SEMIPARAMETRIC SINGLE-INDEX MODELS

by

Joel L. Horowitz
Department of Economics
Northwestern University

INTRODUCTION

- Much of applied econometrics and statistics involves estimating a conditional mean function:

  $$E(Y \mid X = x)$$

  - Y may be continuous or binary
  - If binary, then $E(Y \mid X = x)$ is $P(Y = 1 \mid X = x)$
    - In binary response model, Y may indicate an individual's choice among two alternatives, occurrence or non-occurrence of an event, etc.

- Possible approaches
  - Fully parametric
  - Fully nonparametric
  - Semiparametric

FULLY PARAMETRIC MODELING

- In fully parametric model, $\boldsymbol{E}(Y \mid X = x)$ is known up to a finite-dimensional parameter:

$$\boldsymbol{E}(Y = 1 \mid X = x) = F(x, \theta)$$

  - F is known function
  - θ is unknown, finite-dimensional parameter
  - Example: binary probit or logit model
- Advantages: If F is correctly specified
  - Maximizes estimation efficiency
  - Permits extrapolation of x beyond range of data
  - Often has natural behavioral interpretation
- Disadvantages:
  - F rarely known in applications
  - Can be highly misleading if F is misspecified

FULLY NONPARAMETRIC MODELING

- $\boldsymbol{E}(Y \mid X = x) \equiv G(x)$ assumed to be smooth function of x
  - Nothing assumed about shape of G.
  - G estimated by nonparametric mean regression of Y on X
- This minimizes *a priori* assumptions and likelihood of specification error
- Disadvantages:
  - Hard to incorporate behavioral hypotheses drawn from economic or other theory models
  - Estimation precision is exponentially decreasing function of dimension of X
  - Extrapolation not possible

SEMIPARAMETRIC MODELING

- Achieves greater precision than nonparametric models but with weaker assumptions than parametric models

  - Does this by restricting $G(x)$ so as to reduce effective dimension of x.

  - Risk of specification error greater than with fully nonparametric model but less than with parametric one

- Examples:

  - Single-index model:

    $$G(x) = F(x\beta),$$

    where F is unknown

  - Additive model:

    $$G(x) = H[f_1(x_1) + ... + f_d(x_d)],$$

    where H is known or unknown function and f_i's are unknown

IDENTIFICATION OF SINGLE-INDEX MODELS

$$\boldsymbol{E}(Y \mid X = x) = G(x\beta)$$

- β not identified if G is constant function.

- Sign, scale, and location normalizations needed to identify β

  - To implement assume X has no intercept and $\beta_1 = 1$.

- X_1 must be continuously distributed conditional on other components of X.

  - Let $X = (X_1, X_2)'$ and $X'\beta = X_1 + \beta_2 X_2$.

  - G and β_2 can be anything that satisfy:

(X_1, X_2)	$G(X_1 + \beta_2 X_2)$	$E(Y \mid X)$
(0,0)	$G(0)$	0
(1,0)	$G(1)$	1
(0,1)	$G(\beta_2)$	3
(1,1)	$G(1 + \beta_2)$	4

OPTIMZATION ESTIMATORS

- If G known, β can be estimated by nonlinear least squares.

$$\underset{b}{\text{minimize:}}\, n^{-1}\sum_{i=1}^{n} w(X_i)[Y_i - G(X_i b]^2$$

  where $w(\cdot)$ is a weight function.

- When G unknown, replace $G(X_i b)$ with non-parametric estimator of $E(Y|X_i b)$ (e.g., kernel).
  - Estimator now solves

$$\underset{b}{\text{minimize:}}\, n^{-1}\sum_{i=1}^{n} w(X_i)[Y_i - G_n(X_i b)]^2$$

- w may be chosen to
  - Keep denominator of G away from 0
  - Achieve asymptotic efficiency

ASYMPTOTIC NORMALITY

- Ichimura (1993) gives conditions under which

$$n^{1/2}(b_n - \beta) \to N(0,V)$$

  where b_n is weighted NLS estimator

  - Proof based on standard Taylor series methods of asymptotic distribution theory

- Estimator has $n^{-1/2}$ rate of convergence

- Hall and Ichimura (1991) derived asymptotic efficiency bound for β in

$$Y_i = G(X_i\beta) + \sigma(X_i\beta)U_i$$

  where the U_i are iid with mean 0

- Hall and Ichimura also derived asymptotically efficient estimator

  - Uses estimate of $\sigma(X_i\beta)^{-1}$ as weight function in NLS objective function and kernel estimator of G.

