

Mitigating the Impact of Endogeneity in Mental Healthcare Data via Multilevel Models

Paul H. Johnson, Jr.

Ph.D. Candidate
University of Wisconsin – Madison
School of Business

Research funded by the National Institute of Mental Health:
PHS Grant Number: 2 T32 MH18029-22



Purpose

- To examine whether racial disparities occur in the utilization of inpatient mental health services
- To introduce the health services literature to advanced multilevel modeling techniques

Racial Disparities in Mental Health

- Racial disparities: Differences in healthcare treatments and outcomes by race after considering **all** other individual and organizational factors
- The National Research Council (1997) and Institute of Medicine (2002) have found extensive evidence of racial disparities in healthcare treatments and outcomes
- Racial disparities have been attributed to (unobserved):
 - Socio-economic status
 - Health insurance coverage
 - Patient preferences / cultural beliefs
 - Physician bias / discrimination
 - Quality of the local healthcare system



Outline

- 1 Background
 - Multilevel Models
 - Endogeneity
 - Data
- 2 Multilevel Model Endogeneity Analysis
 - Two Level Model
 - Estimation Strategies
 - Empirical Results
- 3 Summary



Outline

- 1 Background
 - Multilevel Models
 - Endogeneity
 - Data
- 2 Multilevel Model Endogeneity Analysis
 - Two Level Model
 - Estimation Strategies
 - Empirical Results
- 3 Summary



Outline

- 1 Background
 - Multilevel Models
 - Endogeneity
 - Data
- 2 Multilevel Model Endogeneity Analysis
 - Two Level Model
 - Estimation Strategies
 - Empirical Results
- 3 Summary



Multilevel Models

- Modeling technique originally used in educational research, now used in other fields (Raudenbush & Bryk 2002, Goldstein 2003)
- Can be used when data has an inherent hierarchical structure, such as students within schools or patients within hospitals
- Advantages of multilevel models include:
 - More precise estimates of fixed effects (regression coefficients)
 - Provide appropriate standard errors for confidence intervals and hypothesis tests



Endogeneity

- Correlation of observed model variables with model random errors (Wooldridge 2002)
- Conceive of endogeneity as omitted variables
- Omitted variables can bias fixed effects estimates of observed model variables and lead to incorrect inferences



Data

- Healthcare Cost and Utilization Project's Nationwide Inpatient Sample (NIS)
 - Sponsored by the Agency for Healthcare Research and Quality
- Area Resource File (ARF)
 - Sponsored by the Bureau of Health Professions
- Sample: 97,378 adults (age 18 to 64) admitted to a hospital in 2003 and discharged with a mental illness, from 331 hospitals and 231 counties
- Goal: use data to determine whether there are racial disparities in inpatient mental healthcare utilization, measured by hospital total charges (TOTCHG)

