Session 5: The consequences of ghettoization

Jacob L. Vigdor

Duke University and NBER

8 September 2010

Outline

- Challenges in estimating neighborhood effects.
- Efforts to bound the impact of neighbors.
- Timing-based identification.
- Instrumental variables estimates.
- Quasi-experimental estimates.

Neighborhood effects

- Selection into neighborhoods (internal validity).
- Uncertainty regarding the "correct" specification of neighborhood or the neighborhood characteristics that matter (construct validity).
- Manski's reflection problem: when estimating individual's y as a function of neighbors' y. (internal validity)
- Uncertainty regarding the generalizability of results (external validity).
- Effects on adults vs. effects on children: experimental estimates of the former but not yet the latter.

Upper bounds

- Find correlation in outcomes between individuals who grew up in the same neighborhood. This incorporates:
  - Causal effect of neighborhood conditions on outcomes.
  - Similarity in unobserved conditions between neighbors.
- Bounding exercise is interesting if the bound is small, but not if the bound is high.
- Use correlation between siblings as a comparison, though the interpretation of that is unclear.
- Solon, Page, and Duncan (2000) for the US.

Raaum, Salvanes, and Sorensen (2006)

- Norwegian longitudinal population data.
- Examine cohorts born 1946-1955 and 1956-1965.
- Use bootstrapping procedure to impute the variance of the bound estimate.

Table 3

Correlation in Education and Adult Earnings Among Siblings and Neighbouring Children

	Siblings		Neighbours	
	1946–55	1956–65	1946–55	1956–65
Education 1995				
Males	0.4150	0.4213	0.1121	0.0612
	(0.0088)	(0.0075)	(0.0261)	(0.0075)
Adjusted for parental education (PE)			0.0590	0.0245
			(0.0111)	(0.0030)
Adjusted for family structure (FS)			0.1105	0.0602
			(0.0260)	(0.0076)
Adjusted for PE and FS			0.0494	0.0206
			(0.0094)	(0.0032)
Adjusted for PE, FS and parental income				0.0163
				(0.0034)
Females	0.4561	0.4739	0.1027	0.0653
	(0.0064)	(0.0080)	(0.0213)	(0.0095)
Adjusted for parental education (PE)			0.0493	0.0245
			(0.0062)	(0.0046)
Adjusted for family structure (FS)			0.1013	0.0642
			(0.0046)	(0.0050)
Adjusted for PE and FS			0.0405	0.0202
			(0.0205)	(0.0095)
Adjusted for PE, FS and parental income				0.0153
				(0.0050)

Table 3

Correlation in Education and Adult Earnings Among Siblings and Neighbouring Children

	Siblings		Neighbours	
	1946–55	1956–65	1946–55	1956–65
Average log Earnings 1990–95				
Males	0.2032	0.1845	0.0591	0.0283
	(0.0082)	(0.0059)	(0.0074)	(0.0051)
Adjusted for parental education (PE)			0.0499	0.0252
			(0.0059)	(0.0047)
Adjusted for family structure (FS)			0.0584	0.0280
			(0.0077)	(0.0048)
Adjusted for PE and FS			0.0470	0.0245
			(0.0053)	(0.0052)
Adjusted for PE, FS and parental income				0.0221
				(0.0051)
Females	0.1480	0.1645	0.0292	0.0201
	(0.0053)	(0.0043)	(0.0055)	(0.0024)
Adjusted for parental education (PE)			0.0225	0.0141
			(0.0041)	(0.0022)
Adjusted for family structure (FS)			0.0287	0.0197
			(0.0051)	(0.0024)
Adjusted for PE and FS			0.0206	0.0127
			(0.0036)	(0.0021)
Adjusted for PE, FS and parental income				0.0104
				(0.0021)

Timing-based identification

- Use parent-generation measures of local characteristics as a predictor of child outcomes.
- Alternatively, examine the impact of local conditions on young adults, presuming that youth locations are determined by parent decisions.
- Not particularly convincing strategies if parent-generation unobservables correlate with both location characteristics of interest and outcomes.

