A Comparison of Block Group and Census Tract Data in a Hedonic Housing Price Model
Author(s): Allen C. Goodman
Source: *Land Economics*, Vol. 53, No. 4 (Nov., 1977), pp. 483-487
Published by: University of Wisconsin Press
Stable URL: http://www.jstor.org/stable/3145991
Accessed: 11-05-2018 15:20 UTC

SHORT PAPERS

A Comparison of Block Group and Census Tract Data in a Hedonic Housing Price Model

Allen C. Goodman

The consideration of neighborhood effects in urban and regional models has been constrained by data inadequacies. Analysts have customarily used data aggregated at the census tract level to characterize areas differentiated by public service provision or socioeconomic composition. This paper introduces use of the "block group," an aggregation unit typically 20 to 25 percent of the size of a census tract, formulated for general use in the 1970 censuses for all urbanized areas.[1]

Following a brief conceptual treatment of neighborhood definition, the measurement unit of the block group is considered. Its design and purpose are discussed and its empirical properties are tested. It is found to increase the accuracy with which the heterogeneity of neighborhoods within census tracts can be described. When applied to hedonic price models of urban housing, it is shown to improve the explanatory power of a regression as a whole and, in particular, the significance of racial variables, which appear to be extremely sensitive to the areal nature of the neighborhood specified.

The proper characterization of a neighborhood in urban analysis has been elusive, principally because the concept of a neighborhood is used to introduce spatial differentiation among goods such as land or housing that are often treated as homogeneous. Since a neighborhood represents an external influence on a physical bundle, it should be defined as a small urban area within which the residents receive or perceive a common set of socioeconomic effects and neighborhood services.[2] In particular, such an areal aggregation should be large enough to summarize the neighborhood effects that are common to small groups of residents, yet small enough to distinguish significant differences in these effects among neighborhoods across metropolitan areas. Although it is apparent that these effects should be considered at an aggregation level larger than that of the

---

The author is Assistant Professor of Economics, Lawrence University, Appleton, Wisconsin. He is indebted to A. Thomas King and Samuel Korper for providing the data used in this analysis and to John Quigley, Guy Orcutt, Eric Hanushek, Corry Azzi and an anonymous referee for their comments and suggestions. Support by U.S. Department of Housing and Urban Development doctoral dissertation grant H-2304 is gratefully acknowledged. The views expressed do not represent those of HUD.

[1] One of the first delineations of spatial areas by level of public service is proposed by Tiebout [1956]. Socioeconomic characteristics are often considered in hedonic price models, surveyed by Ball [1973], who also notes the shortcomings of grouped data. The block group measure proposed to ease this problem is originally described by the Bureau of the Census [1971]. The data base is customarily referred to as the First Count A File and is also available as the Fifth Count Intermediate Enumeration District and Block Group File.

[2] Schelling [1972] notes that a resident can be conscious of various neighborhoods in which he may live, work, play or have financial interests. He refers to both natural boundaries and socioeconomic barriers that can serve to define neighborhoods.

individual household, determination of the proper level is an empirical matter, with the researcher generally limited by the availability of suitable data.

Census Bureau block data would define neighborhoods of one city block each. Since it seems clear that almost any perception of neighborhood should begin with a household's block face and the one facing it, the arbitrary separation imposed by block aggregations is probably unrealistic. Two additional problems are related to Bureau policy. As a block may contain from 20 to 60 households, samples at the 15 or 20 percent survey levels would be collected from as few as two or three households, leading to unreliable results. Even if such surveys were taken, suppression rules designed for withholding data to avoid disclosure of information for individual housing units would limit their availability.[3]

The tract and block group aggregations have been developed at higher levels of aggregation, in consideration of such difficulties. Both measures include listings from the 100, 20, 15 and 5 percent census questionnaires on race, age, sex, ethnicity, employment classification, transportation and housing, although the block group level statistics have not generally been released in summary form. In discussing the proper level of aggregation for summary indicators of socioeconomic variables, Census Bureau statisticians note that if the units are as large as tracts, the summarization might "tend to hide the diversity within the areas."[4]

Table 1 uses the determination ratio, R^2, to reveal the considerable within-tract variation in neighborhood components that is hidden by the use of tract measurements for the New Haven SMSA. The four socioeconomic variables are *BLACK* (percentage black population), *POOR* (percentage of families with income less than $5,000), *EDUC* (percentage of population over age 25 with 13 or more years of education) and *UNEM* (percentage of population over age 14 that is unemployed). For the SMSA as a whole, for example, as much as 54.7 percent ($100[1 - R^2]$) of the variation for *UNEM* is unexplained by tract measurements. The heterogeneity is even more pronounced within the central city. The R^2 measures are lower for all of the variables; in particular, 47.8 percent of the variation of *POOR* and 65.5 per-

TABLE 1
DETERMINATION RATIO OF CENSUS TRACT MEASURES AS INDICATORS OF SOCIOECONOMIC NEIGHBORHOOD COMPOSITION

Category	Mean (%)	R^2
New Haven SMSA (87 Tracts)		
BLACK	10.9	.831
POOR	14.0	.695
EDUC	24.8	.784
UNEM	3.5	.453
City of New Haven (28 Tracts)		
BLACK	26.3	.790
POOR	23.7	.522
EDUC	20.4	.769
UNEM	4.7	.345

Note: See Table 3 for definition of variables.