Variables and Descriptions

Discharge-Level

Hospital total charges

Age at admission

Gender of patient

Race of patient

Primary expected payer

APR-DRG code

Risk of mortality subclass

Severity subclass

Hospital-Level

Bed size of hospital

Ownership of hospital

Location of hospital

Teaching status of hospital

County-Level

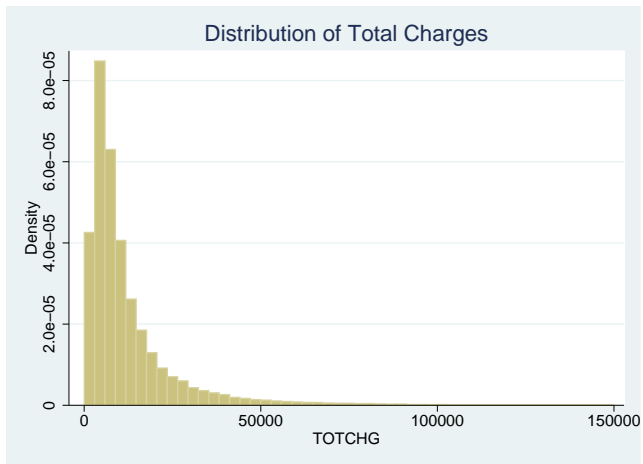
HPSA primary care physician code

HPSA mental health professional code

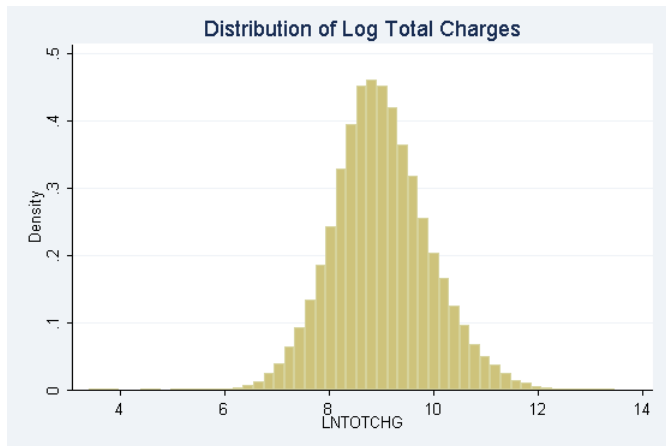
Total hospital admission rate

Per capita income

Distribution of TOTCHG



Distribution of LN(TOTCHG)



Two Level Model: Discharges within Hospitals

- Discharge Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)} \beta_1 + \beta_h^{(1)} + \epsilon_{hd}^{(1)}$$

- Hospital Model:

$$\beta_h^{(1)} = \mathbf{X}_h^{(2)} \beta_2 + \epsilon_h^{(2)}$$

- Combined Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)} \beta_1 + \mathbf{X}_h^{(2)} \beta_2 + \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$$

- Let $\mathbf{X}_{hd} = (\mathbf{X}_{hd}^{(1)} : \mathbf{X}_h^{(2)})$, $\beta = (\beta_1' : \beta_2')'$, $\delta_{hd} = \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$, $\mathbf{V}_h = \text{var}(\delta_h)$
Fully specify the two level model as:

$$\mathbf{V}_h^{-1/2} \mathbf{y}_h = \mathbf{V}_h^{-1/2} \mathbf{X}_h \beta + \mathbf{V}_h^{-1/2} \delta_h$$



Two Level Model: Discharges within Hospitals

- Discharge Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)}\beta_1 + \beta_h^{(1)} + \epsilon_{hd}^{(1)}$$

- Hospital Model:

$$\beta_h^{(1)} = \mathbf{X}_h^{(2)}\beta_2 + \epsilon_h^{(2)}$$

- Combined Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)}\beta_1 + \mathbf{X}_h^{(2)}\beta_2 + \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$$

- Let $\mathbf{X}_{hd} = (\mathbf{X}_{hd}^{(1)} : \mathbf{X}_h^{(2)})$, $\beta = (\beta_1' : \beta_2')'$, $\delta_{hd} = \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$, $\mathbf{V}_h = \text{var}(\delta_h)$
Fully specify the two level model as:

$$\mathbf{V}_h^{-1/2} \mathbf{y}_h = \mathbf{V}_h^{-1/2} \mathbf{X}_h \beta + \mathbf{V}_h^{-1/2} \delta_h$$



Two Level Model: Discharges within Hospitals

- Discharge Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)} \beta_1 + \beta_h^{(1)} + \epsilon_{hd}^{(1)}$$

- Hospital Model:

$$\beta_h^{(1)} = \mathbf{X}_h^{(2)} \beta_2 + \epsilon_h^{(2)}$$

- Combined Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)} \beta_1 + \mathbf{X}_h^{(2)} \beta_2 + \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$$

- Let $\mathbf{X}_{hd} = (\mathbf{X}_{hd}^{(1)} : \mathbf{X}_h^{(2)})$, $\beta = (\beta_1' : \beta_2')'$, $\delta_{hd} = \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$, $\mathbf{V}_h = \text{var}(\delta_h)$
Fully specify the two level model as:

$$\mathbf{V}_h^{-1/2} \mathbf{y}_h = \mathbf{V}_h^{-1/2} \mathbf{X}_h \beta + \mathbf{V}_h^{-1/2} \delta_h$$



Two Level Model: Discharges within Hospitals

- Discharge Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)}\beta_1 + \beta_h^{(1)} + \epsilon_{hd}^{(1)}$$

