Endogeneity & Accounting

part I

Doug Schroeder
Ohio State University

Purdue University workshop
October 18-19, 2007

Outline

I. Brief overview of accounting and econometric analysis of endogeneity

II. Some (highly-stylized) examples

III. Road map for remaining discussions

Endogeneity & accounting

Managers actively make production-investment, financing, and accounting choices.
These choices are intertwined and far from innocuous.
Design of accounting (like other information systems) is highly dependent on the implications and responses to accounting information in combination with other information.
As these decisions are interrelated, their analysis is inherently endogenous (Demski, 2004, AAA Presidential address).
Endogeneity presents substantial challenges for statistical/econometric analysis.

A brief overview of some history of (statistical-based) inquiries into endogenous relations in accounting, business, and economics

– *where have we been?*

– *where are we?*

– *where might we go from here?*

---

where have we been?

linear models and linear IV (Larcker and Rusticus provide an extensive review of accounting literature)

discrete choice and random utility model (RUM) – statistical analysis of choice

simultaneous equations

---

where have we been?

structural models and causal effects (Cowles commission)

Ragnar Frisch in inaugural editorial in *Econometrica* (1936) set the stage for the Cowles commission's view that economics focus on structural modeling by simply reminding us that "econometrics" stands for the unification of economic theory, statistics, and mathematics.

where are we?

much of where we've been and where we're at (perhaps even where we might go) is summarized in the 2000 Nobel lectures of James Heckman and Daniel McFadden
(both offer rich and deep insights into the development of econometric analysis of choice)

causal effects, the ceteris paribus response to a change in variable or parameter, remain the objective of much business & economic analysis; endogeneity often makes it infeasible to “turn one dial at a time”

where are we?

average treatment effects and counterfactuals

identification of treatment effects & ignorable treatment or “selection on observables”

identification of treatment effects & homogeneous response

identification of treatment effects & heterogeneous response

more later

where are we?

fixed effects analysis of panel data with endogenous effects

- Nikolaev and Van Lent study variation through time in a firm’s disclosure quality and its marginal cost of debt where unobservable cross firm heterogeneity (presumed largely constant through time) is accommodated via firm fixed effects

where are we?

nonparametric identification of treatment effects

where are we?

propensity score approaches (including propensity score matching; Rosenbaum and Rubin)
and propensity score IV approaches

- often functional form assumptions on the conditional means of the unobservables are employed to estimate treatment effects
- alternatively, the propensity for treatment (estimated by the probability of treatment conditional on the regressors, $p(X)$) can serve this role

where are we?

control function approaches
and control function IV approaches

- functions which control for endogeneity (selection) bias (assuming regularity conditions are satisfied)

where are we?

bivariate probit – two interdependent choices

simultaneous probit – a probit model with endogenous regressors

– Bagnoli, Liu, and Watts study debt contracts in which interest rates and inclusion of covenants are jointly determined

where are we?

continuous endogenous treatment

– relaxes the uniformity, monotonicity, or single index structure of choice assumption

strategic probit

– statistical analysis of strategic multi-agent games

where are we?

latent IV

general equilibrium analysis

where are we?

local IV (LIV) estimation of marginal treatment effects

– nonparametric estimation of treatment effects where impact of treatment is essentially heterogeneous

Bayesian data augmentation and predictive distributions of treatment effects

– posterior distributions simulated via McMC methods

more later

where we might go from here?

three cornerstones of data compression are

theory

data

diagnostic checking – discovery of the data generating process (DGP)

– all 3 require attention and creativity

where we might go from here?

theory frames the problem so that (economically) meaningful causal effects can be deduced

Heckman criticizes the selection literature for periods of preoccupation with devising estimators with nice statistical properties (especially asymptotic consistency) but little economic import

Heckman's more recent work attempts to position the treatment effects literature vis-a-vis the more ambitious structural modeling for policy evaluation of the Cowles commission (where theory is paramount)

where we might go from here?

data

Zvi Griliches ("Economic data issues," *Handbook of Econometrics*, volume IIIc, 1986)

"every econometric study is incomplete."

muses on the love/hate relationship with data

reminds us that the "quality of the data depends on both its source and use"

and that creativity is required to embrace the data issue

---

where we might go from here?

data (cont.)

