SUPPLEMENTAL APPENDIX

DO "CAPITALIZATION EFFECTS" FOR PUBLIC GOODS REVEAL THE PUBLIC'S WILLINGNESS TO PAY?

Section I uses Tinbergen's linear-quadratic-normal model to provide numerical examples of the relationship between measures of MWTP derived from preference functions and capitalization effects derived from the adjustments to the reduced form of the equilibrium that clears the market following an exogenous shock. Section II describes robustness checks on our main empirical results. Section III presents summary statistics for the data describing Portland, OR; Los Angeles, CA; Philadelphia, PA; and Detroit, MI. Section IV provides instructions for obtaining our data and code.

I. *Numerical Examples of Conflation Bias: Linear-Quadratic-Normal Model*

Using the model from section 3.3, we select the following values for the structural parameters,

$$k = [g, x_1, x_2], \qquad \mu_\alpha = [20 \quad 50 \quad 25], \qquad \mu_k = [5 \quad 10 \quad 0],$$

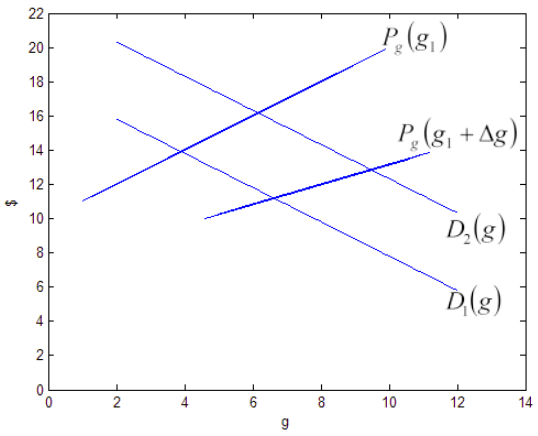
$$\sigma_\alpha = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \qquad \zeta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \text{and} \quad \sigma_k = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

These values imply that all three characteristics are normal goods and the demand for each is downward sloping. The multivariate normal distributions are used to take one million random draws. After evaluating the price function in period 1, we shock g using $\Delta g \sim N(3, 0.25)$ and $\text{cov}(\Delta g, x_1) = \text{cov}(\Delta g, x_2) = 0$. Then we evaluate the price function in period 2 and determine the capitalization effect.

Figure A1 reports results for two different values of $\text{cov}(\Delta g, g_1)$. Each panel shows the implicit marginal price functions for g before and after the shock, as well as demand curves for two households. Because demand is downward sloping, a positive shock increases the price of housing but decreases the MWTP for a further improvement.

In panel A, g_1 and Δg are negatively correlated so that areas with the lowest baseline

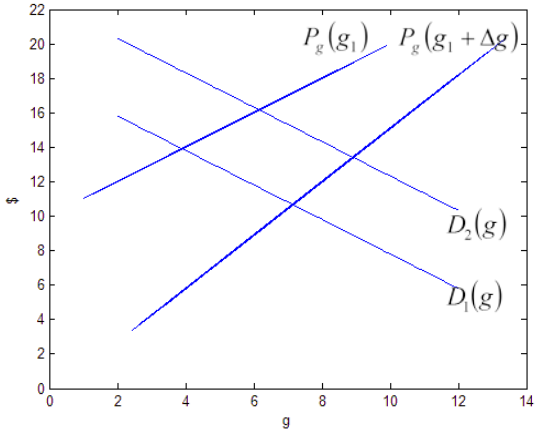
levels of g receive the largest improvements. This results in a sufficiently large upward bias on the capitalization-based estimate for MWTP ($18.12), that it exceeds the true average MWTP in the pre-shock equilibrium ($15). Intuitively, this is gentrification. The households who value g the most bid up prices in improved areas by more than the average resident is willing to pay.



