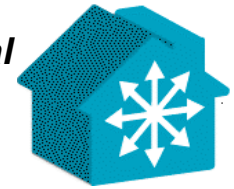
Estimating the Effects of School Indoor Air Quality on Academic Outcomes

Tess Stafford

Department of Economics
University of Texas

IGERT: Indoor Environmental Science & Engineering
The University of Texas



Why should we care about IAQ?

- ~90% of time spent indoors
- Demand for cheaper building materials and consumer products is increasing
- Buildings have been getting tighter and tighter

Why is this project important?

- 20% of US schools suffer from poor IAQ (GAO 1995)
- Children spend a large fraction of time in school
- Indoor air pollution poses a greater risk to children
- "Novel" methodological approach

Model

Academic performance $\longrightarrow$ $P_{i,s,t} = f\left(\mathbf{x}_{i,s,t}, \mathbf{c}_i, \mathbf{s}_s; \beta\right) + u_{i,s,t}$

i subscripts individual
s subscripts school
t subscripts school year

Model

Academic performance → $P_{i,s,t} = f\left(\mathbf{x}_{i,s,t}, \mathbf{c}_i, \mathbf{s}_s; \beta\right) + u_{i,s,t}$

Time dependent variables

i subscripts individual
s subscripts school
t subscripts school year

Model

Academic performance → $P_{i,s,t} = f\left(\mathbf{x}_{i,s,t}, \mathbf{c}_i, \mathbf{s}_s; \beta\right) + u_{i,s,t}$

Individual characteristics

i subscripts individual
s subscripts school
t subscripts school year

Model

Academic performance → $P_{i,s,t} = f\left(\mathbf{x}_{i,s,t}, \mathbf{c}_i, \mathbf{s}_s; \beta\right) + u_{i,s,t}$

School characteristics

i subscripts individual
s subscripts school
t subscripts school year

Model

Academic performance $\longrightarrow$ $P_{i,s,t} = f\left(\mathbf{x}_{i,s,t}, \mathbf{c}_i, \mathbf{s}_s; \beta\right) + u_{i,s,t}$

Indoor air quality

i subscripts individual
s subscripts school
t subscripts school year

Model

$$P_{i,s,t} = \mathbf{x}_{i,s,t}\beta_x + \beta_q q_{s,t} + \mathbf{c}_i\beta_c + \mathbf{s}_s\beta_s + u_{i,s,t}$$

Parameter of interest

i subscripts individual
s subscripts school
t subscripts school year

Model

$$P_{i,s,t} = \mathbf{x}_{i,s,t}\beta_x + \beta_q q_{s,t} + \mathbf{c}_i\beta_c + \mathbf{s}_s\beta_s + u_{i,s,t}$$

Concerns:

- Unobserved heterogeneity

i subscripts individual
s subscripts school
t subscripts school year

Model

$$P_{i,s,t} = \mathbf{x}_{i,s,t}\beta_x + \beta_q q_{s,t} + \mathbf{c}_i\beta_c + \mathbf{s}_s\beta_s + u_{i,s,t}$$

Concerns:

- Unobserved heterogeneity
- Can we observe *q*?

i subscripts individual
s subscripts school
t subscripts school year

Model

$$P_{i,s,t} = \mathbf{x}_{i,s,t}\beta_x + \beta_q q_{s,t} + \mathbf{c}_i\beta_c + \mathbf{s}_s\beta_s + u_{i,s,t}$$

Concerns:

- Unobserved heterogeneity → Panel data
- Can we observe *q*?

i subscripts individual
s subscripts school
t subscripts school year

Model

$$P_{i,s,t} = \mathbf{x}_{i,s,t}\beta_x + \beta_q q_{s,t} + \mathbf{c}_i\beta_c + \mathbf{s}_s\beta_s + u_{i,s,t}$$

Concerns:

- Unobserved heterogeneity → Panel data
- Can we observe *q*? → Renovation projects

i subscripts individual
s subscripts school
t subscripts school year

Quasi-Natural Experiment

- **Texas School District Renovations**
  - $50+ million spent on IAQ related renovations at ~all campuses in district during last 7 years
- **Quasi-natural experiment**
  - Projects have been spread out over time
  - For each school year, there is a control group and a treatment group

Data

- Student-level data
- Teacher-level data
- Renovation schedules and budgets

Data

- Student-level data
  - ✓ School year
  - ✓ Grade level
  - ✓ School attended
  - ✓ Attendance (by the 6-week period)
  - ✓ Test scores, test date (for 5 tests)
  - ✓ Demographic info
- Teacher-level data
- Renovation schedules and budgets