MLE FOR BINARY RESPONSE MODEL

- If $Y = 0$ or 1, $G(x\beta) = P(Y=1|X=x)$

- If G known, log likelihood is

$$\log L(b) = \sum_{i=1}^{n} \left\{ \log G(X_i b) + (1 - Y_i) \log\left[1 - G(X_i\, b)\right] \right\}$$

- If G unknown, replace it with estimator G_n

$$\log L(b) =$$
$$\sum_{i=1}^{n} \tau_i \left\{ \log G_n(X_i b) + (1 - Y_i) \log\left[1 - G_n(X_i\, b)\right] \right\}$$

- τ_i trims away observations for which $G_n(X_i b)$ is too close to 0 or 1.

- Klein and Spady (1993) gave conditions under which semiparametric MLE estimator is $n^{1/2}$-consistent and asymptotically normal

- Chamberlain (1986) and Cosslett (1987) derived asymptotic efficiency bound for case in which G is a CDF

    - Semiparametric MLE achieves bound

DIRECT ESTIMATORS

- NLS and ML estimators are hard to compute
- Direct estimators avoid need to solve optimization problem
  - Direct estimators are not asymptotically efficient
  - Efficient estimator can be obtained easily by one-step method
- If X is continuous random vector, β proportional to average derivative of G
  - $\beta \propto \boldsymbol{E}\left[w(X)\partial G(X\beta)\partial X\right]$

  where w is a weight function
- Only weighted average derivative needed because β identified only up to scale
- If w is identity function, get average derivative estimator of β (Härdle and Stoker 1989)
  - This estimator is hard to analyze because of its random denominator

DENSITY WEIGHTED AVERAGE DERIVATIVE ESTIMATORS

- Random denominator problem can be overcome by setting $w(x) = f(x)$, density of X

- Integration by parts gives

$$\delta \equiv \boldsymbol{E}\left[f(X)\partial G(X\beta)\partial X\right]$$

$$= -2E\left[G(X\beta)\partial f(X)/\partial X\right]$$

$$= -2E\left[Y\partial f(X)\partial X\right]$$

- Estimate δ by replacing E with sample average and f with kernel estimator to get

$$\delta_n = (-2/n)\sum_{i=1}^{n} Y_i\left[\frac{\partial f_i(X_i)}{\partial x}\right]$$

where f_i is leave-one-out kernel estimator of $f(x)$.

- Powell, Stock, and Stoker (1989) gave conditions under which $n^{1/2}(\delta_n - \delta) \to N(0, V)$

METHOD OF PROOF

- Write δ_n as U statistic of order 2 with bandwidth-dependent kernel

- U statistic is asymptotically equivalent to its projection, which gives

$$\delta_n = (2/n)\sum_{i=1}^{n} r_n(Y_i, X_i) + o_p(n^{-1/2}),$$

where

$$r_n(Y_i, X_i) =$$

$$-\int \left(\frac{1}{h}\right)^{k+1} K'\left(\frac{X_i - x}{h}\right)\left[Y_i - E(Y \mid X = x)\right] f(x)dx$$

- Changing variables in integral shows that leading term of r_n does not depend on h or n

- So δ_n is asymptotically equivalent to a sum of iid random variables

- $n^{-1/2}$-consistency and asymptotic normality follow from Lindeberg-Levy theorem

TECHNICAL DETAILS

- Must use higher-order K with undersmoothing to insure that asymptotic distribution of $n^{1/2}(\delta_n - \delta)$ is centered at 0.

- Härdle and Tsybakov (1993) and Powell and Stoker (1996) describe methods for selecting h in applications.

- Horowitz and Härdle (1996) show how to include discrete components of X in direct estimator.

ESTIMATOR WITH DISCRETE COVARIATES

- Write model as $E(Y|X = x, Z = z) = G(X\beta + Z\alpha)$, where X is continuous and Z is discrete with M points of support.
  - Identification requires a continuous covariate
  - Assume estimator of β, b_n is available, possibly average of average derivative estimates computed at each point in support of Z.
- Suppose there are finite numbers c_0, c_1, v_0, v_1 such that
  - $G(v + z\alpha)$ is bounded for all $v \in [v_0, v_1]$ and $z \in \text{supp}(Z)$.
  - $v \le v_0 \Rightarrow G(v + z\alpha) \le c_0$ for each $z \in \text{supp}(Z)$
  - $v \ge v_1 \Rightarrow G(v + z\alpha) > c_1$ for each $z \in \text{supp}(Z)$
- Define

$$J(z) = \int_{v_0}^{v_1} \{c_0 I[G(v + z\alpha) < c_0] + c_1 I[G(v + z\alpha) > c_1]$$

$$+ G(v + z\alpha) I[c_0 \le G(v + z\alpha) \le c_1]\} dv$$

DISCRETE COVARIATES (cont.)

