Traditional Instrumental Variables and Likelihood (Bayesian) Approaches for Comparing Antipsychotic Medications

A. James O'Malley
Harvard Medical School

Co-authors: Sharon-Lise T. Normand and Richard G. Frank
(NIH-NIMH grants MH061434 and MH069721)

Offset of Mental Health Costs

- **Atypical antipsychotics, a new class of mental health drugs, are more expensive than tradition treatments but are likely to improve tolerability $\Rightarrow$ better patient compliance $\Rightarrow$ lower subsequent medical costs.**
- **Hypothesis: Atypical antipsychotics reduce the net cost of mental health care compared to conventional antipsychotics.**
- **Data Source: Continuously-enrolled (6+ months) Florida Medicaid population, July 1994 - June 2001.**

Predictor variables

- **Treatment: atypical (Zyprexa, Seroquel, Geodon) vs. conventional**
- **Gender: male vs. female**
- **Ethnicity: white, black vs. other**
- **Substance abuse**
- **Area of residence (11 regions of Florida); Miami most populous**
- **Recipient of supplementary security income (SSI)**
- **Year**

Need for Instrumental Variables

- **Endogeneous variables/selection effects:**
  - **Existing medical and mental health comorbidities**
  - **Severity of illness**
  - **Patient preferences**
- **Natural variation in availability of atypicals across Florida from physician prescribing habits.**
  - **Availability related to treatment but not directly related to cost.**
  - **Instrumental variables: drug release indicators and interactions with patient residence.**

Notation

- y_i = log(total spending)
- z_i = use of an atypical
- x_i = vector of exogenous predictors
- u_i = vector of instrumental variables
- c_i = unmeasured confounding variable

Instrumental Variable (IV) Assumptions

- u_i associated with z_i: $\text{cor}(z_i, u_i) \neq 0$.
- u_i is independent of c_i.
- u_i is independent of y_i given z_i and c_i.
- Note: associations suffice with continuous data.

Binary Treatment Variable z_i

- **Traditional IV approach consistent (Battacharya et al., 2006).**
- **We compare:**
  - **Ordinary least squares (OLS)**
  - **Two-stage (linear and nonlinear) least squares**
  - **Two-stage residual inclusion**
  - **Bayesian parametric modeling**

when assumptions hold and under violations.

Ordinary Least Squares

- **Ordinary least squares (OLS) regression model:**

$$y_i = z_i\beta_1 + x_i^T\beta_2 + \epsilon_{y,i}.$$

- **For inference it is often assumed that**

$$\epsilon_{y,i} \sim N(0, \sigma_y^2),$$

where $N(0, \sigma_y^2)$ is the normal density function with mean 0 and variance σ_y^2.

- **Require $\text{cor}(z_i, \epsilon_{y,i}) = 0$ for unbiased inferences.**

Two-stage Least Squares

- **Linear first stage model:**

$$\boldsymbol{z_i = u_i^T \theta_1 + x_i^T \theta_2 + \epsilon_{z,i}} \tag{1}$$

- **Compute predicted values of $\boldsymbol{z_i}$, denoted $\boldsymbol{\hat{z}_i}$.**
- **Fit second-stage (outcome) model:**

$$\boldsymbol{y_i = \hat{z}_i \beta_1 + x_i^T \beta_2 + \epsilon_{y,i}}$$

- **Alternative: Nonlinear first-stage model (2SPS):**

$$\boldsymbol{\Phi^{-1}(z_i) = u_i^T \theta_1 + x_i^T \theta_2} \tag{2}$$

- **Interpretation of $\boldsymbol{\beta_1}$ same under (1) and (2).**

Two-stage Residual Inclusion (2SRI)

- **Hausman (1978); Terza et al (2008)**
- **Nonlinear first stage model:**

$$\Phi^{-1}(z_i) = u_i^T \theta_1 + x_i^T \theta_2$$

- **Estimate c_i from "bad" variation in z_i:**

$$\hat{c}_i = \Phi(u_i^T \hat{\theta}_1 + x_i^T \hat{\theta}_2) - z_i$$

- **Augmented outcome model:**

$$y_i = z_i \beta_1 + x_i^T \beta_2 + \hat{c}_i \beta_3 + \epsilon_{y,i}$$

Structural Model

- **Fit simultaneous system of equations**

$$y_i = z_i\beta_1 + x_i^T\beta_2 + \epsilon_{y,i}$$

$$z_i^* = u_i^T\theta_1 + x_i^T\theta_2 + \epsilon_{z,i}$$

$$\epsilon_i = \begin{pmatrix} \epsilon_{y,i} \\ \epsilon_{z,i} \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_z \\ \rho\sigma_y\sigma_z & \sigma_z^2 \end{pmatrix}\right\},$$

where $z_i = I(z_i^* \geq 0)$.