[3] Block data refers to the Bureau of the Census Third Count File of Urbanized Areas. The rules noted concern the Census Bureau's suppression of all information outside of simple population counts that could possibly be used to discern the earnings or house value, for example, of an individual or household.

[4] There are approximately 10 blocks per block group, and between one and eight block groups per census tract. Block groups usually contain between 200 and 600 households and between 500 and 1,600 persons, compared with the average census tract population of 4,000. For further information, see Bureau of the Census [1971, p. 12].

TABLE 2
THREE REPRESENTATIVE CENSUS TRACTS AND THEIR DIVISIONS INTO BLOCK GROUPS

Tract	Block Group	Pop.	*BLACK*	*EDUC*	*UNEM*	*POOR*
1418[a]		4715	26.3	53.8	3.3	20.3
	1	532	1.9	74.7	2.0	12.6
	2	1766	22.8	54.4	3.6	25.6
	3	842	1.7	82.1	3.4	17.7
	4	1575	51.7	27.9	3.6	18.2
1654[b]		4796	15.8	14.4	3.1	14.7
	1	1332	42.6	15.8	3.2	13.4
	2	712	20.2	14.5	3.4	14.2
	3	1137	0.0	9.4	5.7	13.7
	4	1615	2.8	16.9	0.6	16.7
1806[c]		7111	0.0	14.1	3.6	6.6
	1	2377	0.0	21.7	3.3	6.9
	2	561	0.0	11.7	2.0	6.6
	3	1273	0.0	7.8	2.0	7.1
	4	1755	0.0	7.6	5.6	7.0
	5	1145	0.0	16.5	3.7	4.5

Note: See Table 3 for definition of variables.
[a]New Haven.
[b]Hamden.
[c]East Haven.

cent of the variation of *UNEM* are unexplained by tract measurements.[5]

Examples of such information loss due to aggregation are shown in Table 2 for three selected census tracts within the SMSA. Tract 1418 displays substantial variation in racial and educational characteristics, with *BLACK* ranging from 1.7 to 51.7 and *EDUC* from 27.9 to 82.1. Tract 1656 exhibits a similar range in racial characteristics as well as a spread of 5.1 percentage points of unemployment. Examination of Tract 1806 indicates that a racially homogeneous area can show considerable range in other neighborhood dimensions such as educational level (7.6 to 21.7) and unemployment (2.0 to 5.6).

Another test of aggregation levels is derived from urban housing analysis. The application of hedonic price models to housing market behavior allows for the inclusion of neighborhood variables. In such a model,

$$P = f(B,N)$$

where P represents the price of a house, B represents a group of components of the physical bundle of the house and lot, and N represents a group of components that are pertinent to the house price, yet external to the physical bundle.[6] The hedonic price of any n_j of N is defined as

[5]The R^2 relates within-tract variation of block groups (as components of census tracts) to the total variation. An assumption of within-tract homogeneity would yield an R^2 of 1. This procedure is discussed by Suits [1963, chap. 5].

[6]These methods have often been associated with Griliches [1971]. For applications to housing markets, see the Ball [1973] survey.

TABLE 3
COMPARISON OF HEDONIC PRICE ESTIMATES WITH CENSUS TRACT AND BLOCK GROUP AGGREGATIONS*

Variable	Tract Aggregation	Block Group Aggregation	Variable	Tract Aggregation	Block Group Variable
SIZE	.00073	.00074	*FP*	7.216	6.814
	(14.48)	(14.00)		(3.704)	(3.530)
BRICK	34.30	34.50	*BLACK*	−.1185	−.2494
	(4.859)	(4.932)		(.8935)	(2.293)
HW	14.16	12.96	*POOR*	−.8011	−.4471
	(4.699)	(4.325)		(2.064)	(1.878)
GAR	11.11	11.07	*EDUC*	.9454	.9240
	(6.657)	(6.727)		(9.140)	(10.63)
AGE	−.8158	−.8147	*TIP*	.8439	−12.78
	(16.08)	(16.21)		(.1840)	(3.219)
RMS	5.931	6.144	*PCN*	14.87	15.47
	(5.051)	(5.285)		(5.348)	(5.888)
BATH	27.18	26.70	*Y68*	15.97	16.94
	(12.60)	(12.50)		(6.718)	(7.197)
LAV	9.456	8.013	*Y69*	39.15	40.00
	(4.137)	(3.528)		(12.70)	(13.10)
SPACE	.05837	.05396	Constant	13.88	61.05
	(11.52)	(10.99)	*N*	1835	1835
LDIS	−41.21	−42.59	R^2	.8353	.8388
	(8.236)	(8.864)	Std. Error	45.76	45.26
SPDIS	19.40	21.10			
	(6.651)	(7.585)			