at present, it seems that creativity in the address of
- omitted, correlated variables,
- heterogeneity in unobservables, and
- instruments

is in short supply in the accounting and business literature

self-critical

---

diagnostic checking (discovery of the DGP)

receives more attention in these notes
but has little to offer if the first two (theory and data) are not carefully and creatively attended

with our current understanding of econometrics it seems you can't say much about a potential issue – including endogeneity – if you don't accommodate it in the analysis

even then it is typically quite challenging to assess the nature and extent of the problem

---

where do we go from here?

in accounting & business we frequently borrow from other disciplines

endogeneity & selection - labor economics and biostatistics

strategic choice - economics and political science

---

where do we go from here?

strategic choice & selection - information problems perhaps offer a venue for accounting & business to give back to the academy

though in accounting & business we seem to be well behind

may need to take some chances, report failures (along with successes), and generally market our papers less intensely

---

why do we care?

perusal of Heckman's "Causal parameters and policy analysis in economics: A twentieth century retrospective" (*QJE* 2000) reminds us of the modest progress (frequent failures to adequately describe the data) enjoyed by structural modeling, and often evokes
"why bother studying endogenous relations?"

why do we care?

some modest examples

why do we care?

sample selection & Simpson's paradox example

Example:

Suppose a firm has two production facilities, *A* and *B*.

Facility *A* is perceived to be more efficient (higher proportion of non-defectives).

Consequently, production has historically been skewed in favor of facility *A*.

The firm is interested in improving production efficiency and particularly for facility *B*.

They have identified new production technology and are interested in whether the new technology improves production efficiency.

Production using the new technology is skewed toward facility *B*.

This "experiment" generates the following data:

	Facility *A*		Facility *B*		Total	
Technology	New	Old	New	Old	New	Old
Successes	10	120	133	25	143	145
Trials	10	150	190	50	200	200
% successes	100	80	70	50	71.5	72.5

Is the new technology more effective than the old technology?

What is the technology treatment effect?

Now, suppose an analyst collects the data but is unaware that there are two different facilities
(the analyst only has the last two columns of data).

What conclusion regarding the technology treatment effect is likely to be reached?

This illustration of *Simpson's paradox* is produced via a sample selection problem.

The data are not generated randomly but rather in pursuit of management's selective "experimentation" on production technology.

why do we care?

self-selection & treatment effects (& Simpson's paradox) examples

– to appreciate the examples need to define treatment effects

---

Treatment effects

Suppose treatment is binary
($D = 1$ for treatment, $D = 0$ for untreated).

The treatment effect with observable outcome
$y = Dy_1 + (1-D)y_0$ is
$\Delta = y_1 - y_0 = \mu_1 + u_1 - \mu_0 - u_0 = (\mu_1 - \mu_0) + (u_1 - u_0)$,
an individual's (potential) outcome response to a change in treatment from regime 0 to regime 1.

Note $(u_1 - u_0)$ is the individual-specific gain.

Treatment effects

while treatment effects focus on potential gains for an individual, the unobservable nature of counterfactuals leads to focus on population level parameters

---

ATE

The average treatment effect,

ATE = $E[\Delta] = E[y_1 - y_0]$,

is the average response to treatment for a random sample from the population.

Note: $E[y_1 - y_0] = E[y_1|D=1]Pr(D=1)+E[y_1|D=0]Pr(D=0)$

$-\{E[y_0|D=1]Pr(D=1)+E[y_0|D=0]Pr(D=0)\}$

ATT

The average treatment effect on the treated,

ATT = $E[\Delta|D = 1] = E[y_1 - y_0|D = 1]$,

is the average response to treatment for a sample of individuals that choose (or are assigned) treatment.

selection (treatment) is assumed to follow some RUM, $U_D = Z - V_D$ where U_D is latent utility index associated with treatment, V_D is the part unobserved by the analyst, and $D = 1$ if $U > 0$ or $D = 0$ otherwise.

ATUT

The average treatment effect on the untreated,

ATUT = $E[\Delta|D = 0] = E[y_1 - y_0|D = 0]$,

is the average response to treatment for a sample of individuals that choose (or are assigned) no treatment.

OLS

exogenous dummy variable regression estimates

$\text{OLS} = E[y_1 | D = 1] - E[y_0 | D = 0]$.