Structural Model (Cont.)

- z_i^*, propensity to prescribe treatment, unobserved.
- For model identifiability set $\sigma_z^2 = 1$.
- Special case of Heckit model (Heckman, 1978).

Bayesian Analysis

- **Likelihood function is proportional to:**

$$L = \prod_{i=1}^{n} \phi(y_i; \mu_{y,i}, \sigma_y^2)\Phi(\mu_{y|z,i})^{z_i}(1 - \Phi(\mu_{y|z,i}))^{1-z_i},$$

where

$$\mu_{y,i} = z_i\beta_1 + x_i^T\beta_2,$$

$$\mu_{y|z,i} = \frac{u_i^T\theta_1 + x_i^T\theta_2 + \rho(y_i - \mu_{y,i})/\sigma_y^2}{(1-\rho^2)^{1/2}}.$$

- **Diffuse prior:**

$$p(\beta, \theta, \sigma_y^2, \rho) \propto I(\sigma_y^2 \geq 0, \rho \in [-1, 1])/\sigma_y^2.$$

Special Cases of Bayesian Model

- If $\rho = 0$ obtain separated linear and probit regression models.
- Posterior mode under alternative (flat) prior

$$p(\beta, \theta, \sigma_y^2, \rho) \propto I(\sigma_y^2 \geq 0, \rho \in [-1, 1])$$

corresponds to maximum likelihood estimator (MLE).

- Use standard optimization methods to obtain posterior mode.

Quantities Estimated

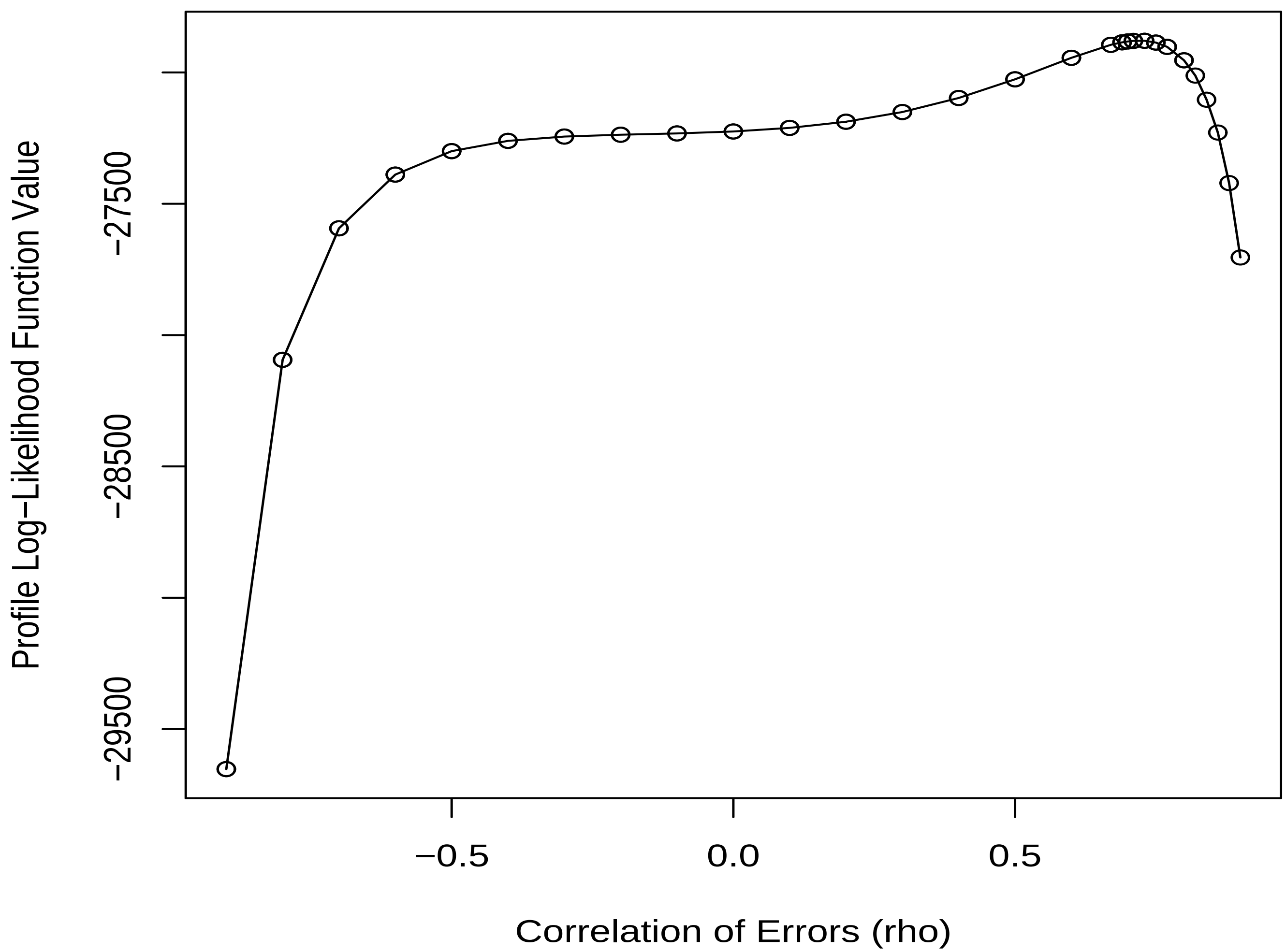
- **Conditional effect (error term held fixed) - β_1.**
- **Average treatment effect (ATE) - marginal effect:**

