Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: The Case of School and Neighborhood Effects

Joseph G. Altonji — Yale University

Richard K. Mansfield — Cornell University

July 23, 2015

Motivation

- Many contexts in which individual outcomes depend on both individual and group-level inputs.
- Wish to evaluate the treatment effects associated with particular groups (or group characteristics)...
- But groups are chosen endogenously either by individuals or administrators, partly based on individual inputs.

Sorting Bias

- Classic sorting problem
- Standard solutions: Experiments, IV
- Wish to evaluate the variation in treatment effects across many possible groups
- Hard to find an instrument for each group.

Contribution

- Methodological
  - Show how the inclusion of group average characteristics as controls in a standard value-added regression . . .
  - . . . can purge estimates of group treatment effects of bias from sorting on unobservable characteristics
  - Approach yields meaningful lower bound estimate of the variance in group treatment effects.
- Empirical
  - How much does it matter where a student grows up and attends school?

Summary of Empirical Results

- Moving a randomly selected student from a school at the 10th percentile of the school treatment effect distribution to the 90th percentile would ...
  - raise the predicted probability of graduation by at least .07 (with larger effects for the most vulnerable)
  - raise the predicted probability of enrollment in a four-year college by at least .15.

Relevant Literature

- **Empirical:**
  - **Experimental/Quasi-Experimental**: Oreopoulos (2003), Jacob (2004), Cullen et al. (2006), Kling et al. (2007), Deming et al. (2014), Chetty et. al. (2015)
  - **Observational**: Coleman (1966), Jenks and Brown (1975), Hoxby (2000), Hanushek et al. (2003), Speakman and Welch (2006), Altonji and Mansfield (2011), Chetty and Hendren (2015)
- **Methodological:**
  - **Control Functions**: Metcalf (1974), Altonji (1982), Olley and Pakes (1999), Levinsohn and Petrin (2003), Ackerberg et. al. (2006)
  - **Equilibrium Sorting Models**: Rosen (1974), Epple and Sieg (1999), Bayer and Ross (2006), Bayer et al. (2007), Browning et al. (2014), Lindenlaub (2015)

Alternative Value-Added Estimating Equations

- True Production Function:
  - $Y_{is} = \overbrace{X_i\beta + X_i^U\beta^U}^{Student\ Contribution} + \overbrace{Z_{s,i}\Gamma + Z_{s,i}^U\Gamma^U}^{School\ Contribution\ (VA_s)}$
  - Focus on sch. avg. treatment effects
  - Object of interest: $\Rightarrow Var(VA_s) \equiv Var(Z_s\Gamma + Z_s^U\Gamma^U)$

- Alternative school report card estimating equations:
  - $Y_{is} = X_iB_1 + \phi_s + v_{is} \Rightarrow \hat{VA}_s = \hat{\phi}_s$
  - $Y_{is} = X_iB_2 + v_{is} \Rightarrow \hat{VA}_s = Y_s - X_s\hat{B}_2$
  - $Y_s = \mathbf{X_s G_1} + Z_{2s}G_2 + v_s \Rightarrow \hat{VA}_s = Y_s - X_s\hat{G}_1 \equiv Z_{2s}\hat{G}_2 + \hat{v}_s$
  - Show that $Var(Z_{2s}\hat{G}_2 + \hat{v}_s)$ places lower bound on true variance $Var(Z_s\Gamma + Z_s^U\Gamma^U)$.

- Teacher value-added:
  - $Y_{ic} = X_iB_3 + Z_{2c}G_3 + v_{is} \Rightarrow \hat{VA}_k = \hat{G}_{3k}$
  - $Y_{ic} = X_iB_4 + \mathbf{X_c G_4} + Z_{2c}G_5 + v_{ic} \Rightarrow \hat{VA}_k = \hat{G}_{5k}$
  - Show that vector of teacher fixed effects $\hat{G}_{5k}$ likely to contain minimal sorting bias.

Starting Point . . .

- If more able individuals did not differentially value the different amenities offered by groups . . .
- . . . no sorting would take place.
- $\Rightarrow$ Variation in differently-valued amenities across groups drives the sorting.
- $\Rightarrow$ Group means of individual characteristics are *functions* of the amenity factors driving choice.