Notes: SIZE: Lot size in square feet. *BRICK:* "1" if house is all brick, "0" otherwise. *HW:* "1" if hardwood floors, "0" otherwise. *GAR:* Number of covered garage spaces. *AGE:* Age of house in years. *RMS:* Number of rooms excluding bathrooms, lavatories. *BATH:* Number of full bathrooms. *LAV:* Number of lavatories. *SPACE:* Indoor living space in square feet. *LDIS:* Logarithm of distance in miles from central business district. *SPDIS: SPACE* multiplied by *LDIS. FP:* Number of fireplaces. *BLACK:* Percentage black population. *POOR:* Percentage families with income less than $5,000. *EDUC:* Percentage of population over 25 with 13 or more years of education. *TIP:* "1" if *BLACK* is greater than 5% and less than 15%, "0" otherwise. *PCN:* Principal components measure of neighborhood attitudes. *Y68*: "1" if house was sold in 1968, "0" otherwise. *Y69*: "1" if house was sold in 1969, "0" otherwise. (Both *Y68* and *Y69* equal "0" if house sold in 1967.)
*t-statistics in parentheses ($t_{.05}$ = 1.645). Dependent Variable: Price (in hundreds of dollars).

$\partial P/\partial n_j$ (similarly, $\partial P/\partial b_i$ for any b_i of B). Often specified in linear form, the set of hedonic prices is not necessarily one of long-run equilibrium supply prices, but rather of market prices that reflect the composition and location of existing stocks of residential capital and neighborhood components.[7]

In estimating such a relationship, the proper areal aggregation for neighborhood components is vital. Consider, for example, two physically identical houses in the same census tract, selling for different prices. If the tract is homogeneous with respect to socioeconomic variables, this difference is related either to market

[7] Kain and Quigley [1975] consider a model in which a coefficient for a given component in a linear hedonic price model is considered as the sum of its production cost and a quasi-rent reflecting the degree of competition in the market.

noise or to omitted variables in the model. If, however, the tract varies in socioeconomic composition, this price differential might be explained by within-tract variance in these neighborhood components. If block group measurements are a better specification of neighborhood effects, then both the R^2 for the estimates and the significance levels for the neighborhood coefficients should be improved.

As Table 3 indicates, the R^2 rises from .8353 to .8388, and the significance levels for all neighborhood variables but *POOR* rise. Of special note is the behavior of the racial variables, positive and insignificant for the tract data. They assume their expected signs when block group data is used. Aggregation at too high a level appears, as such, to mask behavior related to racial differences.[8]

This paper has examined a set of data that is available at a level of aggregation much smaller than that previously available. The block group appears to be particularly useful in cases where researchers feel that census tract aggregations hide the considerable within-tract heterogeneity that exists among neighborhoods within urban areas. Tests on this data set for both descriptive and analytical uses verify that it can contribute additional explanatory power when employed to measure such neighborhood variables.

[8] This would tend to attribute some of the racial effects to variables with which *BLACK* and *TIP* are correlated—in particular, *POOR*; hence an explanation of the fall in magnitude and significance of that variable.

References

Ball, Michael J. 1973. "Recent Empirical Work on the Determinants of Relative House Prices." *Urban Studies* 10 (June): 213–31.

Griliches, Zvi, ed. 1971. *Price Indexes and Quality Change.* Cambridge: Harvard University Press.

Kain, John F., and Quigley, John M. 1975. *Housing Markets and Racial Discrimination.* New York: Columbia University Press.

Schelling, Thomas C. 1972. "The Process of Residential Segregation: Neighborhood Tipping." In *Racial Discrimination in Economic Life,* ed. Anthony H. Pascal. Lexington, Mass.: D.C. Heath and Company.

Suits, Daniel B. 1963. *Statistics: An Introduction to Quantitative Research.* Chicago: Rand McNally.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (Oct.): 416–24.

U.S. Bureau of the Census 1971. *Census Use Study: Health Information System II.* Report no. 12, Washington, D.C.