Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Urban Economics 54 (2003) 474–498

JOURNAL OF Urban Economics

www.elsevier.com/locate/jue

Understanding gentrification: an empirical analysis of the determinants of urban housing renovation

Andrew C. Helms

Department of Economics, University of Georgia, Athens, GA 30602, USA

Received 19 September 2000; revised 12 June 2003

Abstract

The "back-to-the-city" phenomenon presented an unpredicted countercurrent in the prevalent tide of suburbanization, and this process of upper-income resettlement in the inner city has been thoroughly analyzed in the urban economic literature. Housing renovation, a process that always accompanies gentrification and constitutes a significant portion of residential housing investment, has been studied much less. Contrary to the expectation that "location matters," the existing empirical studies have concluded that most neighborhood amenities and structural attributes are insignificant as determinants of renovation. Using a detailed parcel-level data set that documents all residential renovation activity in Chicago between 1995 and 2000, this paper establishes that the characteristics of a building and its neighborhood do indeed influence the likelihood that it will be renovated.

1. Introduction

Surprisingly and encouragingly, recent data from the 2000 Census [32] have revealed that after a half-century of continual population loss, the cities of New York and Chicago gained residents between 1990 and 2000. Some demographers dismally maintain that the scales were tipped by high birth rates (especially among inner-city immigrants), not by the widespread adoption of pro-urbanist preferences among consumers. At best, the population growth is the result of a gradual but steady shift in residential behavior patterns.

A countercurrent in the tide of suburbanization was first detected in the late 1960s: some inner-city neighborhoods were unexpectedly being resettled by middle- and upper-income

E-mail address: ahelms@terry.uga.edu.

0094-1190/$ – see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/S0094-1190(03)00081-0

"pioneers," who were typically young, childless, and well educated.[1] Gentrification, as the phenomenon was dubbed, garnered breathless media coverage, attracted academic attention, and raised the hopes of city governments. Though gentrification did not herald the end of suburbanization, neither was it a transitory trend. It has steadily persisted, if not gathered momentum, over the past three decades. During this time, gentrification has revealed itself to be less often a one-way migration back to the city than a continual circulation through the city: as one demographer straightforwardly explained (about Chicago), "You've got all these 20-year-olds coming in, and all these 30-year-olds going out."[2] In addition, gentrifiers include the so-called "empty-nesters," who return to the city and stay throughout the second childless phase of their lives.

Even though the city often loses the younger cohort of (re)settlers to the suburbs after they start families, it retains the physical improvements that they made to their residences, and also benefits from the upgrading investments of the returning empty-nesters. Housing rehabilitation, which is certainly the most visible evidence of gentrification, improves the city's physical health by forestalling further decay of the housing stock and improves its fiscal health by boosting the property tax base. The sheer volume of expenditures on residential improvements is notable: in the year 2000, when US households spent $160 billion on construction of new single-family homes, owner-occupiers of existing single-family homes spent *more than half* that amount ($81 billion) on home improvements, not including routine maintenance and repair.[3] In cities, the ratio is even more striking: in Chicago between 1995 and 2000 (the city and time period analyzed in this paper), investment in new construction and investment in the improvement of existing housing were *nearly equal*.[4]

Of course, not all inner-city renovation activity is gentrification-based; much of it is performed by existing city residents.[5] This "incumbent upgrading" is a relatively predictable and continual occurrence in historically stable areas. While its effects are certainly not negligible, they are usually gradual and often nearly invisible. By definition, incumbent upgrading does not significantly alter the demographic or socioeconomic composition of a neighborhood. Consequently, it does not dramatically change neighborhoods, let alone catalyze city-wide revitalization. Though gentrification is also unlikely to singlehandedly revitalize America's inner cities, it does markedly transform neighborhoods, both phys-

[1] The preferences and demographics of the "first-generation" gentrifiers are thoroughly documented by Kern [17] and by many sociologists (e.g., Clay [11] and Gale [13]). It is generally agreed that the personal characteristics of their current counterparts are quite similar.