Key Insight

- If ...
  - Unobservable and observable individual characteristics both affect tastes for a common set of group amenities ...
  - and the set of relevant amenities is not too large ...
- Then ...
  - $\Rightarrow$ Equilibrium sorting leads both group-averages of observable and unobservable individual characteristics to span the same space as the vector of group amenities.
  - $\Rightarrow$ Group-averages of observables can control for group-averages of unobservables.

Simple Case

- To build intuition, consider a simple case:
  - Single index of school/neighborhood quality: A_s
  - Graduation from high school (outcome of interest) depends only on:
    - School/neighborhood input Z_s
    - Mother's education M_i (observable)
    - Athletic talent T_i (unobservable)
  - Assume athletic ability uncorrelated with mother's education unconditionally (WLOG)

Simple Case (Cont.)

- Students/parents sort to school systems/communities based on willingness to pay for quality
- $\Rightarrow$ house prices adjust to clear the market ...
- Willingness to pay: $WTP_{is} = \lambda_i A_s + \epsilon_{is}$
- Project marginal value of additional neighborhood quality λ_i onto individual characteristics:
- $\lambda_i = M_i \Theta_1 + T_i \Theta_2 + \kappa_i$
- $\Rightarrow WTP_{is} = (M_i \Theta_1 + T_i \Theta_2 + \kappa_i) A_s + \epsilon_{is}$

Simple Case (Cont.)

- Since mother's education is uncorrelated with the other factors determining willingness to pay ...
- Expected willingness to pay for quality will be monotonically increasing in mother's education:
- $E[\lambda_i|M_i] \equiv M_i\delta_1 + T_i\delta_2 + \kappa_i|M_i] = M_i\Theta_1$
- And chosen school/neighborhood quality is monotonically increasing in WTP.
- So average mother's education will increase monotonically with school/neighborhood quality:
- $E[M_i|i \text{ attends } s] \equiv M_s = f(A_s)$
- Can invert: $A_s = f^{-1}(M_s)$

Simple Case (Cont.)

- Analogously, expected WTP for quality is monotonically increasing in athletic talent:
- $E[M_i\Theta_2 + T_i\Theta_2 + \kappa_i | T_i] = T_i\Theta_2$
- $\Rightarrow$ Average athletic ability will monotonically increase with school/neighborhood quality:
- $E[T_i | i$ attends $s] \equiv T_s = g(A_s)$
- $\Rightarrow$ school-average athletic ability will monotonically increase with school-average mother's education:
- $T_s = g(f^{-1}(M_s))$

Simple Case (Cont.)

- Key takeaways:
  1. Can perfectly control for average (unobserved) athletic ability using a flexible function of average (observed) mother's education.
  2. Control function approach works even though the two student-level variables are uncorrelated unconditionally.
  3. Athletic ability and mother's education will be negatively correlated within a school
- What happens if there are multiple amenity dimensions?

General Case: Model of School Choice

- Multinomial model of school choice
- Set of L^O observable outcome-relevant student characteristics, $\{X_{1i}, \ldots, X_{L^O i}\}$
- Set of L^U unobserved outcome-relevant student characteristics, $\{X^U_{1i}, \ldots, X^U_{L^U i}\}$
- Set of K underlying desired/undesired school/neighborhood-level amenities $\{A_{1s}, \ldots, A_{Ks}\}$
    - K latent factors capture all variation in school/neighborhood characteristics differentially valued by more/less able students/parents.
- $WTP_i(s) = (X_i\Theta + X_i^U\Theta^U + \kappa_i)A_s + \epsilon_{si}$

General Case

- Analytic solution based on continuous case.
- If:
  1. The amenities driving sorting on observables are less numerous than observable characteristics
  2. For each X^U_{li}, either
     - $Cov(X_{ki}, X^U_{li}) \neq 0$ for some X_{ik}, or
     - The amenity factors driving sorting on X^U_{il} are a subset of those driving sorting on X_i
- Then:
  - The vector of group-average observables X_s can be used as a control function that fully absorbs the vector of group-average unobservables X^U_s.
  - $\Rightarrow$ Purges G_{2s} and v_s of all sorting bias.
- Derived from FOC, holds even when amenities are endogenous.
- Additive separability of WTP $\Rightarrow$ Linear control function.