Borjas (1995)

- Blend of a traditional ethnic capital model and a neighborhood effects model.
  - Is "ethnic capital" really just standing in for a measure of neighborhood-level inputs?
  - Does ethnicity play a stronger role for individuals residing in an enclave?
- Two datasets: 1970 Census microdata (neighborhood=census tract where you live) and NLSY (neighborhood=ZIP code when you were 14-22).

TABLE 1—RESIDENTIAL SEGREGATION IN THE 1970 CENSUS

Neighborhood characteristics of average person in:	Percentage of population in neighborhood that is:				
	First generation	First or second generation	Black	Hispanic	Sample size
First generation	15.3	32.7	6.9	10.2	63,099
Second generation	6.7	28.2	4.3	5.2	156,134
Third generation	3.8	13.8	11.7	3.9	905,213
Hispanics:					
First generation	22.2	36.7	6.5	35.0	10,713
Second generation	9.4	27.3	5.1	33.0	10,801
Third generation	8.9	21.9	11.4	28.8	25,202
Third generation:					
Blacks	3.1	8.0	54.7	3.7	109,533
Whites	3.7	14.4	5.6	3.1	771,359

Notes: The "white" sample includes all nonblack, non-Hispanic third-generation workers. The population proportions are as follows: immigrants, 4.8 percent; first or second generation, 16.6 percent; blacks, 11.1 percent; and Hispanics, 4.4 percent.

TABLE 6—ESTIMATES OF INTERGENERATIONAL CORRELATION IN 1970 CENSUS

	Regressions using neighborhood file				Regressions using county group file		
Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Education:							
Mean of group in 1940	0.3649	—	0.1707	0.2670	0.3628	—	0.3316
	(0.0828)		(0.0457)	(0.0557)	(0.0833)		(0.0709)
Includes neighborhood fixed effects	no	—	yes	no	—	—	—
Includes county fixed effects	—	—	—	—	no	—	yes
Includes neighborhood characteristics	no	—	no	yes	no	—	no
Log wage:							
Mean of group in 1940	0.4549	0.3974	0.2191	0.2474	0.4607	0.3710	0.3938
	(0.0781)	(0.0662)	(0.0578)	(0.0362)	(0.0874)	(0.0694)	(0.0772)
Includes skill-adjusted wage level	no	yes	no	yes	no	yes	no
Includes neighborhood fixed effects	no	no	yes	no	—	—	—
Includes county fixed effects	—	—	—	—	no	no	yes
Includes neighborhood characteristics	no	no	no	yes	no	no	no
Log wage, adjusted for education:							
Mean of group in 1940	0.2038	0.1767	0.1101	0.1020	0.2132	0.1589	0.1701
	(0.0400)	(0.0321)	(0.0413)	(0.0193)	(0.0511)	(0.0352)	(0.0440)
Includes skill-adjusted wage level	no	yes	no	yes	no	yes	no
Includes neighborhood fixed effects	no	no	yes	no	—	—	—
Includes county fixed effects	—	—	—	—	no	no	yes
Includes neighborhood characteristics	no	no	no	yes	no	no	no

Notes: Standard errors are reported in parentheses; the sample size is 53,703. All regressions include a second-order polynomial in the worker's age. The neighborhood characteristics included in column (iv) are the fraction of persons in the neighborhood with at least 12 years of schooling, the fraction with at least 16 years of schooling, the labor-force participation rates of men and women, the unemployment rate, the fraction of persons working in professional occupations, the fraction of families below the poverty line, and the fraction of families that earn at least $15,000 annually. The regressions use a random-effects estimator.