[2] Ken Johnson in P. Reardon, "Floating in Data and Loving it," *Chicago Tribune*, Nov. 8, 2001.

[3] The Census Bureau separates "residential improvements and repairs" into two categories: "maintenance and repairs," and "improvements." For the purposes of this paper, the terms *renovation, rehabilitation*, and *alteration* are considered synonymous with the latter category [31].

[4] The valuation/construction cost of new construction between 1/1/96 and 12/31/00 was $1.84 billion (US Census Bureau [31]); expenditures on renovations, excluding additions, between 10/19/95 and 10/25/00 were $1.75 billion (Chicago Department of Buildings [8]).

[5] As a very rough approximation, about 56% of the renovations performed in Chicago between 1995 and 2000 (the sample analyzed by this paper) was "incumbent upgrading": on average, renovation activity occurred in neighborhoods in which 44% of the residents were recent in-movers, according to block-level Census data from 1990 [30].

ically and demographically (with some side-effects, as sociologists have noted).[6] As a result, the housing renovation that accompanies gentrification is a process that is important to understand. However, there exist only a few empirical studies of residential renovation, and none of them provides a rigorous and conclusive answer to the central questions: What exactly are the determinants of urban housing renovation? Which local amenities and structural characteristics attract renovators to certain neighborhoods?

Some sociologists have hypothesized answers to these questions in their case studies of gentrified neighborhoods.[7] From this literature, the mainstream press, and even casual observation, the common characteristics of gentrified areas are easily identifiable. Most of the neighborhoods consist of historic, low-density, architecturally distinctive houses, and they frequently feature parks and pleasant views. They are usually quite proximate to the central business district (CBD) and convenient to mass transit, and they are almost always far away from highways, public housing projects, and other disamenities. The houses in neighborhoods like these can intuitively be expected to experience a high level of renovation activity. A neighborhood's demographic characteristics (such as racial composition, average income, age distribution, and ethnicity) are also likely to affect gentrification and renovation activity, but the exact nature and extent of their influence are difficult to conclusively determine from anecdotal analyses.

Most existing empirical studies of renovation either fail to adequately account for the idiosyncratic attributes of individual buildings and neighborhoods, or find that these attributes are statistically insignificant predictors. Two exceptions are the studies by Mayer [20] and Melchert and Naroff [22]. Mayer analyzes renovation activity among rental properties in Berkeley, California. The results of his logit regression confirm the expected effects of buildings' characteristics: older, smaller, owner-occupied units that were structurally sound (but not necessarily good-looking) and had not been recently renovated were the most likely to be rehabilitated.[8] However, the effects of many of the neighborhood characteristics—including noise and traffic levels, non-residential land uses, population density, and distance from the university campus (Berkeley's counterpart of a CBD)—are statistically insignificant. Melchert and Naroff use a wide variety of block-level Census data to characterize the buildings and neighborhoods in their study of neighborhood gentrification in Boston. Of the 34 explanatory variables that they consider, five—distance to downtown, proximity to a small or medium-sized open area, pre-1900 construction, and average rent in each block—have statistically significant coefficients in their final regression.[9]

[6] The essays collected by Laska and Spain [18] discuss a variety of these issues. One of the most serious and commonly-cited side effects, the displacement of lower-income residents, is thoroughly analyzed by Nelson [27].

[7] See Clay [11] and Laska and Spain [18].

[8] Mayer's results establish that these characteristics influence landlords' decisions to renovate rental housing, but there is no evidence (nor claim) that they similarly influence homeowners' renovation decisions. However, the positive coefficient of owner-occupancy suggests that non-absentee landlords might perform more renovations because they "can tailor improvements on their own dwelling units to their own tastes" (p. 85). If this hypothesis is correct, then the results of Mayer's study—in which 92% of the buildings in the sample were owner-occupied—are indeed relevant.

