, apply Fubini's THEOREM to write $E(\partial G/\partial X_1)$ as

$$\text{(IV)} \quad \int_\Omega \frac{\partial G(X)}{\partial X_1} p(X)\,dv = \int \left[\int_{\omega(X_0)} \frac{\partial G(X)}{\partial X_1} p(X)\,dv_1(X_1)\right] dv_0(X_0).$$

The result that $E(\partial G/\partial X_1) = \mathrm{E}(l(X)\tilde{y})$ is implied by the validity of the following equation:

$$\text{(V)} \quad \int_{\omega(X_0)} \frac{\partial G(X)}{\partial X_1} p(X)\,dv_1(X_1) = -\int_{\omega(X_0)} G(X) \frac{\partial p(X)}{\partial X_1} dv_1(X_1)$$

since the right-hand side of (V) simplifies to

$$\text{(VI)} \quad -\int_{\omega(X_0)} G(X) \frac{\partial p(X)}{\partial X_1} dv_1(X_1) = \int_{\omega(X_0)} G(X) \left[-\frac{\partial \ln p(X)}{\partial X_1}\right] p(X)\,dv_1(X_1).$$

By inserting (VI) into (IV), $E(\partial G(X)/\partial X_1) = E(l_1(X) G(X))$ is established, and by iterated expectation, $E(l_1(X)G(X))=E(l_1(X)\tilde{y})$. To establish (V), note that the convexity of $\Omega$ implies that $w(X_0)$ is either a finite interval $[a, b]$ (where $a$, $b$ depend on $X_0$), or an infinite interval of the form $[a, \infty]$, $(-\infty, b]$, or $(-\infty, \infty)$. Supposing that $w(X_0)=[a, b]$, integrate the left-hand side of (V) by parts as

$$\text{(VII)}\quad \int_a^b \frac{\partial G(X)}{\partial X_1} p(X)\,dv_1(X_1) = -\int_a^b G(X)\frac{\partial p(X)}{\partial X_1}dv_1(X_1) + G(b, X_0)p(b, X_0) - G(a, X_0)p(a, X_0).$$

The latter two terms represent $G(X)\,p(X)$ evaluated at boundary points, which vanish by ASSUMPTION 3, so that (V) is established for $w(X_0)=[a, b]$. For the unbounded case $w(X_0)=[a, \infty)$, note first that the existence of $E(\tilde{y})$, $E(\partial G/\partial X_1)$, and $E(l_1(X)\tilde{y})$ respectively imply the existence of $E(G(X)|X_0)$, $E(\partial G/\partial X_1/X_0)$, and $E(l_1(X)G(X)|X_0)$. Now let us consider the limit of (VII) over intervals $[a, b]$, where $b\to\infty$, rewritten as

$$\text{(VIII)}\quad \lim_{b\to\infty} G(b, X_0)p(b, X_0)=G(a, X_0)$$

$$=G(a, X_0)p(a, X_0)+\lim_{b\to\infty}\int_a^b \frac{\partial G(X)}{\partial X_1}p(X)\,dv_1(X_1)+\lim_{b\to\infty}\int_a^b G(X)\frac{\partial p(X)}{\partial X_1}dv_1(X_1)$$

$$=G(a, X_0)p(a, X_0)+p_0(X_0)E\left[\frac{\partial G}{\partial X_1}|X_0\right]-p_0(X_0)E(l_1(X)G(X)|X_0).$$