$$\text{ATE} = \beta_1 + \frac{\rho\sigma_y}{n}\sum_{i=1}^{n}\frac{\phi(u_i^T\theta_1 + x_i^T\theta_2)}{\Phi(u_i^T\theta_1 + x_i^T\theta_2)(1-\Phi(u_i^T\theta_1 + x_i^T\theta_2))},$$

$\phi(.)$ and $\Phi(.)$ are the PDF and CDF of $N(0,1)$.

- **Involves inverse Mills-ratio; relies on normal assumption.**
- **Effects are in terms of log-spending.**

Profile Likelihood Function of ρ



Inferences for Florida Medicaid Data

		Moment-Based				Bayesian posterior	
Term	**Quantity**	**OLS**	**2SLS**	**2SPS**	**2SRI**	**Mode**	**Mean**
β_1	**Estimate**	**1.022**	**-0.028**	**0.237**	**0.193**	**-0.793**	**-0.721**
	Standard Deviation	**0.010**	**0.169**	**0.144**	**0.145**	**0.031**	**0.035**
ATE	**Estimate**					**1.049**	**1.044**
	Standard Deviation					**0.010**	**0.010**
ρ	**Estimate**					**0.722**	**0.702**
	Standard Deviation					**0.003**	**0.010**

Note: The marginal variance in the Bayesian model was 2.314 (0.029).

Summary: Analysis of Florida Data

- **Bayesian (likelihood) methods more precise than 2SLS!**
- **2SRI more precise than 2SLS methods (as in Terza et al, 2008).**
- **Posterior mode (MLE) $\hat{\rho} = 0.722$ and posterior mean $E[\rho \mid Y, Z, X, U] = 0.702$ are large!**

Simulation Experiment

- **Univariate x_i and u_i equated to combined effects in Florida Medicaid; set $\sigma_y^2 = 1$.**
- **Generative model:**

$$y_i = z_i\beta_1 + x_i\beta_2 + u_i\nu_3 + \epsilon_{y,i}$$

$$z_i^* = u_i\theta_1 + x_i\theta_2 + \epsilon_{z,i}, \text{ where } z_i = I(z_i^* \geq 0) \text{ and}$$

$$\epsilon_i = \begin{pmatrix} \epsilon_{y,i} \\ \epsilon_{z,i} \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}.$$

- **Key parameters: β_1, θ_1, ρ, and ν.**

Simulation Results: Normal Distribution

Simulation 1: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	-0.793
Bias	0.003	-0.011	0.012	0.012	0.004	0.003
MSE	0.002	0.270	0.042	0.042	0.035	0.003
Coverage	0.970	0.950	0.944	0.942	0.934	0.960

Simulation 2: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0.722$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	1.349
Bias	1.812	-0.031	0.003	0.003	0.001	0.005
MSE	3.287	0.272	0.046	0.045	0.006	0.002
Coverage	0.000	0.958	0.932	0.906	0.962	0.944

Note: Mode-Q denotes the posterior mode of Q from the Bayesian analysis with the joint flat prior.