[9] Melchert and Naroff's dependent variable is an appraisal by the Boston Redevelopment Authority of whether or not each block had experienced gentrification. While this approach has the advantage of separating

From the rest of the empirical literature, only one result consistently emerges: the likelihood of renovation increases with a building's age. Mendelsohn's [23] study, one of the earliest empirical examinations of renovation, includes no building or neighborhood characteristics other than age. Shear [28] uses American Housing Survey (AHS) data to examine a household's decision to move or to stay and renovate. Of the building attributes that he includes as explanatory variables, only age and a dummy variable for structural soundness are significant. Montgomery [25] also analyzes the move vs. renovate decision using AHS data. Of the seven dwelling and neighborhood characteristics that she includes, building age is the only variable that has a statistically significant influence on the likelihood that a household will improve its property.[10] AHS data are also used by Bogdon [6], who focuses on a homeowner's decision to hire outside help for renovations. In her regressions, age and square footage are the only significant building attributes, and none of the neighborhood attributes is significant. Galster [14] gathers and analyzes detailed survey data on housing "upkeep." In his results, the significant predictors include many of the homeowners' characteristics but only one of the building and neighborhood characteristics (again, building age). Likewise, Chinloy [10] includes 14 housing characteristics in his estimation of maintenance expenditures (excluding improvements), but only building age turns out to be a significant explanatory variable.

In spite of these results, this paper rejects the apparent conclusion that all structural characteristics, neighborhood attributes, and local amenities except building age are insignificant determinants of residential renovation. Empirically, the effects of 27 of these variables (most of which are measured at the parcel or block level) are definitively established by analyzing a set of microdata that documents all renovation activity among Chicago buildings over the years 1995 to 2000. The estimation results confirm that building age does indeed have a significant influence on improvement activity. Additionally, seven variables establish the effects of housing density and vacancy (both at the building and neighborhood level) on the likelihood of renovation. Accessibility to the CBD, measured by three variables, and three neighborhood (dis)amenities also have significant effects. Finally, most of the demographic variables, which describe the racial composition, average income level, and age distribution of each neighborhood, have significant coefficients as well.

In Section 2 of this paper, the theoretical literature that is relevant to gentrification and renovation is summarized, and a simple model of household renovation behavior is developed. In Section 3, the data are described and hypotheses about the variables are discussed. Empirical results are presented and reviewed in Section 4. Finally, Section 5 discusses the implications and conclusions of the analysis.

gentrification activity from incumbent upgrading, it relies upon a subjectively-determined dummy variable instead of "hard" data on actual consumer behavior.

[10] When the dependent variable includes routine maintenance in addition to improvements, four of the seven dwelling and neighborhood variables have significant coefficients. Six of the seven coefficients are significant when the dependent variable is the dollar amount of improvement expenditures (instead of a dummy variable indicating the occurrence of improvement activity).

2. Theory

2.1. Gentrification as a prediction of the urban model

Gentrification encompasses the two distinct processes of upper-income resettlement and housing renovation, which are usually modeled separately as independent phenomena. While this paper is concerned with the latter, studies of spatial income patterns can indirectly provide insight into the process of housing renovation. The models that explicitly include urban amenities are particularly relevant to this paper.

Early models of a monocentric city by Alonso [1], Mills [24], and Muth [26] predict a spatial equilibrium in which income increases with distance from the center. This outcome relies on the assumption that housing demand is more income-elastic than commuting costs. Wheaton [33] empirically tests this assumption and finds that the two income elasticities are very similar. Consequently, the bid-price functions are almost identical across income groups, making the model's income segregation predictions "statistically unreliable" (p. 631). This conclusion lends credence to the suspicion that urban spatial income patterns, including the upper-income resettlement component of gentrification, are strongly influenced by factors that are omitted from the simple urban model.

By focusing on changes in transport mode choice, LeRoy and Sonstelie [19] attempt to explain the spatial income patterns of three distinct phases in the life cycle of a city: "paradise," when the rich live downtown; "paradise lost," when the rich flee to the suburbs; and "paradise regained," when they resettle downtown. To capture the effects of transportation innovations, they extend the Alonso–Muth model to include a bimodal choice of transit. As income growth occurs and commuting costs vary, mode-switching may occur differentially across income groups (e.g., the rich adopting streetcars while the poor continue to walk to work). This switching can lead to location reversals and generate spatial equilibria that mirror all three phases above. However, LeRoy and Sonstelie's empirical results support only the first two phases: data from the early 20th century (before automobiles became widely affordable) uphold the model's predictions, but data from the 1950s–1970s yield inconclusive results. These results suggest that gentrification, unlike earlier shifts in residential location patterns, is not a simple consequence of transportation innovation.