So that $C\equiv\lim G(b, X_0)p(b, X_0)$ exists, where $p_0(X_0)$ is the marginal density of $X_0$. Now suppose that $C>0$. Then there exists scalars $e$ and $B$ such that $0<e<C$ and for all $b>B$, $|G(b, X_0)p(b, X_0)-C|<e$. Therefore, $G(X_1, X_0)p(X_1, X_0)>(C-e)\,I_{[B,\infty)}$, where $I_{[B,\infty)}$ is the indicator function of $[B, \infty)$. But this implies that $p_0(X_0)E(G(X)|X_0)=$ $\int G(X_1, X_0)p(X_1, X_0)\,dv_1(X_1)>(C-e)\int I_{[B,\infty)}dv_1(X_1)=\infty$, which contradicts the existence of $E(G(X)|X_0)$. Consequently, $C>0$ is ruled out. $C<0$ similarly contradicts the existence of $E(G(X)|X_0)$. Since $C\equiv\lim G(b, X_0)p(b, X_0)=0$, and $G(a, X_0)p(a, X_0)=0$ by ASSUMPTION 3, equation (V) is valid for $w(X_0)=[a, \infty)$. Analogous arguments establish the validity of (V) for $w(X_0)=(-\infty, a]$ and $w(X_0)=(-\infty, \infty)$. The second equality of (III), $E(l_1(X)\tilde{y})=\mathrm{Cov}(l_1(X), \tilde{y})$, is true becasue the mean of $l_1(X)$ is $0$[4]. The proof is completed by repeating the same development for derivatives of $G(X)$ wiht respect to $X_2, \ldots\ldots, X_M$. Q.E.D.

THEOREM 1 is of significant theoretical interest. It says that the average behavioral derivative $E(\partial G/\partial X)$ can be written as the covariance between $\tilde{y}$ and a function of $X$; namely $l(X)$. The form of $l(X)$ does not depend on the behavioral relation $E(\tilde{y}|X)=G(X)$; $l(X)$ is determined by the marginal density $p(X)$. Thus, THEOREM 1 establishes a general link between behavioral derivatives and covariance estimators that does not depend on assumptions on the form of behavior[5]. The proof is extremely simple, based on integration-by-parts. An useful intuition for THEOREM 1 can be obtained from its connection to results in aggregation theory. In particular, THEOREM 1 reflects the local aggregate effects on $E(\tilde{y})$ of translating the base density $p(X)$. To see this connection, consider the unbounded case where $\Omega=R^M$. Suppose that the base density is translated by an $M$-vector $\theta$: $p(X)$ is altered to $p(X-\theta)$ for all $X$. The value of $E(\tilde{y})$ after this translation is given as

$$\text{(IX)} \quad E(\tilde{y}|\theta)=\int_\Omega G(X)p(X-\theta)dv.$$

By a change of variances, $E(\tilde{y}|\theta)$ can be written as

$$\text{(X)} \quad E(\tilde{y}|\theta)=\int_\Omega G(X+\theta)p(X)dv.$$

The local aggregate effects of the translation are the derivatives $\partial E(\tilde{y}|\theta)/\partial\theta$ evaluated at $\theta=0$[6]. Differentiating (IX) under the integral sign and evaluating at $\theta=0$ gives

$$\text{(XI)} \quad \frac{\partial E(\tilde{y}|0)}{\partial\theta}=\int_\Omega G(X)\frac{\partial p}{\partial\theta}dv=$$

$$\int_\Omega G(X)\frac{\partial \ln p}{\partial\theta}p(X)dv=\int_\Omega G(X)l(X)p(X)dv$$

where the latter equality reflects that $l(X)$ equals $\partial\alpha\ln p(X-\theta)/\partial\theta$ evaluated at $\theta=0$. Similarly, differentiating (X) gives

$$\text{(XII)} \quad \frac{\partial E(\tilde{y}|0)}{\partial\theta}=\int_\Omega\frac{\partial G}{\partial\theta}p(X)dv=$$

$$\int_\Omega\frac{\partial G}{\partial X}p(X)dv.$$

Collecting the equalities of (XI) and (XII) gives $E(G(X)l(X))=E(\partial G/\partial X)$, which underlies equation (III) of THEOREM 1.

THEOREM 1 has a simple geometric explanation. For evaluating the mean $E(\tilde{y})$ under translation, one can average $G(X)$ over the distribution $p(X)$ shifted by $\theta$ (equation (IX)), or one can shift $G(X)$ by $-\theta$ and average over the distribution $p(X)$ (equation (X)). The local effects on $E(\tilde{y})$ can be computed from either perspective (equation (XI) and (XII)) to yield the same value. Equation (III) exhibits this equivalence.