TABLE 8—ESTIMATES OF THE ETHNIC-CAPITAL MODEL IN THE NLSY

Variable	Regression (i)	(ii)	(iii)	(iv)
Education:				
Parental skills	0.2404	0.2005	0.1745	0.1784
	(0.0666)	(0.0669)	(0.0718)	(0.0849)
Ethnic capital	0.2004	0.1356	0.0376	0.1480
	(0.0465)	(0.0301)	(0.0288)	(0.0504)
Includes county fixed effects	no	yes	no	no
Includes neighborhood fixed effects	no	no	yes	no
Includes neighborhood characteristics	no	no	no	yes
Log wage:				
Parental skills	0.3774	0.2645	0.2500	0.2460
	(0.0371)	(0.0398)	(0.0418)	(0.0480)
Ethnic capital	0.3190	0.3107	0.0458	0.0229
	(0.1559)	(0.1116)	(0.1331)	(0.1636)
Includes county fixed effects	no	yes	no	no
Includes neighborhood fixed effects	no	no	yes	no
Includes neighborhood characteristics	no	no	no	yes
Log wage, adjusted for education:				
Parental skills	0.1765	0.1158	0.1214	0.1221
	(0.0369)	(0.0394)	(0.0410)	(0.0476)
Ethnic capital	0.0759	0.1581	−0.0231	−0.0584
	(0.1571)	(0.1141)	(0.1289)	(0.1621)
Includes county fixed effects	no	yes	no	no
Includes neighborhood fixed effects	no	no	yes	no
Includes neighborhood characteristics	no	no	no	yes

Notes: Standard errors are reported in parentheses. The sample size is 7,569 for the educational-attainment regressions and 4,261 for the log-wage regressions. All regressions include variables indicating the worker's age, gender, whether the person is first-generation or second-generation, and whether the person is enrolled in school in 1990. The neighborhood characteristics included in column (iv) are the average educational attainment and the average log wage of parents in the neighborhood. The regressions use a random-effects estimator.

TABLE 9—ESTIMATES OF INTERGENERATIONAL CORRELATION, BY ETHNIC COMPOSITION OF NEIGHBORHOOD

	Education				Log wage			
	(i)		(ii)		(i)		(ii)	
Ethnic composition of neighborhood	Parental skills	Ethnic capital	Parental skills	Ethnic capital	Parental skills	Ethnic capital	Parental skills	Ethnic capital
A. *1970 Census*								
Percentage with same ethnicity:								
0 percent	—	0.2458	—	0.1467	—	0.2567	—	0.1322
		(0.1195)		(0.0781)		(0.1020)		(0.0447)
Between 0 percent and 15 percent	—	0.3206	—	0.2261	—	0.4702	—	0.2920
		(0.1410)		(0.0930)		(0.1320)		(0.0653)
More than 15 percent	—	0.5325	—	0.2711	—	0.6769	—	0.3782
		(0.2338)		(0.2166)		(0.1496)		(0.1091)
B. *NLSY*								
Percentage with same ethnicity:								
Less than 5 percent	0.2748	0.1482	0.2071	0.0491	0.4636	0.1850	0.3178	0.0290
	(0.0126)	(0.0791)	(0.0131)	(0.0257)	(0.0719)	(0.2085)	(0.0758)	(0.1422)
Between 5 percent and 33 percent	0.2933	0.2699	0.2014	0.0439	0.4198	0.2189	0.3292	0.0152
	(0.0116)	(0.0863)	(0.0125)	(0.0267)	(0.0654)	(0.2092)	(0.0737)	(0.1440)
More than 33 percent	0.1965	0.2998	0.1311	0.1188	0.3828	0.2958	0.2586	0.1429
	(0.0105)	(0.0848)	(0.0105)	(0.0268)	(0.0575)	(0.2094)	(0.0618)	(0.1253)
Includes neighborhood fixed effects?	no		yes		no		yes	

Notes: Standard errors are reported in parentheses. The Census regressions include a second-order polynomial in the worker's age. The NLSY regressions control for the worker's age, gender, whether the person is first- or second-generation, and whether the person is enrolled in school in 1990. The Census regressions have 53,703 observations; the NLSY education regressions have 7,569 observations, and the NLSY log-wage regressions have 4,261 observations. The regressions use a random-effects estimator.