By introducing locational amenities into the Alonso–Muth framework, the models of Berry and Bednarz [5] and Brueckner et al. [7] help explain upper-income resettlement. The equilibrium location pattern is determined not only by the relative income elasticities of housing demand and commuting costs, as in the standard model, but also by the slope of the amenity gradient and the rate at which consumers' marginal valuation of amenities rises with income. If the central city has a strong and growing amenity advantage over the suburbs and amenity valuation is highly income-elastic, then the rich will (re)locate downtown.

Kern [15,16] additionally assumes that some goods and services are obtainable only at the city center, in contrast to the standard composite consumption good that can be purchased anywhere in the city. To consume these goods, which represent urban amenities such as cultural, social, and entertainment activities, residents must make extra trips downtown in addition to their regular commutes. By also including a taste parameter to heterogenize preferences within the upper-income group, Kern's model generates a realistic separating spatial equilibrium in which the rich are stratified into the traditional suburbanites and the inner-city resettlers.

2.2. A renovation model

Most of the empirical work on housing renovation (reviewed earlier) is based on simple optimization models in which a homeowner or landlord chooses the level of capital investment to maximize some objective function. Mayer [20] presents a capital-stock adjustment model to study rental housing rehabilitation. Other authors extend this theoretical framework to examine specific elements of the renovation decision. Mendelsohn [23] and Bogdon [6] focus on the decision to hire outside help; Shear [28] and Montgomery [25] consider move decisions; and Chinloy [10] analyzes measurement issues regarding depreciation. Finally, Sweeney [29], Dildine and Massey [12], and Arnott et al. [3,4] apply optimal control methods to analyze the time path of maintenance and renovation. These theoretical models, while complex, are more realistic than a static optimization model, since housing maintenance and renovation are inherently dynamic processes.

To motivate the empirical work, consider a model similar to that of Mayer. Since this paper seeks to explain gentrification-based renovation instead of rental rehabilitation, the appropriate agent is not a profit-maximizing landlord but a utility-maximizing household.[11] As explained below, the household derives utility only from consumption and housing services. Though real-world owner-occupants invariably also take into account the asset value of their property when they make renovation decisions, this model makes the simplifying assumption that households' returns to their housing capital consist only of the utility they derive from consuming the housing services that their investment provides.[12]

First, let k_0 denote a building's initial (pre-renovation) level of housing capital, and let r denote the level of housing investment made during renovations. The post-investment capital level is therefore $k_0 + r$. Although negative values of r are not observable, the consumer choice problem is formulated to allow a negative r to be chosen. If such a choice is optimal, then the actual level of r will be zero.[13] This approach is useful in establishing the empirical framework, as seen below.

[11] The renovations in the data (described in the next section) are indistinguishably performed by both owner-occupiers and landlords, since information about the building owners/renovators was unavailable. Among the buildings in the data set, the average neighborhood owner-occupancy rate is 57.5% (US Census Bureau [31]).

[12] The model could be straightforwardly extended to incorporate a household's financial assets. First-order conditions would equate the marginal rates of substitution between improvements, consumption, *and* assets with the corresponding price ratios. Rates of return in the real estate and financial markets may affect the household's optimal "bundle," changing the *amount* of renovation expenditures but not affecting the *relative influences* of the structural and neighborhood characteristics on the likelihood of renovation activity, which are the concern of this paper.

[13] A negative value of r could presumably manifest itself as "deconstruction" (removal of housing capital), e.g., stripping architectural ornamentation or other valuable structural elements. However, this paper's model assumes that observations of r are nonnegative—see Eq. (4).