( 2 ) Consistent Estimation of Scaled Coefficients ......

This section indicates how to estimate $\beta$ up to scale for single index models of the form ( I ). The ITEM ( 2 ) - 1 indicates the basic approach and proposes a covariance estimator and an instrumental variables estimator. The ITEM ( 2 ) - 2 discusses immediate extensions of the basic results.

ITEM ( 2 ) - 1 : The average derivative approach to estimation ——:

Begin by considering a precise empirical implication of the single index model form $E(y|X)=F(\alpha+X'\beta)$. Clearly, the conditional mean of $y$ depends on $X$ through the value of $X'\beta$. By exploiting differentiability, a precise restriction of the single index form is given as

$$\text{(XIII)} \quad \frac{\partial E(y|X)}{\partial X}=\frac{\partial F(\alpha+X'\beta)}{\partial X}=\left[\frac{dF}{dZ}\right].$$

Thus, $\partial E(y|X)/\partial X$ is proportional to $\beta$, although the scale factor $dF/dZ$ will depend on the value of $X$ chosen, The basic approach in this paper is to focus on the average of the constraint (XIII):

$$\text{(XIV)} \quad E\left[\frac{\partial E(y|X)}{\partial X}\right]=E\left[\frac{\partial F}{\partial X}\right]=E\left[\frac{dF}{dZ}\right]\cdot\beta=\gamma\beta$$

Where $\gamma=E(dF/dZ)$ exists and is nonzero by ASSUMPTION 4. Clearly, any consistent estimator of the average derivative $E(\partial F/\partial X)$ is a consistent estimator of $\beta$ up to scale. Two natural consistent estimators are suggested by THEOREM 1. Firstly, I can define the estimator $\hat{d}_0$ as the sample covariance between $y_i$ and $l_i$:

$$\text{(XV)} \quad \hat{d}_0=S_{ly}.$$

The second estimator is more closely related to standard regression estimators, such as the OLS coefficients of $y$ regressed on $X$. Define $\hat{d}$ as the instrumental variables coefficients of the regression

(XVI) $y_i = \hat{c} + X_i \hat{d} + \hat{u}_i$

obtained using $(1, l_i)'$ as the instrumental variable, namely

(XVII) $\hat{d} = (S_{lX})^{-1} S_{ly}.$

The consistency of $\hat{d}_0$ and $\hat{d}$ for $\alpha\beta$ follows immediately from THEOREM 1, as in THEOREM 2.

THEOREM 2: Given ASSUMPTIONS 1 - 4, $\hat{d}_0$ and $\hat{d}$ are strongly consistent estimators of $\gamma\beta$, where $\gamma = E(dF/dZ)$.

* Proof of THEOREM 2 ——:

The Strong Law of Large Numbers implies that lim $S_{ly} = \Sigma_{ly}$. THEOREM 1 and (XIV) imply that lim $\hat{d}_0 = \gamma\beta$ a.s.. If lim $S_{lX} = \Sigma_{lX} =$ I, an $M*M$ identity matrix, then lim $\hat{d} = \gamma\beta$ a.s. follows. In view of ASSUMPTION 4 (b), THEOREM 1 can be applied with $\tilde{y} = X_j$, for each $j = 1, \ldots, M$. Carrying this out gives

$$\Sigma_{lX} = I. \qquad \text{Q. E. D.}$$

The two estimators $\hat{d}_0$ and $\hat{d}$ appear very similar; however in general they are not first-order $(\sqrt{N})$ equivalent. In particular,