Cutler, Glaeser and Vigdor (2005)

- Alternative statement of Borjas hypothesis: the effects of seregation depend on the characteristics of the segregated group.
- Use earlier data: how does group/city level segregation in 1910 affect adult outcomes for 2nd generation in 1940?
- Allow effect of segregation to vary by parent-generation skills, measured using occupation data.
- Answer: segregation appears beneficial for high-skill groups, but not for low-skill groups.

Cutler, Glaeser, and Vigdor (2009)

- Similar exercise to (2005), with some modifications:
  - Examine outcomes of young adults in 1990 Census data, which includes information on tract of residence. Not a public use dataset.
  - Test for effects of group share in tract as well as segregation level.
  - Instrumental variables to address endogenous selection.

CGV (2009)

- Instrument for group share in tract:
  - Treat number of non-group members in tract, number of group members in city, group's national occupation distribution, and distribution of occupations across tracts within the city as given.

Predicted group size = $p_{ij}R_iO_j$

p_{ij} is scalar, group i population in city j

R_i is t by k matrix, describes distribution of occupations across tracts

O_j is k by 1 vector of group's distribution across occupations.

CGV (2009)

- The use of segregation – a city-level measure – as a substitute for tract group share is itself a reduced-form strategy to address selection (cf. Evans, Oates and Schwab 1992).
- Instrument for segregation with a measure of group's average years since immigration. Control directly for individual's own years since immigration.

Table 2
Tract-level group share and outcomes: Results with MSA and country-of-origin fixed effects

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Group share of tract population	−0.076***	0.038***	−0.195***	0.048***
	(0.006)	(0.006)	(0.033)	(0.006)

Table 3
Tract-level group share and outcomes: Instrumental Variable specifications

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Group share of tract population	−0.103***	0.030	−0.274***	0.003
	(0.012)	(0.020)	(0.054)	(0.027)

Table 4
Testing for heterogeneity in the effect of tract share

Dependent variable	OLS/Fixed effects		IV	
	Main effect	Interaction	Main effect	Interaction
English ability	−0.285***	0.038***	−0.234***	0.024***
	(0.031)	(0.006)	(0.012)	(0.008)
School enrollment	0.090*	−0.010	0.018	0.002
	(0.053)	(0.010)	(0.073)	(0.012)
ln(earnings)	−0.008	−0.034	−0.623***	0.063**
	(0.168)	(0.029)	(0.222)	(0.032)
Idle	0.088**	−0.007	0.210*	−0.038**
	(0.043)	(0.008)	(0.125)	(0.018)

Notes. Main effect is the coefficient on tract share in each specification. Interaction term is between tract share and the mean education level of immigrants of the same nationality within the same metropolitan area. Standard errors, corrected for within-tract/group clustering, in parentheses.

* Significance at the 10% level.
** Idem, 5%.
*** Idem, 1%.

Table 5

Panel A: Dissimilarity and outcomes: Results with MSA and country-of-origin fixed effects

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Dissimilarity index	0.017	−0.021	−0.225*	−0.064**
	(0.036)	(0.069)	(0.120)	(0.033)

Panel B: Isolation and outcomes: Results with MSA and country-of-origin fixed effects

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Isolation index	0.232**	−0.188	0.250	0.133
	(0.107)	(0.202)	(0.512)	(0.204)

Table 6

Panel A: Dissimilarity and outcomes: Results from Instrumental Variable specifications

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Dissimilarity index	0.104**	0.067	0.371**	0.047
	(0.045)	(0.108)	(0.170)	(0.044)

Panel B: Isolation and outcomes: Results from Instrumental Variable specifications

Independent variable	English ability (age 9–18)	Enrollment (age 16–18)	ln(Earnings) (age 20–30)	Idle (age 20–30)
Isolation index	0.970**	0.506	2.963**	0.388
	(0.402)	(0.813)	(1.525)	(0.361)