- Hospital Model:

$$\beta_h^{(1)} = \mathbf{X}_h^{(2)}\beta_2 + \epsilon_h^{(2)}$$

- Combined Model:

$$y_{hd} = \mathbf{X}_{hd}^{(1)}\beta_1 + \mathbf{X}_h^{(2)}\beta_2 + \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$$

- Let $\mathbf{X}_{hd} = (\mathbf{X}_{hd}^{(1)} : \mathbf{X}_h^{(2)})$, $\beta = (\beta_1' : \beta_2')'$, $\delta_{hd} = \epsilon_{hd}^{(1)} + \epsilon_h^{(2)}$, $\mathbf{V}_h = \text{var}(\delta_h)$
Fully specify the two level model as:

$$\mathbf{V}_h^{-1/2}\mathbf{y}_h = \mathbf{V}_h^{-1/2}\mathbf{X}_h\beta + \mathbf{V}_h^{-1/2}\delta_h$$



Estimation with No Omitted Variables

- Typically, analysts estimate fixed effects β using the ordinary least squares (OLS) estimator ($\mathbf{V}_h = \sigma^2 \mathbf{I}_h$):

$$\mathbf{b}_{\text{OLS}} = \left(\sum_h \mathbf{X}'_h \mathbf{X}_h \right)^{-1} \sum_h \mathbf{X}'_h \mathbf{y}_h$$

- However, the most efficient unbiased estimator of β accounts for the hierarchical structure of the data, the generalized least squares (GLS) estimator:

$$\mathbf{b}_{\text{GLS}} = \left(\sum_h \mathbf{X}'_h \mathbf{V}_h^{-1} \mathbf{X}_h \right)^{-1} \sum_h \mathbf{X}'_h \mathbf{V}_h^{-1} \mathbf{y}_h$$



Estimation with No Omitted Variables

- Typically, analysts estimate fixed effects β using the ordinary least squares (OLS) estimator ($\mathbf{V}_h = \sigma^2 \mathbf{I}_h$):

$$\mathbf{b}_{\text{OLS}} = \left(\sum_h \mathbf{X}'_h \mathbf{X}_h \right)^{-1} \sum_h \mathbf{X}'_h \mathbf{y}_h$$

- However, the most efficient unbiased estimator of β accounts for the hierarchical structure of the data, the generalized least squares (GLS) estimator:

$$\mathbf{b}_{\text{GLS}} = \left(\sum_h \mathbf{X}'_h \mathbf{V}_h^{-1} \mathbf{X}_h \right)^{-1} \sum_h \mathbf{X}'_h \mathbf{V}_h^{-1} \mathbf{y}_h$$



Introducing Omitted Variables

- At hospital-level, omitted variables: hospital capacity, physician practice patterns, healthcare demand and supply, community-level socio-economic status, and the quality of the healthcare system
 - Instead of $\epsilon_h^{(2)}$, model contains $\epsilon_h^{(2)*} + \mathbf{u}_h^{(2)}$
- At discharge-level, race-related omitted variables: socio-economic status, health insurance, patient preferences, and physician bias
 - Instead of $WHITE_{hd}$, model contains $WHITE_{hd}^* + \mathbf{u}_h^{(1WHITE)}$
 - $\mathbf{u}_h^{(1j)}$ = omitted effects for race j and hospital h



Introducing Omitted Variables

- At hospital-level, omitted variables: hospital capacity, physician practice patterns, healthcare demand and supply, community-level socio-economic status, and the quality of the healthcare system
 - Instead of $\epsilon_h^{(2)}$, model contains $\epsilon_h^{(2)*} + \mathbf{u}_h^{(2)}$
- At discharge-level, race-related omitted variables: socio-economic status, health insurance, patient preferences, and physician bias
 - Instead of $WHITE_{hd}$, model contains $WHITE_{hd}^* + \mathbf{u}_h^{(1WHITE)}$
 - $\mathbf{u}_h^{(1j)}$ = omitted effects for race j and hospital h