Assume that a building's condition (after any renovations have been made) is given by the function $c(b, k_0 + r)$, where b is a vector of its structural characteristics. These characteristics are the building's inherent attributes, such as its age and number of stories, which cannot be affected by any renovation work. The inclusion of these characteristics in the condition function c reflects the fact that the marginal return to housing investment can differ dramatically by building type. For any building, it is reasonable to assume that renovations always improve its condition, so that $c_r > 0$ (subscripts denote partial derivatives).

The total "housing services" $h(q, c(b, k_0 + r))$ provided by a dwelling are a function of its size q (measured by floor space) and condition c. Increased floor space or improved condition increases a building's housing service level, so that $h_q > 0$ and $h_c > 0$. Neighborhood attributes and amenities are not included in the housing services function, since houses of the same size and condition should provide the same level of housing services regardless of their locations. Instead, neighborhood characteristics, represented by the vector n, directly enter a household's utility function.[14] This specification also allows different households to have different preferences for neighborhood amenities. Finally, households derive utility from a numeraire composite consumption good z. The full utility function is therefore

$$u\big(h\big(q, c(b, k_0 + r)\big), n, z\big) \equiv v(s, r, n, z), \tag{1}$$

where s denotes the vector of structural characteristics (q, b).

Denoting income by y and the price of capital by p^K, a household's budget constraint is $y = z + p^K r$. Utility maximization over z and r then yields the first-order condition

$$v_r / v_z = p^K, \tag{2}$$

or $u_h h_c c_k / u_z = p^K$, which indicates that the marginal rate of substitution between home improvements and consumption must equal the price of capital. This equation implicitly defines the household's optimal consumption level z^* and housing capital investment r^*. The latter can be expressed as

$$r^* = r\big(s, n, p^K\big). \tag{3}$$

As explained above, whether a household will actually perform renovations depends upon the magnitude of r^*. If $r^* \leqslant 0$, none will be undertaken. Therefore, letting $\tilde{r}$ denote the household's renovation expenditure, it follows that

$$\tilde{r} = \begin{cases} r^* & \text{if } r^* > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Equation (4) can be used to motivate two different empirical models. The first is a tobit model, which explicitly accounts for the truncated nature of the dependent variable $\tilde{r}$. The second is a probit model, which does not make use of the actual level of renovation expenditures, but simply distinguishes between the cases in which $\tilde{r} > 0$ and $\tilde{r} = 0$.

14 Though n is not an argument in $h(.)$, note that the marginal utility of housing services u_h is likely to depend upon n.

For both models, optimal housing investment r^* is assumed to follow the relationship

$$r_i^* = \beta s_i + \gamma n_i + \varepsilon_i, \tag{5}$$

where s_i is the vector of building i's structural attributes and n_i is the vector of neighborhood characteristics.

In the probit model, the dummy variable a is defined to reflect the presence or absence of any renovation activity $\tilde{r}$ on the building:

$$a_i = \begin{cases} 1 & \text{if } \beta s_i + \gamma n_j + \varepsilon_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

3. Data and hypotheses

3.1. Dependent variables

The Chicago Department of Buildings requires building owners to obtain a building permit for all improvements except minor repairs. Applicants must describe the type and estimated cost of their proposed work.[15] This information was collected for all 65,536 permits that were issued between October 25, 1995 and October 19, 2000. Permits were discarded if they were issued for new construction, demolition, repairs ordered by the City, or unidentified repairs, or if the renovation cost was not recorded. If multiple permits were issued for the same building, the expenditures were summed. In the final data set, aggregate expenditures during this five-year period are $586 million, averaging $18,448 on each of the 31,774 buildings that were renovated.

From the renovation costs recorded in these permits, three dependent variables are defined. In the first tobit model, *RENBLDG* is the cumulative expenditure per building over the five-year period; in the second tobit, *RENUNIT* is the average expenditure per dwelling unit. In the probit model, the dummy variable *RENACT* indicates whether any renovation activity occurred at all. Summary statistics for these three dependent variables are presented in Table 1.