$\text{cov}(\Delta g, g_1) = -0.5$

mean pre-shock MWTP: $15.00
mean post-shock MWTP: $11.97
capitalization effect: $18.12

Panel A: *Gentrification*



$\text{cov}(\Delta g, g_1) = 0.5$

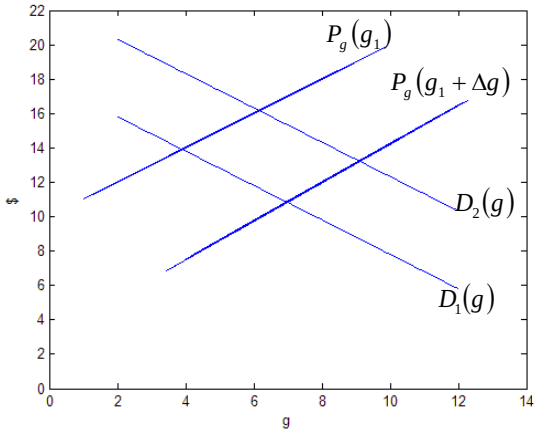
mean pre-shock MWTP: $15.00
mean post-shock MWTP: $11.85
capitalization effect $6.03

Panel B: *Preferential attachment*

FIGURE A1.—Difference between capitalization effects and MWTP.

Panel B demonstrates the opposite case where areas with the highest baseline levels of g receive the largest improvements. In this case, the capitalization-based estimate for MWTP ($6.03) understates average MWTP in the post-shock equilibrium ($11.85). This example of preferential attachment is consistent with Starrett's (1981) observation that there is little upward pressure on prices when the highest quality neighborhood is improved. Together, panels A and B illustrate how the market sorting process that underlies a hedonic equilibrium can drive a wedge between capitalization effects and MWTP.

Finally, figure A2 reports results for a case where $\text{cov}(\Delta g, g_1) = 0$. This setup mimics a randomized experiment. Δg is independent of initial levels and changes in all other variables. While the marginal price function shifts, the capitalization effect ($12.08) provides a close approximation to average MWTP in the new equilibrium ($11.92). The econometrics behind this result are explained in section 4 of the paper.



$\text{cov}(\Delta g, g_1) = 0.0$

mean pre-shock MWTP:	$15.00
mean post-shock MWTP:	$11.92
capitalization effect	$12.08

FIGURE A2.—Difference between capitalization effects and MWTP.

II. *Robustness Checks for Sample Selection Bias and Aggregation Bias*

In addition to conflation bias and time-varying omitted variables, two other features of our research design have the potential to explain the large differences between our baseline estimates for capitalization effects and the hedonic price function parameters reported in table 3: sample selection and data aggregation. First consider the scope for sample selection. Houses located outside the 0.2 mile boundary zones are included in the capitalization model in panel C of table 3, but excluded from the hedonic regressions in panels A and B

of the same table. The excluded houses comprise a large share of housing sales in each metro area, from 35% in Portland to 92% in Los Angeles. Differences between the capitalization and hedonic results could arise from differences in the distribution of properties located in the excluded and included areas. To evaluate this possibility, we repeat estimation of the basic hedonic model (without boundary fixed effects) using all of the micro data that were used to construct the block group averages for the capitalization model. Results are reported in columns 6-10 of table A1. They essentially mirror the original hedonic estimates from panel A of table 3, which is repeated for convenience in columns 1-5 of table A1. Given the large sample sizes, it is remarkable that only two of the ten coefficients are statistically different (Fairfax and Detroit in 2003). We interpret these results as evidence against the hypothesis that sample selection is driving the differences between our hedonic and capitalization-based estimates.

In principle, the difference between our hedonic and capitalization results could also arise from aggregation bias created by averaging micro data over Census block groups (table 3, panel C) or 0.2 mile boundary zones (table 3, panel D). The concern is that the "average" house in a given block group or boundary zone need not correspond to any point on the hedonic price surface. It is difficult to predict the direction and magnitude of the resulting bias. Past studies that have used Census aggregates have assumed the bias is sufficiently small to ignore (Chay and Greenstone 2005, Greenstone and Gallagher 2008, Baum-Snow and Marion 2009). To evaluate this assumption, we first aggregate the micro data from panel B of table A1 into block groups and repeat the estimation. Results are reported in panel C of the table. Comparing the two panels reveals that aggregation does not affect the general pattern of results. The magnitudes of the coefficients do change a bit, but the differences are mostly insignificant.