(XVIII) $\sqrt{N}(d - \hat{d}_0) = \sqrt{N}(S_X^{-1} - \mathrm{I})_0^{d}.$

Since lim $\hat{d}_0 = \alpha\beta \neq 0$, and $\sqrt{N}(S_X^{-1} - \mathrm{I})$ in general has a nontrivial limiting distribution, $\sqrt{N}(\hat{d} - \hat{d}_0)$ will not vanish as $N \to \infty$. For expository purposes, I will refer to $\hat{d}$ for the remainder of the exposition; however, all consistency results can be extended to $\hat{d}_0$[7]. The connection to the aggregate effects of translation permits a further interpretation of the scale factor $\gamma = E(dF/dZ)$. The structure of the single index model (I) implies that the local aggregate effects of translation are proportional to the parameters of interest. In particular, insert (XIII) into (XI), giving

(XIX) $\frac{\partial E(y|0)}{\partial \theta}=\int_{\Omega}\frac{dF}{dZ}\cdot\beta p(X)dv=\alpha\beta.$

This appearance of $\beta$ is due to the correspondence between density translation and the linear form of the index $Z=\alpha+X'\beta$. To interpret $\gamma$, note that under translation, the marginal distribution of Z is shifted by the parameter $\eta=\theta'\beta$, with the mean of $Z$ increased by $\eta$. (XIX) can be regarded as the chain rule formula $\partial E(y)/\partial\theta=(dE(y)/d\eta)(\partial\eta/\partial\theta)$, where $\partial\eta/\partial\theta$ is equal to $\beta$. The scale factor $\gamma$ is equal to $dE(y)/d\eta$, the effect on $E(y)$ induced by a change in the mean $E(Z)$ of the index variable $Z$.

ITEM (2)-2: Extraneous variables and multiple index models ——:

The approach of parameter estimation via average derivatives easily extends to more general models than those relying on a single index. In this ITEM, I consider some immediate extensions, namely to models with extraneous variables and multiple index models. Begin by expanding the notation to consider two sets of explanatory variables; an $M_1$-vector $X_1$ and an $M_2$-vector $X_2$, distributed with density $p(X_1, X_2)$[8].

Consider first the case where $X_2$ represents extraneous variables, in that the behavioral model for $y$ implies

(XX) $E(y|X_1, X_2)=F(\alpha_1+X_1'\beta_1, X_2)=F(Z_1, X_2)$

for some function $F$ with constant coefficients $\alpha_1$, $\beta_1$, and $Z_1=\alpha_1+X_1'\cdot\beta_1$. In this case, $\beta_1$ is proportional to the (partial) derivative of F with respect to $X_1$,

(XXI) $\frac{\partial E(y|X_1, X_2)}{\partial X_1}=\frac{\partial F}{\partial X_1}=\left[\frac{\partial F}{\partial Z_1}\right]\cdot\beta_1$

so that the average derivative is proportional to $\beta_1$:

(XXII) $E\left[\frac{\partial F}{\partial X_1}\right]=E\left[\frac{\partial F}{\partial Z_1}\right]\beta_1=\gamma_1\cdot\beta_1.$

THEOREMS 1 - 2 can be applied as long as the appropriate analogues of ASSUMPTIONS 1 - 4 applied to $X_1$ are valid. In particular, the proof of THEOREM 1 will apply to individual components of $X_1$ provided that no two components of $X_1$, $X_2$ are perfectly correlated, and that the conditional density $p_1^c(X_1|X_2)$ vanishes on the boundary of $X_1$ values for each value of $X_2$. Under these conditions, the sample covariance $\hat{d}_{10}=s_{l_1y}$ consistently estimates $\gamma_1 \cdot \beta_1$, where the partial score $l_{1i}=l_1(X_{1i}, X_{2i})$ is defined via

$$\text{(XXIII)} \quad l_1(X_1, X_2) = -\frac{\partial \ln p(X_1, X_2)}{\partial X_1} = -\frac{\partial \ln p_1^c(X_1|X_2)}{\partial X_1}$$