Table 7
Testing for heterogeneity in the effects of segregation

Dependent variable	Dissimilarity				Isolation			
	OLS/Fixed effects		IV		OLS/Fixed effects		IV	
	Main effect	Interaction	Main effect	Interaction	Main effect	Interaction	Main effect	Interaction
English	0.207*	−0.021*	−0.254	0.041**	0.239	−0.001	−2.770**	0.350**
ability	(0.109)	(0.012)	(0.154)	(0.018)	(0.448)	(0.051)	(1.214)	(0.138)
School	−0.115	0.011	−0.208	0.033	0.852	−0.130	−1.950	0.239
enrollment	(0.333)	(0.036)	(0.370)	(0.045)	(1.00)	(0.116)	(2.579)	(0.293)
ln(earnings)	0.216	−0.049	−1.571***	0.238***	2.965*	−0.331*	−14.021***	1.637***
	(0.544)	(0.058)	(0.589)	(0.080)	(1.600)	(0.183)	(5.212)	(0.568)
Idle	−0.099	0.004	−0.048	0.012	0.587	−0.055	−0.740	0.108
	(0.192)	(0.020)	(0.141)	(0.017)	(0.840)	(0.094)	(0.920)	(0.102)

Notes. Main effect is the coefficient on the relevant segregation index in each specification. Interaction term is between segregation and the mean education level of immigrants of the same nationality within the same metropolitan area. Standard errors, corrected for within MSA/group clustering, in parentheses.

* Significance at the 10% level.

** Idem, 5%.

*** Idem, 1%.

Edin, Fredriksson and Åslund (2003)

- Sweden: refugee assignment policy leads to random-conditional-on-observables variation in group share at the municipality level, 1985-1991.
- Test for heterogeneous effects based on enclave "quality," as measured by pre-assignment policy labor market income and self-employment rates.