Introducing Omitted Variables

- At hospital-level, omitted variables: hospital capacity, physician practice patterns, healthcare demand and supply, community-level socio-economic status, and the quality of the healthcare system
 - Instead of $\epsilon_h^{(2)}$, model contains $\epsilon_h^{(2)*} + \mathbf{u}_h^{(2)}$
- At discharge-level, race-related omitted variables: socio-economic status, health insurance, patient preferences, and physician bias
 - Instead of $WHITE_{hd}$, model contains $WHITE_{hd}^* + \mathbf{u}_h^{(1WHITE)}$
 - $\mathbf{u}_h^{(1j)}$ = omitted effects for race j and hospital h



Introducing Omitted Variables

- At hospital-level, omitted variables: hospital capacity, physician practice patterns, healthcare demand and supply, community-level socio-economic status, and the quality of the healthcare system
 - Instead of $\epsilon_h^{(2)}$, model contains $\epsilon_h^{(2)*} + \mathbf{u}_h^{(2)}$
- At discharge-level, race-related omitted variables: socio-economic status, health insurance, patient preferences, and physician bias
 - Instead of $WHITE_{hd}$, model contains $WHITE_{hd}^* + \mathbf{u}_h^{(1WHITE)}$
 - $\mathbf{u}_h^{(1j)}$ = omitted effects for race j and hospital h



Introducing Omitted Variables

- At hospital-level, omitted variables: hospital capacity, physician practice patterns, healthcare demand and supply, community-level socio-economic status, and the quality of the healthcare system
 - Instead of $\epsilon_h^{(2)}$, model contains $\epsilon_h^{(2)*} + \mathbf{u}_h^{(2)}$
- At discharge-level, race-related omitted variables: socio-economic status, health insurance, patient preferences, and physician bias
 - Instead of $WHITE_{hd}$, model contains $WHITE_{hd}^* + \mathbf{u}_h^{(1WHITE)}$
 - $\mathbf{u}_h^{(1j)}$ = omitted effects for race j and hospital h



Estimation with Omitted Variables

- If only $\mathbf{u}_h^{(2)}$, can use *fixed effects estimation* (Kim & Frees 2006) that removes both observed and omitted hospital variables to estimate discharge fixed effects β_1 :

$$\mathbf{b}_{1FE} = \left(\sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1)} \right) - \sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- Again, if only $\mathbf{u}_h^{(2)}$, can obtain unbiased estimates of **all** fixed effects β with *instrumental variables (IV) estimation* (Kim & Frees 2006):

$$\mathbf{b}_{IV} = \left(\sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{X}_h \right) - \sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- If both $\mathbf{u}_h^{(2)}$ and $\mathbf{u}_h^{(1j)}$, can only obtain estimates of $\beta_1^{(1NR)}$, the discharge variables other than race: $\mathbf{b}_{1FE*}^{(1NR)} =$

$$= \left(\sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1NR)} \right) - \sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

Estimation with Omitted Variables

- If only $\mathbf{u}_h^{(2)}$, can use *fixed effects estimation* (Kim & Frees 2006) that removes both observed and omitted hospital variables to estimate discharge fixed effects β_1 :

$$\mathbf{b}_{1FE} = \left(\sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1)} \right) - \sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- Again, if only $\mathbf{u}_h^{(2)}$, can obtain unbiased estimates of **all** fixed effects β with *instrumental variables (IV) estimation* (Kim & Frees 2006):

$$\mathbf{b}_{IV} = \left(\sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{X}_h \right) - \sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- If both $\mathbf{u}_h^{(2)}$ and $\mathbf{u}_h^{(1j)}$, can only obtain estimates of $\beta_1^{(1NR)}$, the discharge variables other than race: $\mathbf{b}_{1FE*}^{(1NR)} =$

$$= \left(\sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1NR)} \right) - \sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

Estimation with Omitted Variables

- If only $\mathbf{u}_h^{(2)}$, can use *fixed effects estimation* (Kim & Frees 2006) that removes both observed and omitted hospital variables to estimate discharge fixed effects β_1 :

$$\mathbf{b}_{1FE} = \left(\sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1)} \right) - \sum_h \mathbf{X}_h^{(1)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- Again, if only $\mathbf{u}_h^{(2)}$, can obtain unbiased estimates of **all** fixed effects β with *instrumental variables (IV) estimation* (Kim & Frees 2006):

$$\mathbf{b}_{IV} = \left(\sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{X}_h \right) - \sum_h \mathbf{X}_h' \mathbf{V}_h^{-1/2} \mathbf{P}(\mathbf{H}_h) \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