3.2. Explanatory variables

The Chicago Property Information Project maintains the “Harris File,” a database that provides information about the basic structural characteristics of every residential building in the city [9]. These characteristics include the number of dwelling units, the number of stories, vacancy status, and age. The 1995 raw data set includes 623,547 buildings at 599,043 street addresses. After deleting observations with missing or inconsistent values for important variables and observations with addresses that could not be geocoded or

15 Unfortunately, no personal characteristics of the applicant were recorded in the data set. Of particular significance is whether the applicant is an owner-occupier or a landlord. The theoretical model (presented in the previous section) assumes that the renovator is an owner-occupier, though the data also include rental properties. The rate of owner-occupancy in each block is included as an explanatory variable (*OWNEROCC*), as described later in this section.

Table 1
Statistical summary

Variable[a]	Tests for equality between groups		Full sample ($N = 435{,}534$)		Renovated buildings ($N_1 = 31{,}774$)		Unrenovated bldgs. ($N_2 = 403{,}760$)	
	F-stat.[b]	$\lvert t\text{-stat.}\rvert$[c]	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<u>*Renovation*</u>								
RENBLDG Total renov. expenditures	–	–	1346	19,099	18,448	68,446	0	0
RENUNIT Renov. exp. per dwl. unit			602	6049	8250	20,939	0	0
RENACT Renov. activity dummy	–	–	0.073	0.260	1	0	0	0
<u>*Building*</u>								
AGE[d] Age in year 1995	1.090	50.390	67.963	26.240	74.829	25.146	67.422	26.248
DWLUNITS # of dwelling units	2.578	9.544	2.632	12.187	3.566	18.527	2.559	11.538
STORIES # of stories	2.424	22.789	1.722	1.210	1.938	1.790	1.705	1.150
VACANT Vacancy dummy	1.812	6.658	0.004	0.064	0.007	0.084	0.004	0.063
<u>*Demographic*</u>								
BLACK % black, non-Hisp (blk)	1.072	29.860	0.349	0.448	0.423	0.462	0.343	0.446
OTHERMIN % min., not blk/Hisp (blk)	1.083	4.433	0.030	0.071	0.032	0.073	0.030	0.071
COLLEGE % coll. grads (22+) (bg)	1.286	13.272	0.142	0.130	0.152	0.146	0.141	0.129
FOREIGN % born outside USA (bg)	1.044	10.711	0.169	0.165	0.160	0.168	0.170	0.164
MFI Med. fam. inc. /1000 (bg)	1.361	16.494	35.035	13.679	33.646	15.746	35.145	13.497
KIDS % age < 18 (blk)[e]	1.091	12.894	0.250	0.091	0.257	0.095	0.249	0.091
YOUNG % age 18–34 (blk)[e]	1.198	24.893	0.281	0.081	0.293	0.088	0.280	0.080
OLD % age 65+ (blk)[e]	1.155	31.018	0.135	0.088	0.121	0.082	0.136	0.088
SINGLE % adults unmarried (blk)	1.403	18.961	0.058	0.096	0.070	0.112	0.057	0.094
<u>*Neighborhood*</u>								
DISTCBD Distance to CBD (mi)	1.009	34.021	1.504	0.569	1.400	0.565	1.512	0.568
ELTRAIN Dist. to el station (mi)	1.221	34.566	0.260	0.219	0.222	0.199	0.263	0.220
HIGHWAY Near hwy dummy (blk)	1.118	2.711	0.019	0.136	0.021	0.143	0.019	0.135
LAKE Near lake dummy	2.545	8.779	0.003	0.054	0.007	0.082	0.003	0.051

(*continued on next page*)

Table 1 (*Continued*)