- Then for $i = 2,\ldots,M$

$$J[z^{(i)}] - J[z^{(1)}] = (c_1 - c_0)[z^{(i)} - z^{(1)}]\alpha.$$

- This is M - 1 linear equations in components of α. To solve, write

$$\Delta J = \begin{bmatrix} J[z^{(2)}] - J[z^{(1)}] \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ J[z^{(M)}] - J[z^{(1)}] \end{bmatrix}; \qquad W = \begin{bmatrix} z^{(2)} - z^{(1)} \\ \ldots\ldots\ldots\ldots\ldots \\ z^{(M)} - z^{(1)} \end{bmatrix}.$$

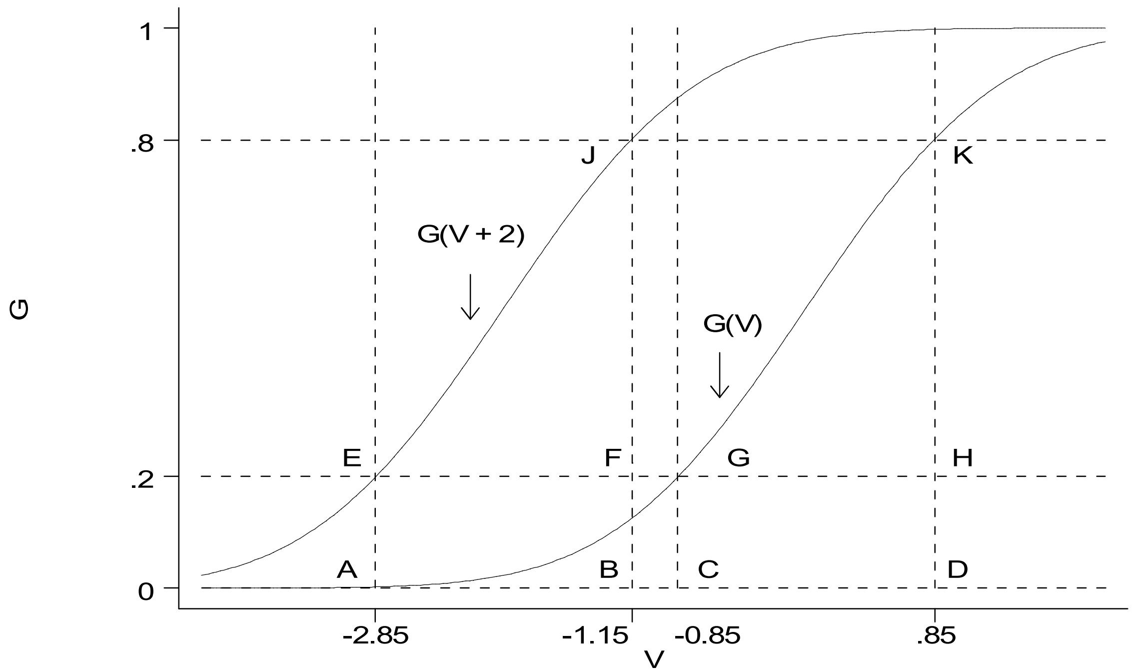
- Then

$$\alpha = (c_1 - c_0)^{-1}(W'W)^{-1}W'\Delta J.$$

- Obtain estimator by replacing G with non-parametric regression estimate of $\boldsymbol{E}(Y \mid Xb_n = v, Z = z)$.
  - Let ΔJ_n be resulting estimator of ΔJ
  - Estimator of α is

$$\alpha_n = (c_0 - c_1)^{-1}(W'W)^{-1}W'\Delta J_n$$

- Horowitz and Härdle (1996) give conditions under which $n^{1/2}(\alpha_n - \alpha) \to^d N(0, V_\alpha)$.



- $(c_0, c_1) = (0.2, 0.8),\ (v_0, v_1) = -(2.85, 0.85)$

$J[z^{(1)}] = ACGE + CDHG + GHK$

- $$= 2c_0 + 1.7c_0 + GHK$$

$J[z^{(2)}] = ABFE + BDKJ + EFJ$

- $$= 1.7c_0 + 2c_1 + EFJ$$

- $J[z^{(2)}] - J[z^{(1)}] = 2(c_1 - c_0) = (c_1 - c_0)[z^{(2)} - z^{(1)}]\alpha$

HIGH-DIMENSIONAL *X*

- Average derivative estimators require G and f to have many derivatives) if X is high dimensional.
  - This is form of curse of dimensionality
  - Implies that finite-sample precision of average derivatives may be low if $\dim(X)$ large.