As an additional test, we repeat estimation of the hedonic models from panels A and B of table 3 after aggregating the micro data into 0.2 mile zones around each boundary. This approach aggregates every variable to the geographic level of resolution at which we observe variation in test scores near a boundary. Results are reported in panels C and D of table A2. Panels A and B are the same as table 3. Comparing panel A with panel C and comparing panel B with panel D illustrates that aggregation does not affect the general

pattern of results. As before, the differences that exist are mostly insignificant. Taken together, the results from tables A1 and A2 suggest that aggregation bias is not driving the large differences we observe between hedonic and capitalization-based estimates for price function parameters.

TABLE A1
ROBUSTNESS CHECKS ON TEST SCORE COEFFICIENTS

	FAIRFAX, VA	PORTLAND, OR	PHILADELPHIA, PA	DETROIT, MI	LOS ANGELES, CA
	A. Test Score Parameters from Hedonic Regressions (micro data from 0.2 mile boundary sample without boundary fixed effects)				
	(1)	(2)	(3)	(4)	(5)
log (test score), 2003 coefficient	0.122	0.456	0.481	0.524	0.274
	(0.027)	(0.020)	(0.045)	(0.036)	(0.012)
log (test score), 2007 differential	0.554	0.034	0.229	0.516	0.084
	(0.056)	(0.032)	(0.067)	(0.086)	(0.023)
R^2	0.74	0.70	0.68	0.68	0.75
Number of observations	6,036	14,443	3,973	6,252	12,287
	B. Test Score Parameters from Hedonic Regressions (micro data from full sample without boundary fixed effects)				
	(6)	(7)	(8)	(9)	(10)
log (test score), 2003 coefficient	0.227	0.540	0.546	0.751	0.260
	(0.023)	(0.016)	(0.017)	(0.019)	(0.004)
log (test score), 2007 differential	0.550	0.024	0.396	0.565	0.041
	(0.044)	(0.024)	(0.028)	(0.042)	(0.008)
R^2	0.75	0.69	0.72	0.68	0.70
Number of observations	10,662	25,294	29,327	32,485	146,783
	C. Test Score Parameters from Hedonic Regressions (block group data from full sample without boundary fixed effects)				
	(11)	(12)	(13)	(14)	(15)
log (test score), 2003 coefficient	0.148	0.388	0.229	0.813	0.321
	(0.068)	(0.045)	(0.046)	(0.052)	(0.012)
log (test score), 2007 differential	0.475	0.086	0.367	0.717	0.082
	(0.121)	(0.070)	(0.083)	(0.108)	(0.022)
R^2	0.84	0.82	0.78	0.77	0.73
Number of observations	889	1,553	2,647	3,333	14,727

NOTE.—All regressions use Eicker-White standard errors and include controls for property taxes, physical housing characteristics (square feet, number of bathrooms, age, lot size, number of bedrooms) and neighborhood characteristics measured at the block group level (population density, percent nonwhite, percent under 18, percent owner occupied, and percent vacant). In cols. 1 through 10, the dependent variable is the natural log of the sale price of the house. All control variables are interacted with a dummy for sales made during the 2007-2008 school year. In cols. 11 through 15 the dependent variable is the change in the natural log of the average sale price in a census block group.