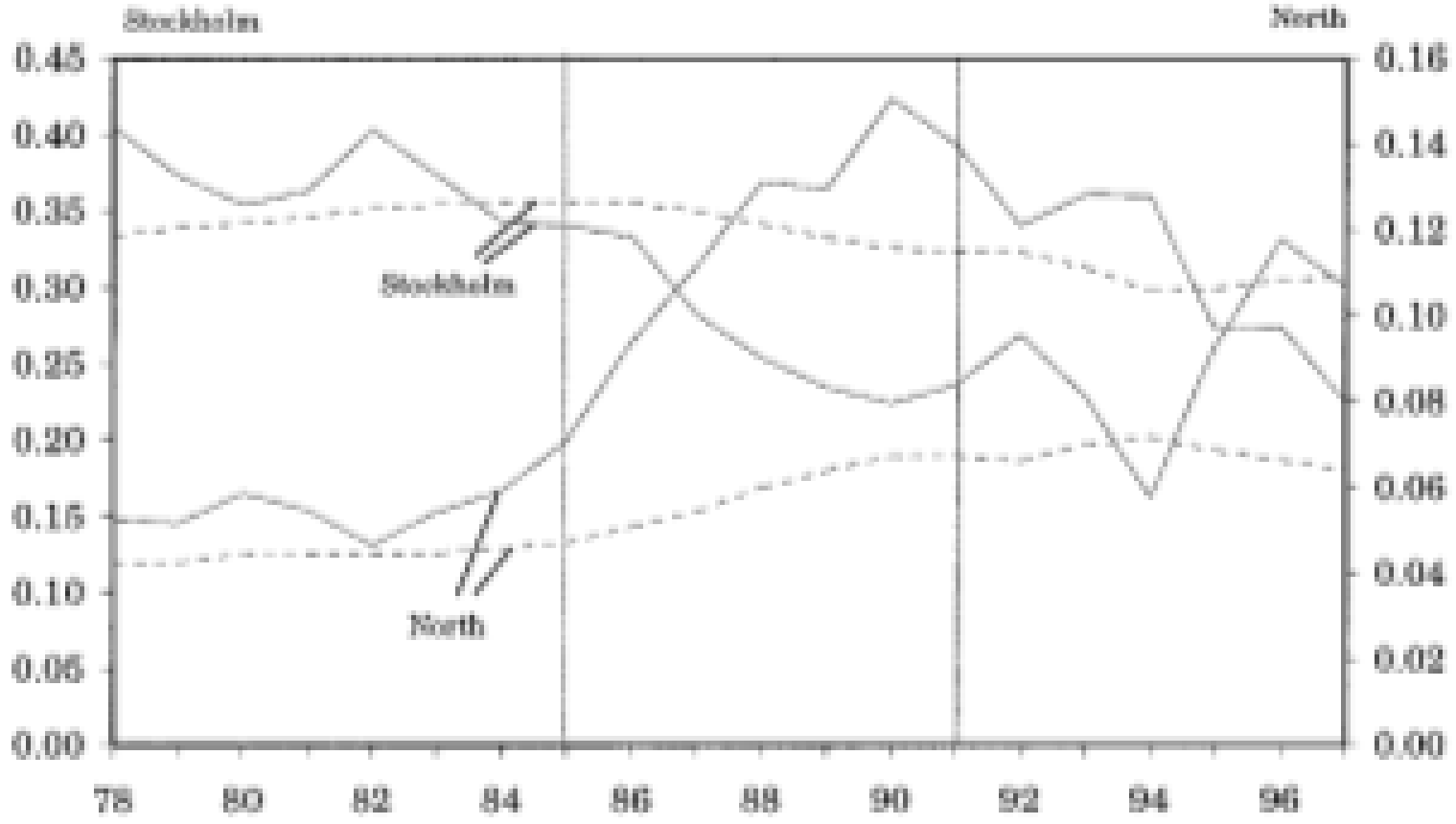

FIGURE I

Share of Non-OECD Immigrant Inflow (solid) and Stock (dashed) Located in Stockholm and in the North of Sweden, Respectively, 1978–1997

"Stockholm" refers to the county of Stockholm, and "North" to the six northernmost counties of Sweden. Authors' calculations using the LINDA immigrant sample.

TABLE 1

INDIVIDUAL CHARACTERISTICS BY INITIAL PLACEMENT

	Initial placement	
	Enclave	No enclave
Female	.44	.45
Age	37.3	37.6
	(7.7)	(7.4)
Years of schooling	11.3	11.7
	(3.0)	(2.9)
Married	.63	.62
Kid $\leq$15 years of age	.55	.57
No. of individuals	3094	3324

TABLE III

BASELINE ESTIMATES—DEPENDENT VARIABLE: ln(EARNINGS)

	Full sample		Low education (10 years or less)		High education (more than 10 years)	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
ln(ethnic group)	−.056	.012	−.053	.174	−.050	−.057
	(.022)	(.050)	(.024)	(.088)	(.030)	(.080)

TABLE IV

THE "QUALITY" OF ENCLAVES—DEPENDENT VARIABLE: ln(EARNINGS)

	Full sample		Low education (10 years or less)		High education (more than 10 years)	
	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
ln(ethnic group)	−.221	−.138	−.031	.027	−.315	−.267
	(.109)	(.071)	(.220)	(.114)	(.161)	(.113)
ln(ethnic group)*ethnic inc	.044		.039		.047	
	(.015)		(.040)		(.019)	
ln(ethnic group)*ethnic self-employment rate		3.212		2.964		4.592
		(.926)		(1.589)		(1.309)

TABLE V

REDUCED-FORM ESTIMATES OF THE "QUALITY" OF ENCLAVES AT DIFFERENT POINTS IN TIME

DEPENDENT VARIABLE: ln(EARNINGS)

Year after immigration	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
$\ln(\text{ethnic group})_t$	−.010	−.065	−.096	.007	−.045	−.115	−.101
	(.042)	(.048)	(.061)	(.064)	(.051)	(.060)	(.050)
$\ln(\text{ethnic group})_t$ * ethnic inc	−.004	.006	.009	.002	.011	.021	.022
	(.006)	(.007)	(.009)	(.011)	(.008)	(.010)	(.009)
No. of individuals	5473	5455	5332	5209	5327	5637	6393
Standard error of regression	1.16	1.28	1.31	1.32	1.30	1.24	1.44

Damm (2009)

- Raises concern with Edin et al.: use of "stock" enclave measure when only the "flow" is exogenous.
- Using Danish data from the refugee-dispersal-policy era, instruments for size of stock with magnitude of flow.

