Instrumental Variables Type Estimation for Nonignorable Selection Bias with a Continuous Instrument

Robert Moffitt
Johns Hopkins University

January, 2006

INTRODUCTION

- **Economists (mostly) assume the existence of nonignorable selection bias**

- **Economists (mostly) use instrumental variables (IV) or some variant of it to adjust for that bias**

- **Economists (mostly) test for existence of nonignorable selection bias by specification tests based on a comparison of treatment effect estimates with and without IV**

- **Economists (mostly) like to use the framework of regression analysis to analyze the problem**

- **(Some) economists are moving toward less parametric estimation of the model**

- **Economists (mostly) admit that there is probably heterogeneity in treatment impact but do nothing about it except appeal to the LATE motivation of Imbens-Angrist (1994)**

- **(A few) economists use continuous instruments and attempt to map out the distribution of responses over as much of the participation probability support as possible**

- **Parametric version: Björklund-Moffitt (1987)**

A FORMULATION OF THE HETEROGENEOUS TREATMENT EFFECT MODEL

- **Potential outcomes in regression framework:**

α_i = treatment effect for individual i

$= Y_{1i} - Y_{0i}$

where Y_{Di} = value of outcome if the individual participates (D_i=1) or does not (D_i=0)

- **Model:**

$y_i = \beta_i + \alpha_i D_i$ **an equivalent formulation**

$D_i^* = k(Z_i, \delta_i)$

$D_i = 1(D_i^* \geq 0)$

where Z_i is an exclusion restriction (potential instrument).

and where $\beta_i = Y_{0i}$ and $\alpha_i = Y_{1i} - Y_{0i}$

- **Ignoring for the moment what the object of estimation is, it looks as though there are two potential sources of bias from unobserved confounders:**

  β_i may not be independent of D_i

  α_i may not be independent of D_i

- **Indirectly, these are statements about independence of D_i from of Y_{1i} and Y_{0i}**

Hence

$$E(y_i \mid Z_i{=}z) \;=\; E(\beta_i \mid Z_i{=}z) \;+\; E(\alpha_i \mid D_i{=}1, Z_i{=}z)\ \text{Prob}(D_i{=}1 \mid Z_i{=}z)$$

$$E(D_i \mid Z_i{=}z) \;=\; \text{Prob}\,[k(z, \delta_i) \geq 0]$$

Make the following assumptions:

A1. $E(\beta_i \mid Z_i{=}z) = \beta$

A2. $E(\alpha_i \mid D_i{=}1, Z_i{=}z) = g[E(D_i{=}1 \mid Z_i{=}z)]$

A3. "Monotonicity" in effect of Z on D in D* equation

(for interpretation)

With these assumptions, we have

$$y_i = \beta + g[F(Z_i)]\, F(Z_i) + \epsilon_i$$

$$D_i = F(Z_i) + \nu_i$$

where F is a proper probability function and where $E(\epsilon_i \mid gF) = E(\nu_i \mid F) = 0$ **by construction**

- **g is the treatment on the treated**

  (mean of α_i for those observed to participate)

- **Average treatment effect is g(1)**

- **derivative of Y eqn. w.r.t. F is the marginal treatment effect (mean of α_i for those on the borderline of participation) (LATE is a discrete approximation to it)**

- **All functions (F and g) are nonparametrically identified at each Z point (g needs a normalization if the support of F in the data does not contain zero)**

- **Can use your favorite nonparametric method for g**

- **My preference: series estimation**

- **Can estimate the two equations with two-step methods or jointly with nonlinear least squares (my preference)**

- **Adding X: similar treatment gives:**

$$y_i = g[F(Z_i,X_i), X_i]\, F(Z_i,X_i) + h(X_i) + \epsilon_i$$

$$D_i = F(Z_i,X_i) + \nu_i$$

- **Three functions to estimate: g(F,X), h(X), and F(Z,X)**

Applications

- **Three from Public Health**

1. **Effect of hospital procedure (cathertization) on health outcomes (mortality). IV: distance to hospital.**

   (Newhouse, Annual Review of Public Health, 1998)

2. **Effect of smoking on health. IV: anti-smoking legislation.**

   (Pearl, 1995)

3. **Effect of teen drinking on y. IV: state beer tax**

   (W. Evans, 2003)