Moreover, $\gamma_1 \cdot \beta_1$ is consistently estimated by the slope coefficient estimates $\hat{d}_1$ of the linear equation

$$\text{(XXIV)} \quad y_i = \hat{c}_1 + X_{1i}\hat{d}_1 + \hat{u}_{1i}$$

obtained by instrumenting with $(1, l_{1i}')'$. Thus, the extraneous variables $X_2$ are accommodated in the estimation of $\beta_1$ by modification of the appropriate instrumental variables, to reflect the joint distribution of $X_1$ and $X_2$. Clearly, if $X_2$ were distributed independently of $X_1$, then $X_2$ can be ignored in the estimation of $\beta_1$ up to scale[9]. This extension provides an initial response as to how to accommodate discrete explanatory variables into the analysis. If $X_2$ is composed of discrete variables, an approach based on average derivatives is not obviously applicable to estimating effects of $X_2$. However, the coefficients of the remaining continuous variables $X_1$ can be estimated up to scale by using the score vectors of the conditional density of $X_1$ (given the observed values of $X_2$) as instrumental variables. Consequently, while the analysis is silent on how to estimate coefficients of discrete variables, their presence does not prohibit the estimation of continuous variable coefficients up to scale. Putting aside this proviso on discrete variables, I now turn to multiple index models[10]. All relevant points are exhibited by two index models, so assume that $X=(X_1', X_2')'$ is composed entirely of continuous variables with $M_1 > 2$ and $M_2 > 2$. Suppose that the behavioral model implies the following two index form:

$$\text{(XXV)} \quad E(y|X) = F(\alpha_1 + X_1'\beta_1, \alpha_2 + X_2'\beta_2) = \mathrm{F}(X_1, Z_2)$$

where $Z_1 = \alpha_1 + X_1'\beta_1$ and $Z_2 = \alpha_2 + X_2'\beta_2$ represent the two index variables. The derivative of the conditional expectation now takes the form

$$\text{(XXVI)} \quad \frac{\partial E(y|X)}{\partial X} = \frac{\partial F}{\partial X} = \left[\frac{\partial F}{\partial Z_1}\right] \cdot \beta_1 \quad \left[\frac{\partial F}{\partial Z_2}\right] \cdot \beta_2,$$

So that the average derivative is

$$\text{(XXVII)} \quad E\left[\frac{\partial F}{\partial X}\right] = \left[\frac{\alpha_1 \beta_1}{\alpha_2 \beta_2}\right]$$

where $\gamma_1 = E(\partial F/\partial Z_1)$ and $\gamma_2 = E(\partial F/\partial Z_2)$ are scalar constants. Thus a consistent estimator of the average derivative will estimate $\beta_1$ and $\beta_2$ up to scale; however the scale factors $\gamma_1$ and $\gamma_2$ will differ in general. Such a consistent estimator has already been established, provided that y, $F$ of (XXV) obey CONDITION A. Namely, the estimator $\hat{d}$ of (XVII) consistently estimates $E(\partial F/\partial X)$, so that its components corresponding to $X$ estimate $\beta$ up to scale, the main modeling limitation of this result is that no two components of $X_1$ and $X_2$ may be functionally related or perfectly correlated. Thus, the index variables $Z_1$ and $Z_2$ may have no common component variables, an exclusion restriction that is required for estimating both $\beta_1$ and $\beta_2$ up to scale using average first derivatives. The following example gives a two index model, where $\gamma_1 = 1$ a priori.

## Conclusion

This paper proposes an approach to parameter estimation based on average behavioral derivatives, and applies the approach to the estimation of $\beta$ up to scale in single index models. The proposed estimators explicitly utilize information on the marginal distribution of the explanatory variables in the model. The framework is illustrated using several examples of limited dependent variables models, and extended to multiple index models. The asymptotic biases in OLS coefficients are characterized vis-a-vis the distribution of explanatory variables. There are two major advantages of the proposed estimator $\hat{d}$. First, $\hat{d}$ is nonparametric to the extent that it is robust to many specific functional form and stochastic distribution assumptions. If a particular application requires only estimates of the ratios of components of $\beta$, then $\hat{d}$ will suffice. In a general application where different sets of assumptions give rise to different estimates of $\beta$, $\hat{d}$ will provide a benchmark estimate for choosing the best specification. Given parametric modeling of the explanatory variable distribution, the precision of the components of $\hat{d}$ can be measured, and tests of scalefree on the value can be performed. The other advantage of $\hat{d}$ is computational simplicity. Once the distribution of explanatory variables is characterized, $\hat{d}$ (as well as $\hat{d}_0$) is a linear estimator, computed entirely from sample covariances. This suggests that implementation may be particularly easy and inexpensive, especially for large data bases.