Table 2
Summary Statistics for Refugees with Positive Earnings 7 Years after Immigration: Means and Standard Deviations (Parentheses)

	ln(Local Inflow of Assigned Conationals)		t-Test of Difference in Means
	Below Average	Above Average	
Education missing and < 10 years	.47	.47	1.00
	(.50)	(.50)	
High school	.39	.40	1.39
	(.49)	(.49)	
University	.13	.13	.56
	(.34)	(.33)	
Female	.24	.21	3.37
	(.43)	(.41)	
Age	26.43	26.10	2.98
	(6.36)	(6.19)	
Number of children	.71	.61	5.08
	(1.17)	(1.12)	
Married	.35	.31	5.20
	(.48)	(.46)	
Number of observations	2,631	3,016	

NOTE.—Variables refer to the values in year t, where t is the year of immigration. Mean (standard deviation) of ln(local inflow of assigned conationals) is 3.45 (1.41).

Table 4
OLS First-Stage Earnings Regression Estimates of 2SLS: Dependent Variable: ln(Local Ethnic Stock)

	Full Sample (1)	Low Education (12 Years or Less) (2)	High Education (More than 12 Years) (3)
ln(local inflow of assigned conationals)	.306***	.308***	.273***
	(.041)	(.042)	(.094)

Table 5
Baseline Estimates: Dependent Variable: ln(Earnings)

	Full Sample		Low Education (12 Years or Less)		High Education (More than 12 Years)	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
ln(local ethnic stock)	−.066***	.216**	−.062***	.224*	−.132***	.213
	(.015)	(.112)	(.017)	(.118)	(.037)	(.332)

Table 8
The "Quality" of Enclaves (IV Estimates): Dependent Variable: ln(Earnings)

	Full Sample				Low Education (12 Years or Less)				High Education (More than 12 Years)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ln(local ethnic stock)	.07	.13	.06	.11	.06	.18	.05	.14	.26	.10	.26	.08
	(.15)	(.15)	(.17)	(.15)	(.16)	(.16)	(.19)	(.16)	(.45)	(.42)	(.48)	(.43)
ln(local ethnic stock) × ethnic earnings ($\times 10^{-4}$)	.07***				.08*				−.02			
	(.03)				(.04)				(.08)			
ln(local ethnic stock) × ethnic self-employment rate		2.67				1.49				3.62		
		(2.28)				(2.60)				(5.36)		
ln(local ethnic stock) × ethnic group share with at least 10 years of education			.55*				.61				−.15	
			(.33)				(.40)				(.75)	
ln(local ethnic stock) × ethnic group share with at least 13 years of education				1.32				1.10				1.21
				(.97)				(1.13)				(1.83)
R^2	.026	.022	.033	.028	.032	.033	.042	.035	.176	.182	.175	.199
Number of individuals	5,647				4,745				902			

NOTE.—Controls are the same as in table 5. The instrument for the log of the local ethnic stock in year $t + 7$ is the log of the inflow of assigned conationals to individual i's municipality of assignment since 1986 until the end of year t. The instrument for the interaction between quality and the log of the local ethnic stock in year $t + 7$ is the interaction between quality and the log of the inflow of assigned conationals to individual i's municipality of assignment since 1986 until the end of year t. Standard errors, reported in parentheses, are corrected for heteroskedasticity and clustering of the residuals by municipality of assignment and year of immigration.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Summary

- For first generation:
  - Negative selection into enclaves.
  - Reasonable efforts to address selection reveal a positive mean impact, with stronger positive effects associated with higher "quality" enclaves.
- For second generation:
  - The experimental papers have yet to be written!
  - "mean" effects somewhat uncertain, but continued evidence that growing up in a higher "quality" enclave is better.