- Hristache, Juditsky, and Spokoiny (2001) proposed method for iteratively improving an average derivative estimator.

- Method uses two bandwidths: a large one in the direction orthogonal to current estimate and a small one in parallel direction.
  - Calculate new estimate of β using average derivatives with the two bandwidths

- This procedure yields estimator that is $n^{-1/2}$-consistent and asymptotically normal regardless of dimension of X when G is twice differentiable.

- Monte Carlo evidence indicates that iterated estimator has smaller finite-sample errors than non-iterated one.

OUTLINE OF ITERATIVE METHOD

- Initialization: Specify parameters ρ_1, $\rho_{\min}$, a_ρ, h_1, $h_{\max}$, a_h, $k=1$, $\hat{\beta}_0$ (initial estimate of β)

- Compute $S_k = (I + \rho_k^{-2}\hat{\beta}_{k-1}\hat{\beta}'_{k-1})^{1/2}$

- For every $i = 1,...,n$, compute $\nabla\hat{f}_k(X_i)$ from

$$\begin{bmatrix} \hat{f}_k(X_i) \\ \nabla\hat{f}_k(X_i) \end{bmatrix} = \left[\sum_{j=1}^{n} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}' K\left(\frac{|S_k X_{ij}|^2}{h_k^2} \right) \right]^{-1}$$

$$\times \sum_{j=1}^{n} Y_j \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K\left(\frac{|S_k X_{ij}|^2}{h_k^2} \right)$$

where $X_{ij} = X_j - X_i$

- Compute $\hat{\beta}_k = n^{-1}\sum_{j=1}^{n} \nabla\hat{f}_k(X_i)$

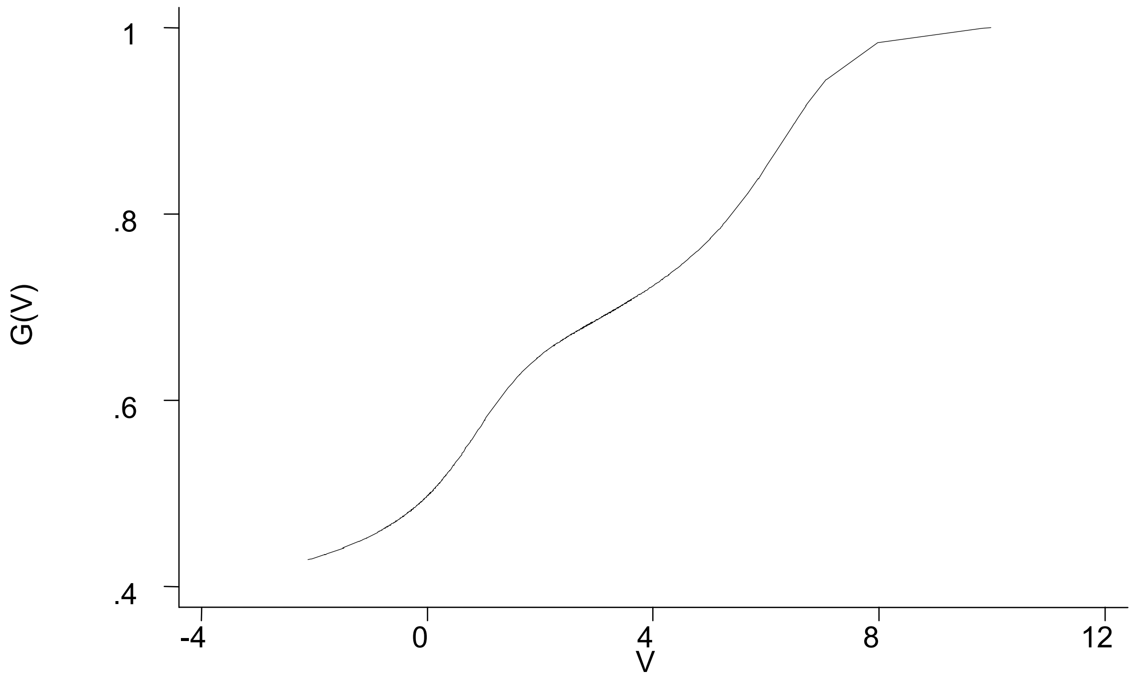
- Set $h_{k+1} = a_h h_k$, $\rho_{k+1} = a_\rho \rho_k$. If $\rho_{k+1} > \rho_{\min}$, set $k = k+1$ and return to step 2. Otherwise, stop.