## Notes

1 ) This framework differs from that of Chung and Goldberger and Deaton and Irish, since those papers only require e to be uncorrelated with $X$.

2 ) The support $\Omega$ is defined as the closure of the set $\{X \cdot \varepsilon R^{M1} | p(X) > 0\}$ .

3 ) The terminology is due to the fact that $l(X)$ is the score vector of p with respect to a translation parameter.

4 ) This is shown by noting that CONDITION A is satisfied by $\tilde{y} = G(X) = 1$ , a constant variable, and by applying (V), (VI).

5 ) A similar link is used to establish the consistency of OLS estimators for the standard linear model, namely, the functional form assumption that $E(\tilde{y}|X) = G(X) = \alpha + X'\beta$ implies $\mathrm{Cov}(X, \tilde{y}) = \Sigma_{XX}\beta$, or $\mathrm{Cov}(\Sigma_{XX}^{-1} X, y) = \beta$. By the same assumption, the behavioral effects are $\beta = \partial G(X) / \partial X = E(\partial G(X) / \partial X)$.

6 ) Stoker gives a general development of local aggregate, or macroeconomic effects.

7 ) Other consistent estimators of $\alpha\beta$ include the production moment estimator $\hat{d} = \Sigma_i y_i / \mathrm{N}$, the reduced form OLS estimator of the slope coefficients of $y_i = c_2 + d_2 \hat{X}_i + u_{2i}$, where $\hat{X}_i = (\Sigma_{1ll}^{-})^{-1} l_i$, and the weighted OLS ESTIMATOR PROPOSED by Ruud. None of these estimators are first-order equivalent to either $\hat{d}$ or $\hat{d}_0$ in general.

8 ) This expanded notation is used in this ITEM only.

9 ) $X_2$ then takes on the same role as the random term e of this ITEM.

10) Notice that $\hat{d}_1$ of (XXIV) consistently estimates $\gamma_1\beta_1$, and that the analogous coefficients from the linear equation with $X_2$ as explanatory variables will estimate $\gamma_2\beta_2$.

## 要　　　約

論題：スケール係数一般の一致推定に関する考察——：

この小論ではたとえば，$E(y|X) = F(\alpha + X'\beta)$ のような指標関数モデルの係数$\beta$に関する推定問題を考える．その各論点は次の通り．１．行動因子変数一般の微分値と共分散推定子との相関度の確定．２．係数$\beta$の推定値の，独立変数 $X$ の限界分布情報できまるスケールをテコに

した算定. **3**. $X$ の分布スコアベクトルを用いた標本共分散推定子と手段変数の係数ベクトルとの確定. **4**. 制約条件つき従属変数モデルを用いた$\beta$の推定作業の枠組みづくり. **5**. この作業の，バイアスつき多元名目型不連続選好モデルを内臓する指標関数モデルへの拡張. **6**. 手段変数の推定子の逐次型分布の，$X$ の分布モデルづくりでの有限型係数化操作による確定. **7** 従属変数 $y$ の，$X$ の OLS 型回帰係数の逐次型バイアスによる算定.

(付記)：本稿は昨年の12月28～30日，米国シカゴ大学で開催された1987年度国際計量経済学会年次大会での筆者の研究報告論文を修正・加筆したものである（なお拙論の同学会での論文審査がパスした日付は1987年 6 月24日である）