Bias in OLS estimate for ATT is

$\text{OLS} = \text{ATT} + \text{bias}_{TT}$

$E[y_1 | D = 1] - E[y_0 | D = 0] =$

$E[y_1 | D=1] - E[y_0 | D=1] + \{E[y_0 | D=1] - E[y_0 | D=0]\}$

Hence, $\text{bias}_{TT} = \{E[y_0 | D = 1] - E[y_0 | D = 0]\}$.

Bias in OLS estimate for ATUT is

$\text{OLS} = \text{ATUT} + \text{bias}_{TUT}$

$E[y_1 | D = 1] - E[y_0 | D = 0] =$

$E[y_1 | D = 0] - E[y_0 | D = 0] + \{E[y_1 | D = 1] - E[y_1 | D = 0]\}$

Hence, $\text{bias}_{TUT} = \{E[y_1 | D = 1] - E[y_1 | D = 0]\}$.

Since $\text{ATE} = E[y_1 | D=1]\Pr(D=1) + E[y_1 | D=0]Pr(D=0)$

$- \{E[y_0 | D=1]Pr(D=1) + E[y_0 | D=0]Pr(D=0)$

$= Pr(D = 1)\ \text{ATT} + Pr(D = 0)\ \text{ATUT}$

$\text{ATE} = p\ \text{ATT} + (1-p)\ \text{ATUT}$

$E[y_1 - y_0] = pE[y_1 - y_0 | D = 1] + (1-p)E[y_1 - y_0 | D = 0]$

bias in OLS estimate for ATE is

$\text{bias}_{ATE} = Pr(D = 1)\ \text{bias}_{TT} + Pr(D = 0)\ \text{bias}_{TUT}$

Some stylized examples

Case 1:

State	*one*		*two*		*three*	
Pr(state)	0.04		0.32		0.64	
D	0	1	0	1	0	1
y_0	0	0	1	1	2	2
y_1	1	1	1	1	1	1
$Pr(y,D \mid \text{state})$	0.68	0.32	1.0	0.0	0.92	0.08
$Pr(y,D,\text{state})$	0.0272	0.0128	0.32	0.0	0.5888	0.0512

Case 1

Results:	Key components:
ATE = $E[y_1 - y_0] = -0.6$	$E[y_1 \mid D=1] = 1.0$
ATT = $E[y_1 - y_0 \mid D = 1] = -0.6$	$E[y_1 \mid D=0] = 1.0$
ATUT = $E[y_1 - y_0 \mid D = 0] = -0.6$	$E[y_1] = 1.0$
OLS = $E[y_1 \mid D = 1] - E[y_0 \mid D = 0] = -0.6$	$E[y_0 \mid D=1] = 1.6$
$\text{bias}_{\text{ATT}} = E[y_0 \mid D = 1] - E[y_0 \mid D = 0] = 0.0$	$E[y_0 \mid D=0] = 1.6$
$\text{bias}_{\text{ATUT}} = E[y_1 \mid D = 1] - E[y_1 \mid D = 0] = 0.0$	$E[y_0] = 1.6$
$\text{bias}_{\text{ATE}} = p\ \text{bias}_{\text{ATT}} + (1-p)\ \text{bias}_{\text{ATUT}} = 0.0$	$p = Pr(D=1) = 0.064$

Case 2:

State	*one*		*two*		*three*	
Pr(state)	0.04		0.32		0.64	
D	0	1	0	1	0	1
y_0	0	0	1	1	2	2
y_1	1	1	1	1	1	1
$Pr(y,D \mid \text{state})$	0.68	0.32	0.70	0.30	0.92	0.08
$Pr(y,D,\text{state})$	0.0272	0.0128	0.224	0.096	0.5888	0.0512

Case 2

Results:	Key components:
ATE = $E[y_1 - y_0] = -0.6$	$E[y_1 \mid D=1] = 1.0$
ATT = $E[y_1 - y_0 \mid D = 1] = -0.24$	$E[y_1 \mid D=0] = 1.0$
ATUT = $E[y_1 - y_0 \mid D = 0] = -0.669$	$E[y_1] = 1.0$
OLS = $E[y_1 \mid D = 1] - E[y_0 \mid D = 0] = -0.669$	$E[y_0 \mid D=1] = 1.24$
$\text{bias}_{\text{ATT}} = E[y_0 \mid D = 1] - E[y_0 \mid D = 0] = -.429$	$E[y_0 \mid D=0] = 1.669$
$\text{bias}_{\text{ATUT}} = E[y_1 \mid D = 1] - E[y_1 \mid D = 0] = 0.0$	$E[y_0] = 1.6$
$\text{bias}_{\text{ATE}} = p\ \text{bias}_{\text{ATT}} + (1-p)\ \text{bias}_{\text{ATUT}} = -.069$	$p = Pr(D=1) = 0.16$