Variable[a]	Tests for equality between groups		Full sample ($N = 435{,}534$)		Renovated buildings ($N_1 = 31{,}774$)		Unrenovated bldgs. ($N_2 = 403{,}760$)	
	F-stat.[b]	$\lvert t\text{-stat.}\rvert$[c]	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
MEDVALUE	1.627	8.128	85.135	60.225	88.386	75.118	84.879	58.886
Med. val. HUs /1000 (blk)								
NEIGHAGE	1.050	19.791	48.973	8.078	49.819	7.894	48.907	8.088
Med. age HUs in ’95 (bg)								
NODRIVE	1.132	46.650	0.305	0.134	0.341	0.142	0.302	0.133
% commute not by car (bg)								
OWNEROCC	1.024	41.890	0.575	0.290	0.508	0.293	0.580	0.290
% HUs own-occup. (blk)								
PARK	1.028	0.839	0.029	0.169	0.030	0.171	0.029	0.169
Adj. to park dummy (blk)								
PUBHSNG	1.047	38.087	0.809	0.463	0.716	0.452	0.816	0.463
Dist. from pub. hsng. (blk)								
UNIT1	1.025	37.241	0.492	0.363	0.420	0.358	0.498	0.363
% 1-unit HUs (blk)								
UNIT2	1.028	10.540	0.221	0.206	0.233	0.203	0.220	0.206
% 2-unit HUs (blk)								
UNIT3OR4	1.100	25.735	0.137	0.149	0.158	0.156	0.135	0.148
% 3- or 4-unit HUs (blk)								
VACANCY	1.377	33.261	0.065	0.074	0.080	0.085	0.063	0.072
% HUs vacant (blk)								

[a] (blk) denotes block-level; (bg) denotes blockgroup-level.

[b] To test H_0: $\sigma_X^2 = \sigma_Y^2$, $F = s_X^2/s_Y^2$. For every variable except *DISTCBD*, H_0 is rejected at the 99% level, indicating inequality of variances; for *DISTCBD*, the p-value of the test is 0.166, indicating equality of variances at any level above 83.4%.

[c] To test H_0: $\mu_X = \mu_Y$ for all variables except *DISTCBD*, $t = (\bar{x} - \bar{y})/(s_X^2/n_X + s_Y^2/n_Y)^{1/2}$ assumes unequal variances; for *DISTCBD*, $t = (\bar{x} - \bar{y})/(((n_X - 1)s_X^2 + (n_Y - 1)s_Y^2)/(n_X + n_Y - 2))^{1/2}(1/n_X + 1/n_Y)^{1/2}$ assumes equal variances. For every variable (including *DISTCBD*) except *PARK*, H_0 is rejected at the 99% level, indicating inequality of means.

[d] *AGE* is right-censored at 150, as described in footnote 18.

[e] *KIDS*, *YOUNG*, and *OLD* reflect the population percentages within these age brackets in 1990, as explained in footnote 18.

matched with the addresses in the renovation data set, the final sample consists of 435,534 buildings with unique addresses.

Using a geographic information system (GIS), the buildings’ addresses were geocoded, and several more variables were defined. The distance from each building to the central business district, to the nearest “El” (elevated commuter railroad) station, and to the nearest public housing project were calculated. Dummy variables identify whether a building is adjacent to a park, near an interstate highway, or near Lake Michigan. Finally, block- and blockgroup-level data from the 1990 Census were used to define an additional 17 variables that characterize the physical and demographic attributes of each neighborhood.

Table 1 provides a statistical summary of the variables in the final data set. To support the selection of explanatory variables, their variances and means are tested for equality between the subsamples of renovated and unrenovated buildings. For all but one of the 27 variables, the F-statistics indicate that the variances differ at the 99% level.[16] Accordingly, the t-tests for equality of the means assume that the variances are unequal. The resulting t-statistics demonstrate that the means are significantly different at the 99% level for all but one of the variables.[17]

The variable that differs most significantly between the two groups is *AGE*: renovated buildings are, on average, 7.4 years older than unrenovated buildings.[18] This finding is supported by the common conclusion of the other empirical studies (reviewed in Section 1) that renovation is more likely to be performed on older buildings. Since structures deteriorate over time, older buildings are more likely to require renovations, and the renovations are likely to be more costly, so a positive coefficient is expected for *AGE* in all of the models. Moreover, gentrification occurs almost exclusively in inner-city neighborhoods that feature historic, architecturally distinctive housing, so a positive coefficient is also expected for *NEIGHAGE* (the median age of housing in the blockgroup) in the probit model. Though renovation may be more likely in older areas, the magnitude of renovation expenditures is determined primarily by a building's own age (*AGE*), not the age of nearby housing, so the coefficient of *NEIGHAGE* should be insignificant in the tobit models.