AN APPLICATION

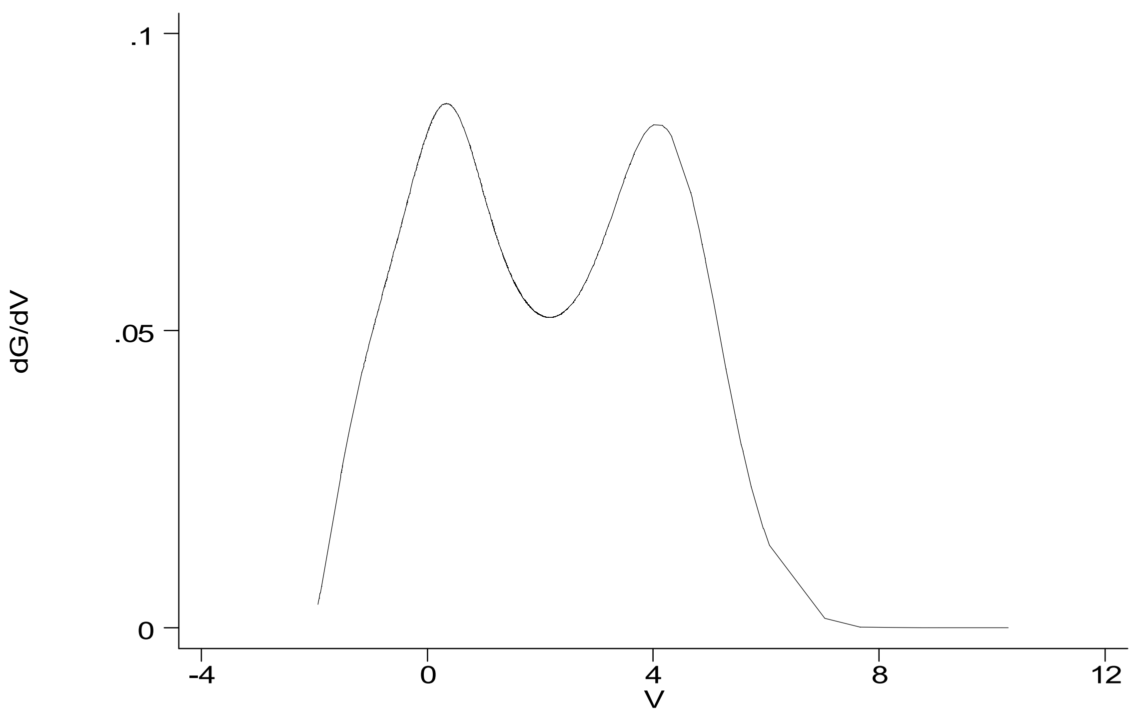
- Model of product innovation by German manufacturers of investment goods
    - Data assembled by IFO Institute in Munich
    - Consist of observations on 1100 manufacturers
- Model: $P(Y=1|X=x) = G(X\beta)$, where
    - $Y = 1$ if manufacturer realized an innovation in a specific product category in 1989 and 0 otherwise
    - Variables: no. of employees in product category (EMPLP), no. of employees in entire firm (EMPLF), indicator of firm's production capacity utilization (CAP), DEM = 1 if firm expected increasing demand for product and 0 otherwise

ESTIMATED COEFFICIENTS FOR MODEL OF PRODUCT INNOVATION

EMPLP	EMPLF	CAP	DEM
	Semiparametric Model		
1	0.032	0.346	1.732
	(0.028)	(0.078)	(0.509)
	Probit Model		
1	0.516	0.520	1.895
	(0.242)	(0.163)	(0.387)



ESTIMATE OF G(V)



CONCLUSIONS

- Single-index models:
  - Provide compromise between restrictions of parametric models and imprecision of fully nonparametric models
  - May be structural (e.g., random utility binary-response model)
- Asymptotic efficiency bounds available in some cases
- Two classes of estimators
  - Nonlinear optimization: provides asymptotically efficient estimator in some cases
  - Direct: Non-iterative, does not require solving nonlinear optimization problem
  - One-step estimation from direct-estimate yields asymptotic efficiency when efficient estimator available
- Example based on real data illustrates usefulness