- If both $\mathbf{u}_h^{(2)}$ and $\mathbf{u}_h^{(1j)}$, can only obtain estimates of $\beta_1^{(1NR)}$, the discharge variables other than race: $\mathbf{b}_{1FE*}^{(1NR)} =$

$$= \left(\sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{X}_h^{(1NR)} \right) - \sum_h \mathbf{X}_h^{(1NR)'} \mathbf{V}_h^{-1/2} \mathbf{Q}_h^* \mathbf{V}_h^{-1/2} \mathbf{y}_h$$

<i>Variables</i>	OLS Model		GLS Model	
	Estimate	t-Statistic	Estimate	t-Statistic
WHITE	-0.186	-9.581	0.035	1.769
BLACK	-0.191	-9.505	0.015	0.692
HISPANIC	0.087	4.054	0.000	0.014
ASIAN	0.056	1.758	0.090	3.151
NAT_AMER	-0.075	-1.239	0.069	1.215
<i>Covariance Components</i>	Estimate	Standard Error	Estimate	Standard Error
$var(\epsilon)$	0.7150	0.003	—	—
$var(\epsilon_{hd}^{(1)})$	—	—	0.4704	0.002
$var(\epsilon_h^{(2)})$	—	—	0.2730	0.025
<i>Model Fit Indices</i>				
-2(Max LogL)	243,683.8		203,940.6	
AIC	243,747.8		204,006.6	



Comparison of Race Fixed Effects Estimates: GLS vs Endogeneity Estimators

<i>Variables</i>	GLS		FE		IV		FE*	
	Est.	t-Stat.	Est.	t-Stat.	Est.	t-Stat.	Est.	t-Stat.
WHITE	0.035	1.769	0.036	1.822	0.036	1.827	—	—
BLACK	0.015	0.692	0.016	0.734	0.016	0.744	—	—
HISPANIC	0.000	0.014	-0.000	-0.006	0.000	0.003	—	—
ASIAN	0.090	3.151	0.090	3.142	0.090	3.152	—	—
NAT_AMER	0.069	1.215	0.069	1.209	0.070	1.217	—	—



Comparison of APR-DRG Fixed Effects Estimates: GLS vs Endogeneity Estimators

<i>Variables</i>	GLS		FE		IV		FE*	
	Est.	t-Stat.	Est.	t-Stat.	Est.	t-Stat.	Est.	t-Stat.
DRG740	0.906	8.918	0.907	8.873	0.906	8.894	0.909	9.052
DRG750	0.363	4.553	0.362	4.487	0.364	4.548	0.362	4.396
DRG751	0.029	0.361	0.027	0.332	0.029	0.367	0.029	0.348
DRG752	-0.129	-1.380	-0.129	-1.372	-0.128	-1.371	-0.125	-1.308
DRG753	0.171	2.129	0.170	2.089	0.172	2.131	0.170	2.050
DRG754	-0.256	-3.174	-0.258	-3.160	-0.256	-3.163	-0.256	-3.070
DRG755	-0.323	-4.004	-0.325	-3.987	-0.323	-3.991	-0.323	-3.885
DRG756	-0.073	-0.879	-0.079	-0.943	-0.075	-0.901	-0.080	-0.936
DRG757	0.201	2.229	0.196	2.163	0.200	2.224	0.200	2.167
DRG758	0.035	0.376	0.033	0.354	0.035	0.378	0.038	0.399
DRG759	0.536	4.243	0.537	4.266	0.537	4.223	0.531	4.278
DRGRM	0.048	3.119	0.048	3.094	0.048	3.100	0.049	3.141
DRGSEV	0.257	20.284	0.256	20.267	0.257	20.286	0.256	20.194

Summary of Current Research

- Multilevel modeling has been applied to the examination of inpatient healthcare utilization outcomes and racial disparities
- Current multilevel model-based methods, fixed effects and IV, have been utilized in the presence of omitted variables
- Empirical analysis found no evidence of significant racial disparities