Simulation Results: T Distribution (7 dof)

Simulation 1: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	-0.793
Bias	0.001	0.027	0.007	0.013	0.009	-0.002
MSE	0.002	0.159	0.014	0.014	0.014	0.003
Coverage	0.946	0.958	0.960	0.962	0.956	0.944

Simulation 2: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0.722$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	1.349
Bias	1.469	-0.025	-0.012	-0.016	-0.033	-0.224
MSE	2.160	0.180	0.021	0.021	0.012	0.052
Coverage	0.000	0.950	0.934	0.888	0.812	0.000

Note: Mode-Q denotes the posterior mode of Q from the Bayesian analysis with the joint flat prior.

Simulation Results: Gamma Distribution

Simulation 1: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	-0.793
Bias	-0.002	-0.005	-0.078	0.009	0.607	-0.066
MSE	0.002	0.359	0.045	0.038	1.308	0.018
Coverage	0.958	0.970	0.914	0.958	0.650	0.654

Simulation 2: Relative $\beta = -0.793$, $\theta = 0.144$, and $\rho = 0.722$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	1.349
Bias	1.990	-0.030	-0.059	-0.174	1.382	-0.056
MSE	3.965	0.343	0.047	0.074	1.916	0.006
Coverage	0.000	0.966	0.942	0.794	0.000	0.662

Note: Mode-Q denotes the posterior mode of Q from the Bayesian analysis with the joint flat prior.

Simulation Results: Exclusion Restriction

Simulation 1: Relative $\beta = -0.793$, $\theta = 0.144$, $\rho = 0.722$, and $\nu = 0$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	-0.793
Bias	1.809	-0.040	0.017	0.017	0.000	0.002
MSE	3.274	0.277	0.041	0.041	0.007	0.002
Coverage	0.000	0.952	0.946	0.922	0.934	0.942

Simulation 2: Relative $\beta = -0.793$, $\theta = 0.144$, $\rho = 0.722$, and $\nu = 0.144$

Criterion	Estimation Method					
	OLS	2SLS	2SNLS	2SRIS	Mode-β_1	Mode-avtrt
Truth	-0.793	-0.793	-0.793	-0.793	-0.793	1.349
Bias	1.841	4.192	0.647	0.647	0.062	0.015
MSE	3.389	17.885	0.465	0.465	0.012	0.002
Coverage	0.000	0.000	0.122	0.096	0.874	0.946

Note: Mode-Q denotes the posterior mode of Q from the Bayesian analysis with the joint flat prior.

Conclusion

- **Parametric model allows both β_1 (conditional effect) and ATE (marginal effect) of treatment to be estimated.**
  - **In general OLS estimates the ATE.**
  - **IV methods estimate β_1.**
- **Bayesian analysis relatively better when:**
  - **Instrument is weak**
  - **Residual correlation (effect of unmeasured confounder) is large**

Conclusion (Cont.)

- **Under departures from normality, Bayesian estimators or MLEs of β_1:**
  - **performed reasonably if distribution symmetric and $\rho < 0.4$**
  - **performed poorly if distribution asymmetric**
- **Results for the ATE are reversed.**
- **2SRI performs worse than 2SPS and substantially worse than 2SLS when the data are asymmetric.**
- **Bayesian (likelihood)-based estimators more robust to violations of exclusion restrictions.**

Policy Implications

- **Linear two-stage least squares safest option?**
- **Lack of evidence to suggest offset of atypicals is positive or negative.**
  - **Consistent with results from Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study (Lieberman, et al 2005).**
- **Incongruence between statistical procedures is alarming!**

Future work

- **Account for repeated measurement of physicians.**
  - **Tentative results: $E[\beta_1 \mid Y, Z, X, U] = -0.492$ (posterior standard deviation $= 0.0197$)**
- **Aid policy-making by evaluating effect of treatment in terms of total spending.**
- **Incorporate historical information on each patient's switching behavior.**
- **Control for innovativeness of each physician using proportion of novel drugs of any drug class they prescribe to their Medicaid patients.**

References

- **Bhattacharya, J., Goldman, D., and McCaffrey, D. (2006). Estimating Probit Models with Self-selected Treatments,** *Statistics in Medicine*, **25, 389–413.**
- **Hausman, J. A. (1978). Specification Tests in Econometrics,** *Econometrica*, **46, 1251–1271.**
- **Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System,** *Econometrica*, **46, 931–960.**
- **Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D.**

O., Keefe, R. S., Davis, S. M., Davis, C. E., Lebowitz, B. D., Severe, J. and Hsiao, J. K. (2005). Effectiveness of Antipsychotic Drugs in Patients with Schizophrenia, *New England Journal of Medicine*, **353, 1209–1223.**

- **Terza, J. V., Basu, A. and Rathouz, P. (2008). Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling,** *Journal of Health Economics*, **27, 531-543.**