Testing the Assumptions Underlying Invertibility

Fraction of Total Variance in X_s
Explained by Various Numbers of Principal Components

		NLS72		NELS88 gr8		ELS2002	
		Baseline	Full	Baseline	Full	Baseline	Full
		(1)	(2)	(3)	(4)	(5)	(6)
(1)	# of Variables in X_s	32	34	39	49	40	51
	# Factors Needed to Explain:						
(2)	75% of Total X_s Var.	7	7	7	9	6	8
(3)	90% of Total X_s Var.	12	12	13	16	11	14
(4)	95% of Total X_s Var.	15	15	17	20	14	19
(5)	99% of Total X_s Var.	20	21	22	26	20	25
(6)	100% of Total X_s Var.	23	24	27	32	26	33

Finite-Sample Properties

- Results from model based on asymptotic case:
  - Continuous space of school characteristics
  - Infinite number of students choosing each school
  - Unrestricted consideration set for each family
  - Full population of students at each school included in sample.
- Simulations suggest that theoretical result may be well-approximated in cases with:
  - $\sim$ 500 students per school
  - Small space of schools (e.g. 100 schools)
  - Parents choosing from overlapping subsets of schools ($\sim$ 15 schools)
- ...but sample averages based on small samples of students per school can be problematic
- Small departures from control function assumptions leave minimal sorting bias.

Potential Collinearity Problem

- Since X_s spans the relevant amenity space A_s ...
- If all productive school inputs were amenities when choice is made ...
- X_s will be collinear with both Z_{2s} and Z_s^U.
- But if some dimensions of school quality may be unobservable to/unappreciated by parents at time of choice ...
- $\Rightarrow$ Breaks collinearity between A_s (and therefore X_s) and realized inputs ($[Z_s, Z_s^U]$)
- Can use $Var(Z_2 G_2 + v_s)$ as a lower bound estimate of the variance in group treatment effects.
- G_2 still contains some OVB.

Data

- Survey data: NLS72, NELS88, ELS2002
- Common features:
  1. Stratified sampling: Can construct school-level averages
  2. Follow individuals to later outcomes: HS graduation, college attendance, log wages (NLS72)
  3. Wealth of individual-level and school-level variables
- Administrative Data: North Carolina student records (2007-2009)
  - Population of students at each school
  - Public/charter schools only
  - Fewer observable characteristics
  - Only outcome: HS graduation

Data: Variables in X_i

- "Baseline" specification: plausibly exogenous characteristics
  - Student Characteristics:
    - Race, 1(English is second language),1(Immigrant)
  - Parental Characteristics:
    - family composition (not in NC), parent's income (free lunch elig. in NC data), parents' education, occupational group dummies (not in NC)
    - religion dummies (not in NC), parental immigrant status, home environment index (not in NC)
- "Full" specification: Additional possibly endogenous characteristics:
  - Math and reading standardized test scores
  - Indicators of student behavior
  - Parent educational expectations

Lower Bound Estimates (With and Without Common Shocks) of the Fraction of Outcome Variance Attributable to School/Neighborhood Quality

Panel B: Fraction of Latent Index Variance Determining Enrollment Attributable to School/Neighborhood Quality						
Lower Bound	NLS		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean	0.27	0.27	0.31	0.31	0.37	0.37
Between Sch.	0.143	0.143	0.224	0.224	0.215	0.215
LB no unobs	0.026	0.019	0.018	0.015	0.022	0.018
$Var(Z_{2s}G_2)$	(0.005)	(0.004)	(0.006)	(0.005)	(0.007)	(0.006)
LB w/ unobs	0.038	0.032	0.040	0.029	0.046	0.031
$Var(Z_{2s}G_2 + v_s)$	(0.007)	(0.006)	(0.008)	(0.006)	(0.009)	(0.007)

Lower Bound Estimates (With and Without Common Shocks) of the Fraction of Outcome Variance Attributable to School/Neighborhood Quality

Panel A: Fraction of Latent Index Variance Determining Graduation Attributable to School/Neighborhood Quality						
Lower Bound	NC		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
Sample Mean	0.76	0.76	0.86	0.86	0.90	0.90
Between Sch. Var.	0.085	0.085	0.170	0.170	0.126	0.126
LB no unobs	0.018	0.013	0.011	0.006	0.025	0.024
$Var(Z_{2s}G_2)$	(0.008)	(0.004)	(0.006)	(0.004)	(0.012)	(0.011)
LB w/ unobs	0.049	0.036	0.028	0.016	0.036	0.025
$Var(Z_{2s}G_2 + v_s)$	(0.014)	(0.008)	(0.009)	(0.005)	(0.012)	(0.011)

Shifts in School Quality

- Variance decompositions may be misleading indicator of the impact of changing schools:
  - $\Rightarrow$ Small variance component still corresponds to substantial standard deviation
  - Many students near the decision margin for dropping out, attending college.
- $\Rightarrow$ We consider average impact of moving from 10th quantile to 50th or 90th quantile of school/neighborhood quality.
- Implementation:
  1. Set $Z_{2s}G_2$ or $Z_{2s}G_2 + v_s$ at appropriate quantile
  2. integrate over distribution of shocks, observable and unobservable student characteristics.