TABLE A2
ADDITIONAL ROBUSTNESS CHECKS FOR AGGREGATION BIAS

	FAIRFAX, VA	PORTLAND, OR	PHILADELPHIA, PA	DETROIT, MI	LOS ANGELES, CA
	A. Test Score Parameters from Hedonic Regressions (micro data from 0.2 mile boundary sample without boundary fixed effects)				
	(1)	(2)	(3)	(4)	(5)
log (test score), 2003 coefficient	0.122	0.456	0.481	0.524	0.274
	(0.027)	(0.020)	(0.045)	(0.036)	(0.012)
log (test score), 2007 differential	0.554	0.034	0.229	0.516	0.084
	(0.056)	(0.032)	(0.067)	(0.086)	(0.023)
R^2	0.74	0.70	0.68	0.68	0.75
Number of observations	6,036	14,443	3,973	6,252	12,287
	B. Test Score Parameters from Hedonic Regressions (micro data from 0.2 mile boundary sample with boundary fixed effects)				
	(6)	(7)	(8)	(9)	(10)
log (test score), 2003 coefficient	0.116	0.200	0.272	0.208	0.140
	(0.040)	(0.028)	(0.071)	(0.047)	(0.015)
log (test score), 2007 differential	0.293	-0.165	-0.120	0.357	0.075
	(0.081)	(0.048)	(0.101)	(0.126)	(0.028)
R^2	0.85	0.77	0.76	0.74	0.85
Number of observations	6,036	14,443	3,973	6,252	12,287
	C. Test Score Parameters from Hedonic Regressions (boundary zone data from 0.2 mile sample without boundary fixed effects)				
	(11)	(12)	(13)	(14)	(15)
log (test score), 2003 coefficient	-0.009	0.348	0.540	0.482	0.243
	(0.051)	(0.050)	(0.189)	(0.121)	(0.065)
log (test score), 2007 differential	0.539	0.182	0.012	0.448	0.126
	(0.116)	(0.077)	(0.234)	(0.249)	(0.098)
R^2	0.786	0.801	0.712	0.764	0.816
Number of observations	887	1,308	393	491	526
	D. Test Score Parameters from Hedonic Regressions (boundary zone data from 0.2 mile sample with boundary fixed effects)				
	(16)	(17)	(18)	(19)	(20)
log (test score), 2003 coefficient	-0.041	-0.004	0.359	0.143	0.105
	(0.068)	(0.047)	(0.221)	(0.116)	(0.037)
log (test score), 2007 differential	0.352	0.065	-0.239	0.394	0.040
	(0.099)	(0.058)	(0.170)	(0.178)	(0.049)
R^2	0.923	0.928	0.893	0.918	0.961
Number of observations	887	1,308	393	491	526

NOTE.—All regressions use Eicker-White standard errors and include controls for property taxes, physical housing characteristics (square feet, number of bathrooms, age, lot size, number of bedrooms) and neighborhood characteristics measured at the block group level (population density, percent nonwhite, percent under 18, percent owner occupied, and percent vacant). In cols. 1 through 10, the dependent variable is the natural log of the sale price of the house. All control variables are interacted with a dummy for sales made during the 2007-2008 school year. In cols. 11 through 20 the dependent variable is the change in the natural log of the average sale price in a 0.2 mile zone adjacent to an attendance zone boundary or district boundary.

III. *Summary Statistics for Portland, Philadelphia, Detroit and Los Angeles*

Summary statistics for our Virginia sample are reported in the main text. Tables A3 through A6 report the corresponding summary statistics for Portland, Philadelphia, Detroit, and Los Angeles.

TABLE A3

SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN PORTLAND, OR

	Full Sample (micro data: N = 25,294)		Sample: 0.20 Mile Boundary Zone (micro data: N = 16,539)			Full Sample (Census block group data: N = 754)			
Portland Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									0.00
2003 price	241,875	142,991	237,021	2,664	0.91	244,315	127,426	0.00	
2007 price	324,181	173,171	316,236	-4,500	-1.14	336,280	156,920		
Average math/reading test result									
2003 score	79.82	10.93	79.89	7.41	44.93	78.94	10.33	-0.28	
2007 score	76.00	10.83	75.96	4.79	28.77	75.24	10.69		
Housing characteristics:									
square feet (100's)	17.88	7.76	17.75	-0.06	-0.48	17.59	5.18	0.04	-0.06
bathrooms	2.22	0.93	2.22	0.04	2.41	2.09	0.59	0.01	-0.03
age	39.67	30.16	38.39	-0.56	-1.17	50.17	22.91	0.00	-0.01
lot acres	0.20	0.29	0.18	0.00	0.72	0.30	0.43	0.08	0.01
bedrooms	3.07	0.94	3.07	-0.01	-0.41	3.03	0.47	0.09	-0.12
Neighborhood characteristics:									
% block group nonwhite	0.17	0.10	0.17	-0.01	-3.64	0.16	0.11	-0.14	0.00
% block group under 18	0.23	0.04	0.24	0.00	-1.14	0.22	0.04	-0.03	-0.02
% block group owner occupied	0.66	0.19	0.66	0.00	1.60	0.61	0.22	0.03	-0.06
% block group vacant	0.05	0.03	0.05	0.00	-7.41	0.05	0.03	-0.05	0.00
block group pop density	0.53	0.29	0.55	-0.01	-1.35	0.56	0.34	-0.08	0.04
tax rate	54.62	8.02	54.74	-0.23	-1.86	54.70	9.01	0.13	