But all three have a continuous Z but estimate a single treatment effect

Table 1

Results of the Initial Estimation

	(1)	(2)	(3)	(4)	(5)
$\underline{\gamma}$					
Constant	-30.9* (4.0)	-77.6* (8.0)	-7.1 (21.5)	-41.4 (37.0)	6.38 (78.7)
F	--	43.0 (6.0)	4.5 (19.2)	67.0 (68.3)	-94.9 (255.4)
Max(0,F-.25F)	--	--	--	--	113.4 (197.6)
Max(0,F-.50F)	--	--	--	-30.6 (51.9)	-11.9 (53.0)
Max(0,F-.75F)	--	--	--	--	18.0 (28.2)
$\underline{\lambda}$					
Education	--	--	-3.16* (.91)	-1.56* (.94)	-1.86 (1.03)
Age	--	--	-1.91 (1.60)	2.41 (1.64)	2.42 (1.64)
Black	--	--	6.98* (3.28)	2.38 (3.48)	3.50 (3.72)
Family6	--	--	.74 (2.42)	-2.75 (2.63)	-1.86 (3.02)
Family	--	--	.98 (.92)	1.06 (1.12)	1.03 (.94)
Nonlabor	--	--	4.16* (1.62)	6.80* (1.93)	5.82* (2.26)
G			-.53 (.62)	-1.14* (.66)	-.98 (.69)

Table 1 (continued)

	(1)	(2)	(3)	(4)	(5)
<u>β</u>					
Constant	21.6* (3.2)	38.8* (5.2)	5.45 (6.89)	7.25 (7.53)	2.78 (10.4)
Education	.91* (.19)	.16* (.31)	2.60 (.43)	2.45* (.42)	2.60* (.48)
Age	1.68* (.31)	1.25* (.41)	1.20* (.53)	1.29* (.53)	1.30* (.54)
Black	-1.86* (.62)	-.14 (.84)	-3.93* (1.28)	-3.72* (1.31)	-4.23* (1.54)
Family6	.06 (.61)	.41 (.70)	-1.29 (1.14)	-1.61 (1.15)	-2.07 (1.41)
Family	-.71* (.19)	-.71* (.23)	-1.06* (.34)	-.98* (.34)	-.98* (.34)
Nonlabor	-.92* (.22)	-1.92* (.27)	-.96* (.31)	-1.02* (.44)	-.67 (.66)
G	-.27* (.14)	-.06 (.17)	-.18 (.25)	-.19 (.25)	-.27 (.27)
<u>δ</u>					
Constant	.62* (.19)	1.06* (.18)	.80* (.17)	.72* (.17)	71* (.17)
Education	-.11* (.01)	-.16* (.01)	-.12* (.01)	-.12* (.01)	-.12* (.01)
Age	-.06* (.03)	-.06* (.03)	-.06* (.02)	-.05* (.02)	-.05* (.02)
Black	.29* (.05)	.31* (.04)	.34* (.04)	.36* (.04)	.36* (.04)

Table 1 (continued)

	(1)	(2)	(3)	(4)	(5)
Family6	.41*	.34*	.38*	.38*	.39*
	(.03)	(.03)	(.02)	(.02)	(.02)
Family	-.04*	-.01	-.03*	-.02*	-.02
	(.02)	(.02)	(.01)	(.01)	(.01)
Nonlabor	-.25*	-.21*	-.20*	-.21*	-.21*
	(.02)	(.02)	(.02)	(.02)	(.02)
G	.05*	.05*	.06*	.07*	.07*
	(.01)	(.01)	(.01)	(.01)	(.02)
<u>η</u>					
Admin	.90	.22	-.41	-2.16*	-1.85*
	(1.19)	(.92)	(1.08)	(1.04)	(1.04)
Pctdenied	-.14	-.08	-.24*	-.31*	-.32*
	(.14)	(.10)	(.12)	(.11)	(.11)
Eligerror	-.03*	-.02*	-.02	-.01	-.01
	(.01)	(.01)	(.01)	(.01)	(.01)

Notes:
Standard errors in parentheses
*: significant at the 10 percent level

Table 5

Parameter Estimates of g Function for Spline Specifications

	Z splines (P eqn only)	X-Z splines	
		P eqn only	y and P eqns[a]
Constant	-27.5*	-20.7*	-20.4*
	(12.4)	(5.2)	(9.37)
F	25.6*	-.41	9.46*
	(8.7)	(4.03)	(5.70)
<u>λ</u>			
Education	-2.24*	--	--
	(.58)		
Education12	--	-7.60*	--
		(1.86)	
Age	.79	-2.29*	--
	(1.39)	(1.07)	
Black	4.59*	6.58*	--
	(2.29)	(1.64)	
Family6	-1.72	.06	--
	(1.42)	(.87)	
Family	1.69*	--	--
	(.81)		
Family>3	--	5.29*	--
		(1.57)	
Nonlabor	5.05*	2.08*	--
	(1.16)	(.66)	
G	-1.09*	-.14	--
	(.51)	(.39)	

Notes:
Standard errors in parentheses
*: significant at the 10% level
[a] Coefficients on X splines shown in the Appendix [available upon request]