Data

- Student-level data
- Teacher-level data
  - ✓ School year
  - ✓ School taught
  - ✓ Salary
  - ✓ Stipend
  - ✓ Years experience (total and within district)
- Renovation schedules and budgets

Data

- Student-level data
- Teacher-level data
- Renovation schedules and budgets
  - ✓ School
  - ✓ Project scope
  - ✓ Project beginning and end
  - ✓ Budget

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = {} & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

Empirical Model

$$P_{i,s,t} = \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t}$$
$$+ \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t$$
$$+ \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t}$$
$$+ \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t}$$
$$+ \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}$$

- Yearly attendance rate (0 - 100)
- Math (reading) score (800 - 2800)
- Math (reading) pass rate (0 or 1)

Empirical Model

$$P_{i,s,t} = \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \} \text{ Renovation status } (q)$$
$$+ \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t$$
$$+ \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t}$$
$$+ \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t}$$
$$+ \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}$$

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = {} & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \quad \} \text{ Year dummies} \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = {} & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

} Grade dummies

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = {} & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \Big\} \mathbf{x}_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = {} & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

} Student & school FE

Empirical Model

$$
\begin{aligned}
P_{i,s,t} = & \beta_0 + \beta_1 during_{s,t} + \beta_2 after_{s,t} \\
& + \beta_3 y2001_t + \beta_4 y2002_t + \beta_5 y2003_t + \beta_6 y2004_t + \beta_7 y2005_t \\
& + \beta_8 grade1_{i,s,t} + \beta_9 grade2_{i,s,t} + \beta_{10} grade3_{i,s,t} + \beta_{11} grade4_{i,s,t} \\
& + \beta_{12} attendance_{i,s,t} + \beta_{13} classsize_{i,s,t} + \beta_{14} years\exp_{i,s,t} + \beta_{15} dist\exp_{i,s,t} \\
& + \beta_c \mathbf{c}_i + \beta_s \mathbf{s}_s + u_{i,s,t}
\end{aligned}
$$

- Student fixed effects remove **c** and **s**
- Clustering by school

Preliminary Results: Attendance Rates

Variable	All Students	Female Only	Male Only
during	-0.271	-.0278	-0.256
	(.303)	(.324)	(.291)
after	0.116	0.100	0.133
	(.121)	(.136)	(.127)
classsize	-0.013***	-0.011**	-0.015***
	(.003)	(.005)	(.005)
yearsexp	-0.003	-0.006	-0.001
	(.002)	(.003)**	(.003)
distexp	0.002	-0.003	0.002
	(.002)	(.003)	(.004)
Constant	96.848***	96.948***	96.762***
	(.109)	(.142)	(.145)
Observations	169975	82622	87300

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Preliminary Results: Math Scores

Variable	All Students	Female Only	Male Only
during	36.49	58.41 **	15.68
	(25.86)	(28.31)	(27.45)
after	29.18	36.76 **	22.48
	(20.82)	(17.12)	(26.32)
attendance	5.13***	5.07***	5.18***
	(.47)	(.73)	(.69)
classsize	-0.74	-0.97	-0.51
	(.49)	(.59)	(.62)
yearsexp	-0.07	0.10	-0.26
	(.32)	(.37)	(.34)
distexp	0.64	0.36	0.93*
	(.43)	(.48)	(.47)
Constant	1843.00 ***	1845.98 ***	1838.28 ***
	(48.54)	(78.84)	(72.58)
Observations	76848	38070	38760

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Preliminary Results: Reading Scores

Variable	All Students	Female Only	Male Only
during	32.49 *	45.67 *	20.26
	(16.9)	(24.01)	(19.65)
after	26.52 *	39.27 **	14.39
	(14.33)	(15.65)	(19.43)
attendance	4.60***	4.28***	4.94***
	(.62)	(.82)	(.72)
classsize	-0.25	-0.11	-0.40
	(.44)	(.52)	(.57)
yearsexp	0.07	0.35	-0.20
	(.27)	(.33)	(.32)
distexp	0.43	0.19	0.65*
	(.34)	(.44)	(.37)
Constant	1789.39 ***	1879.06 ***	1704.34 ***
	(61.41)	(82.4)	(72.84)
Observations	79296	39016	40261

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Preliminary Results: Math Pass Rates