NOTE.—This table reports summary statistics for the key variables included in the analysis for Portland, OR. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score houses on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

TABLE A4

SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN PHILADELPHIA, PA

	Full Sample (micro data: N = 29,333)		Sample: 0.20 Mile Boundary Zone (micro data: N = 3,973)			Full Sample (Census block group data: N = 1,199)			
Philadelphia Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									-0.07
2003 price	295,845	188,924	285,243	32,498	4.73	273,104	148,829	-0.35	
2007 price	334,662	221,967	324,197	49,222	5.44	316,940	170,204		
Average math/reading test result									
2003 score	67.88	13.93	69.43	11.05	30.63	64.38	16.84	-0.74	
2007 score	78.61	10.90	79.51	7.20	25.67	75.99	13.23		
Housing characteristics:									
square feet (100's)	20.87	9.48	20.03	1.21	4.21	19.85	5.95	-0.27	0.00
bathrooms	2.37	1.00	2.28	0.09	3.02	2.15	0.73	-0.35	-0.02
age	42.03	27.85	46.32	3.50	4.23	49.54	21.16	0.03	0.00
lot acres	0.49	0.65	0.44	0.02	1.15	0.45	0.46	-0.15	-0.01
bedrooms	3.38	0.77	3.33	0.07	2.82	3.40	0.44	-0.04	-0.03
Neighborhood characteristics:									
% block group nonwhite	0.12	0.14	0.11	-0.01	-1.30	0.14	0.19	0.22	-0.20
% block group under 18	0.23	0.04	0.22	0.00	2.54	0.22	0.04	0.01	0.10
% block group owner occupied	0.78	0.18	0.79	0.02	3.13	0.74	0.21	-0.12	0.04
% block group vacant	0.03	0.03	0.03	0.00	-3.77	0.04	0.03	0.14	-0.02
block group pop density	0.34	0.39	0.36	-0.04	-3.83	0.46	0.53	0.28	-0.13
tax rate	29.05	14.28	28.38	2.74	6.55	25.47	14.65	-0.30	

NOTE.—This table reports summary statistics for the key variables included in the analysis for Philadelphia, PA. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score houses on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

TABLE A5

SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN DETROIT, MI

	Full Sample (micro data: N =32,486)		Sample: 0.20 Mile Boundary Zone (micro data: N = 6,285)			Full Sample (Census block group data: N = 1,477)			
Detroit Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									0.14
2003 price	219,857	131,658	214,048	14,186	2.92	214,626	123,303	-0.38	
2007 price	166,801	131,839	157,640	11,017	2.44	169,829	104,428		
Average math/reading test result									
2003 score	68.76	12.39	67.91	7.77	27.21	68.01	12.18	-0.60	
2007 score	79.28	10.52	78.51	7.72	32.09	78.80	10.38		
Housing characteristics:									
square feet (100's)	16.57	7.79	16.01	0.66	3.27	16.47	5.93	-0.32	0.02
bathrooms	2.06	1.00	2.00	0.08	3.12	2.03	0.73	-0.38	0.08
age	46.06	23.24	46.78	0.19	0.36	46.87	18.00	0.26	-0.07
lot acres	0.36	0.52	0.30	-0.02	-2.06	0.39	0.46	-0.11	0.01
bedrooms	3.15	0.73	3.11	0.03	1.66	3.15	0.45	-0.22	0.00
Neighborhood characteristics:									
% block group nonwhite	0.13	0.18	0.12	-0.03	-7.53	0.14	0.20	0.04	-0.30
% block group under 18	0.23	0.04	0.23	0.01	8.67	0.23	0.04	0.04	0.16
% block group owner occupied	0.80	0.18	0.82	0.02	3.79	0.78	0.20	-0.15	-0.01
% block group vacant	0.04	0.03	0.03	0.00	0.80	0.04	0.04	0.12	0.02
block group pop density	0.40	0.28	0.46	0.01	2.00	0.40	0.28	0.25	-0.07
tax rate	27.09	11.25	25.90	-0.61	-2.42	27.70	9.73	-0.11	