Case 3:

State	*one*		*two*		*three*	
Pr(state)	0.04		0.32		0.64	
D	0	1	0	1	0	1
y_0	0	0	1	1	2	2
y_1	1	1	1	1	0	0
$Pr(y,D \mid \text{state})$	0.68	0.32	0.70	0.30	0.92	0.08
$Pr(y,D,\text{state})$	0.0272	0.0128	0.224	0.096	0.5888	0.0512

Case 3

Results:	Key components:
ATE = $E[y_1 - y_0] = -1.24$	$E[y_1 \mid D=1] = 0.68$
ATT = $E[y_1 - y_0 \mid D = 1] = -0.56$	$E[y_1 \mid D=0] = 0.299$
ATUT = $E[y_1 - y_0 \mid D = 0] = -1.370$	$E[y_1] = 0.36$
OLS = $E[y_1 \mid D = 1] - E[y_0 \mid D = 0] = -0.989$	$E[y_0 \mid D=1] = 1.24$
$\text{bias}_{\text{ATT}} = E[y_0 \mid D = 1] - E[y_0 \mid D = 0] = -.429$	$E[y_0 \mid D=0] = 1.669$
$\text{bias}_{\text{ATUT}} = E[y_1 \mid D = 1] - E[y_1 \mid D = 0] = 0.381$	$E[y_0] = 1.6$
$\text{bias}_{\text{ATE}} = p\ \text{bias}_{\text{ATT}} + (1-p)\ \text{bias}_{\text{ATUT}} = 0.251$	$p = Pr(D=1) = 0.16$

Case 4:

State	*one*		*two*		*three*	
Pr(state)	0.04		0.32		0.64	
D	0	1	0	1	0	1
y_0	3.34	3.34	-5.8	-5.8	5.52	5.52
y_1	-0.15	-0.15	1	1	2.15	2.15
$Pr(y,D \mid \text{state})$	0.68	0.32	1.0	0.0	0.92	0.08
$Pr(y,D,\text{state})$	0.0272	0.0128	0.32	0.0	0.5888	0.0512

Case 4

Results:	Key components:
ATE = $E[y_1 - y_0] = -0.1204$	$E[y_1 \mid D=1] = 1.69$
ATT = $E[y_1 - y_0 \mid D = 1] = -3.394$	$E[y_1 \mid D=0] = 1.69$
ATUT = $E[y_1 - y_0 \mid D = 0] = 0.1034$	$E[y_1] = 1.69$
OLS = $E[y_1 \mid D = 1] - E[y_0 \mid D = 0] = 0.1034$	$E[y_0 \mid D=1] = 5.084$
$\text{bias}_{ATT} = E[y_0 \mid D = 1] - E[y_0 \mid D = 0] = 3.497$	$E[y_0 \mid D=0] = 1.5866$
$\text{bias}_{ATUT} = E[y_1 \mid D = 1] - E[y_1 \mid D = 0] = 0.0$	$E[y_0] = 1.8104$
$\text{bias}_{ATE} = p\ \text{bias}_{ATT} + (1-p)\ \text{bias}_{ATUT} = 0.224$	$p = Pr(D=1) = 0.064$

Case 4

endogeneity bias can produce a Simpson's paradox result!

while these examples are not as rich and deep as Lucas' critique of econometric policy evaluation, the message is similar – *endogeneity matters!*

concluding (first day) remarks

again thanks for this opportunity to try to gain a foothold on a difficult (if you will, humbling) but important issue

first discussion tomorrow

will focus on more details regarding identification and classical estimation of treatment effects with an emphasis on essential heterogeneity and Heckman's local IV estimator (LIV) of marginal treatment effects (MTE)

second talk tomorrow

will focus on

– Bayesian analysis of self-selection

– data augmentation (to deal with latent variables and counterfactuals),

– deriving marginal posterior distributions via Markov chain Monte Carlo (McMC) methods

– predictive distributions of treatment effects