Lower Bound Estimates of the Impact of Shifts in School Quality on Enrollment in a 4-Year College

Effect on Enrollment Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile

Lower Bound	NLS		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.139	0.118	0.127	0.112	0.155	0.132
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.012)	(0.018)	(0.017)	(0.019)	(0.017)
LB w/ unobs: 10th-90th	0.170	0.152	0.188	0.155	0.216	0.172
Based on $Var(Z_{2s}G_2 + v_s)$	(0.017)	(0.016)	(0.020)	(0.018)	(0.021)	(0.020)
LB no unobs: 10th-50th	0.065	0.056	0.061	0.054	0.075	0.064
Based on $Var(Z_{2s}G_2)$	(0.006)	(0.005)	(0.008)	(0.008)	(0.008)	(0.008)
LB w/ unobs: 10th-50th	0.078	0.071	0.088	0.073	0.103	0.083
Based on $Var(Z_{2s}G_2 + v_s)$	(0.007)	(0.007)	(0.009)	(0.008)	(0.009)	(0.009)

Lower Bound Estimates of the Impact of Shifts in School Quality on High School Graduation

Effect on Graduation Probability of a School System/Neighborhood at the 50th or 90th Percentile of the Quality Distribution vs. the 10th Percentile						
Lower Bound	NC		NELS gr8		ELS	
	Baseline	Full	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)	(5)	(6)
LB no unobs: 10th-90th	0.106	0.084	0.061	0.047	0.070	0.068
Based on $Var(Z_{2s}G_2)$	(0.022)	(0.014)	(0.014)	(0.012)	(0.013)	(0.012)
LB w/ unobs: 10th-90th	0.174	0.152	0.098	0.075	0.083	0.070
Based on $Var(Z_{2s}G_2 + v_s)$	(0.026)	(0.017)	(0.017)	(0.013)	(0.013)	(0.013)
LB no unobs: 10th-50th	0.056	0.044	0.033	0.025	0.040	0.038
Based on $Var(Z_{2s}G_2)$	(0.013)	(0.008)	(0.008)	(0.007)	(0.009)	(0.008)
LB w/ unobs: 10th-50th	0.096	0.083	0.055	0.041	0.048	0.039
Based on $Var(Z_{2s}G_2 + v_s)$	(0.016)	(0.010)	(0.010)	(0.008)	(0.009)	(0.008)

Lower Bound Estimates of the Impact of Shifts in School Quality on Years of Postsecondary Education and Permanent Adult Wages (With and Without Between-School Residual)

Panel C: Years of Postsecondary Education and Permanent Wages (NLS72 data)				
Lower Bound	Yrs. Postsec. Ed.		Perm. Wages	
	Baseline	Full	Baseline	Full
	(1)	(2)	(3)	(4)
LB no unobs: 10th-90th	0.578	0.445	0.152	0.157
Based on $Var(Z_{2s}G_2)$	(0.054)	(0.039)	(0.019)	(0.019)
LB w/unobs: 10th-90th	0.661	0.520	0.177	0.177
Based on $Var(Z_{2s}G_2 + v_s)$	(0.069)	(0.047)	(0.028)	(0.023)
LB no unobs: 10th-50th	0.283	0.222	0.076	0.079
Based on $Var(Z_{2s}G_2)$	(0.027)	(0.019)	(0.010)	(0.010)
LB w/unobs: 10th-50th	0.331	0.260	0.088	0.088
Based on $Var(Z_{2s}G_2 + v_s)$	(0.035)	(0.024)	(0.014)	(0.012)

Conclusions

- If observe a rich/diverse set of student characteristics . . .
- and the relevant set of amenities is not too high dimensional . . .
- Can control for sorting on unobservables using group averages of observables
- Can place a lower bound on the variance in group treatment effects.
- Find that moving from a low to a high quality school/neighborhood environment has a sizeable impact on educational and labor market outcomes (particularly for those on the decision margin).