Variable	All Students	Female Only	Male Only
during	0.052 *	0.049	0.056 ***
	(.03)	(.053)	(.016)
after	0.057 **	0.083 **	0.033 *
	(.026)	(.042)	(.018)
attendance	0.006 ***	0.006 ***	0.006 ***
	(.001)	(.001)	(.001)
classsize	-0.001	-0.002	-0.001
	(.001)	(.001)	(.001)
yearsexp	0.000	0.000	0.000
	(.001)	(.001)	(.001)
distexp	0.002 **	0.002 **	0.002 **
	(.001)	(.001)	(.001)
Constant	0.535 ***	0.474 ***	0.574 ***
	(.079)	(.108)	(.121)
Observations	76848	38070	38760

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Preliminary Results: Reading Pass Rates

Variable	All Students	Female Only	Male Only
during	0.033	0.045	0.023
	(.031)	(.03)	(.056)
after	0.031 **	0.021	0.040
	(.014)	(.031)	(.031)
attendance	0.004 ***	0.004 ***	0.003 ***
	(.001)	(.001)	(.001)
classsize	0.000	0.000	0.000
	(.001)	(.001)	(.001)
yearsexp	0.000	0.000	0.000
	(.001)	(.001)	(.001)
distexp	0.001	0.001	0.001
	(.001)	(.001)	(.001)
Constant	0.518 ***	0.552 ***	0.489 ***
	(.085)	(.119)	(.108)
Observations	79296	39016	40261

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Continuing Work

- Quantile Regression Framework
- Random trends model
- Compare with responses to non IAQ-related renovations
- Cost-Benefit Analysis

Continuing Work:

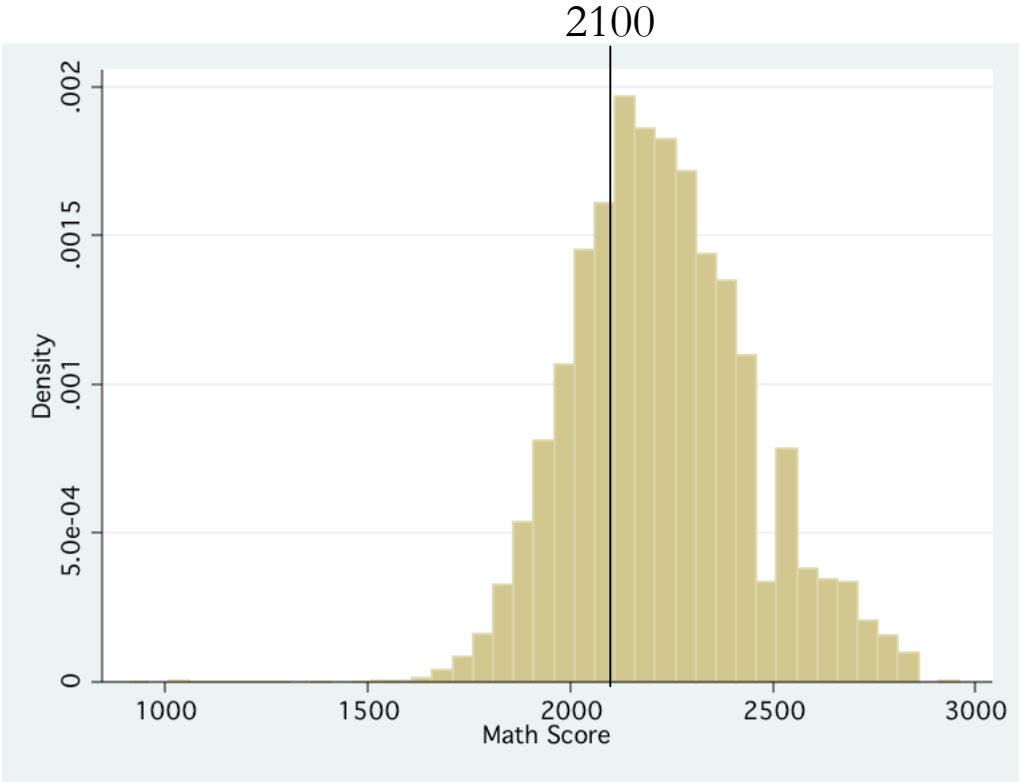
Quantile Regression Framework

- Conditional mean vs. conditional quantile

Continuing Work:

Quantile Regression Framework

- Conditional mean vs. conditional quantile



Continuing Work:

Random Trends Model

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}$$

Continuing Work:

Random Trends Model

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}$$

$$y_{it} = \mathbf{x}_{it}\beta + g_i t + c_i + u_{it}$$

Continuing Work:

Other Renovation Projects

- Want to distinguish between effect of cleaner indoor air and effect of new coat of paint

Continuing Work:

Cost Benefit Analysis

- We know “costs”
- Need to quantify “benefits”