NOTE.— This table reports summary statistics for the key variables included in the analysis for Detroit, MI. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score houses on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

TABLE A6

SUMMARY STATISTICS FOR HOUSING, NEIGHBORHOODS, AND PUBLIC SCHOOLS IN LOS ANGELES, CA

	Full Sample (micro data: N =146,788)		Sample: 0.20 Mile Boundary Zone (micro data: N = 12,287)			Full Sample (Census block group data: N = 6,975)			
Los Angeles Metro Area	mean (1)	standard deviation (2)	mean (3)	difference in means: high score side -low score side (4)	T-statistic on difference in means (5)	mean (6)	standard deviation (7)	correlation: Δscore & variable in 2003 (8)	correlation: Δscore & Δvariable (9)
Sale price									-0.04
2003 price	460,747	353,758	509,207	38,511	4.03	486,539	350,354	-0.15	
2007 price	486,752	463,063	563,566	55,621	3.72	551,685	458,579		
Average math/reading test result									
2003 score	39.75	13.75	41.86	13.92	49.94	39.11	14.14	-0.46	
2007 score	48.81	12.92	51.14	12.81	50.50	48.39	13.22		
Housing characteristics:									
square feet (100's)	17.06	7.67	17.11	0.84	5.81	16.16	5.63	-0.19	0.02
bathrooms	2.13	0.86	2.16	0.13	7.89	1.97	0.61	-0.20	0.02
age	43.06	22.98	44.87	-3.49	-8.98	53.11	19.18	0.21	0.00
lot acres	0.25	0.38	0.20	0.00	0.24	0.22	0.28	-0.13	0.01
bedrooms	3.18	0.87	3.21	0.07	4.66	3.07	0.54	-0.14	0.00
Neighborhood characteristics:									
% block group nonwhite	0.26	0.18	0.31	0.01	3.68	0.28	0.18	0.14	-0.25
% block group under 18	0.25	0.06	0.25	0.00	4.27	0.25	0.06	0.09	0.09
% block group owner occupied	0.68	0.21	0.70	0.01	3.99	0.60	0.25	-0.09	-0.01
% block group vacant	0.06	0.10	0.05	0.00	2.47	0.05	0.07	-0.06	0.00
block group pop density	0.72	0.60	0.82	-0.03	-2.74	0.96	0.75	0.17	0.05
tax rate	84.33	137.66	82.44	1.11	5.61	83.00	16.05	-0.04	

NOTE.— This table reports summary statistics for the key variables included in the analysis for Los Angeles, CA. Cols. 1, 2, 3, 6 and 7 are simply the means and standard deviations for the 3 different samples of data. The boundary zone sample includes all houses located within 0.20 miles of the boundary of another school attendance zone. Col. 4 reports the difference in means between houses located on the "high" test score side of a boundary with the corresponding mean for the "low" test score houses on the opposite side of the boundary. Col. 5 provides a T-statistic on the difference in these means. Cols. 8 and 9 report correlations between the change in test scores and levels and changes in all other variables for the full sample of census block group data.

IV. *Data and Code for Replication*

All data were obtained from the public sources described in the main text, with the exception of micro data on housing transactions, which we purchased from DataQuick (www.dataquick.com). These data can be purchased by contacting DataQuick at 1.888.299.8787. All other data and code needed to replicate our results are posted in the journal's online repository of supplementary material.