Case studies (e.g., Clay [11]) have shown that gentrification favors lower-density structures in higher-density neighborhoods, so the effect of housing density on renovation activity differs depending upon the level at which it is measured. At the individual building level, density is measured by *STORIES* and *DWLUNITS* (the numbers of stories and dwelling units, respectively). If renovators prefer low-density structures with large apartments, then a negative coefficient would be expected for *STORIES* in the absence of the *DWLUNITS* variable. However, with *DWLUNITS* included as an explanatory variable, the coefficient of *STORIES* is expected to be positive. The reason is that, holding *DWLUNITS* constant, more stories imply larger, lower-density units. For the same reason, the coefficient of *DWLUNITS* is expected to be negative: holding *STORIES* constant, more dwelling units imply smaller, higher-density apartments.[19]

At the neighborhood level, density is measured by the variables *UNIT1*, *UNIT2*, and *UNIT3OR4*, which are the percentages of housing in each block with one, two, and three

[16] For every variable except *DISTCBD*, the test statistic $F = s_X^2/s_Y^2 > F^{-1}_{31774,403760}(0.99) = 1.0193$, so H_0: $\sigma_X^2 = \sigma_Y^2$ is rejected.

[17] For every variable except *DISTCBD*, the test statistic (assuming unequal variances) $t = (\bar{x} - \bar{y})/(s_X^2/n_X + s_Y^2/n_Y)^{1/2}$; H_0: $\mu_X = \mu_Y$ is rejected for every variable except *PARK*. For *DISTCBD*, the test statistic (assuming equal variances) $t = (\bar{x} - \bar{y})/(((n_X - 1)s_X^2 + (n_Y - 1)s_Y^2)/(n_X + n_Y - 2))^{1/2}(1/n_X + 1/n_Y)^{1/2} > t^{-1}_{435534}(0.99)$, so H_0: $\mu_X = \mu_Y$ is rejected.

[18] The actual difference is likely to be even greater. *AGE* is right-censored at 150, since the construction dates of very old buildings were evidently identically recorded as "before 1850" in the Harris File. The values of *AGE* and *NEIGHAGE* (from the 2000 Harris File and 1990 Census, respectively) have been adjusted so that the base year for both is 1995 (the year in which the earliest of the renovations occurred).

[19] Indeed, the summary statistics for *DWLUNITS* indicate that renovated buildings have (on average) more dwelling units than unrenovated buildings, but this higher observed mean does not contradict the *ceteris paribus* hypothesis that the coefficient of *DWLUNITS* is negative. This paper does not take into account an interesting and potentially influential idiosyncrasy of Cook County property tax law, which doubles the assessment of buildings with more than six units; presumably, this inhibits renovation of higher density buildings.

or four dwelling units, respectively. Accordingly, these variables' coefficients indicate the effects of lower densities relative to higher densities (five or more units). If renovators are attracted to central-city neighborhoods in which single-family homes and duplexes are less common, then the probit coefficients of *UNIT1* and *UNIT2* are expected to be negative. The probit coefficient of *UNIT3OR4* is expected to be positive, since moderate-density neighborhoods are likely to be the most desirable and therefore the most likely to experience frequent renovation activity. These three variables' coefficients are less likely to be significant in the tobit models, since renovation expenditures are influenced more by a building's own density (*DWLUNITS* and *STORIES*) than by the neighborhood's housing density.

Gentrification seems to be undeterred by housing vacancy, particularly during the early stages when renovators purchase and renovate abandoned buildings that are usually deteriorated but still structurally sound. Most rehabilitation of vacant buildings is probably not gentrification-based, but a positive coefficient is expected for *VACANT* (the dummy variable indicating whether the building is vacant) because *ceteris paribus*, unoccupied buildings tend to be in poorer condition, and because it is easier to undertake large-scale rehabilitation projects in the absence of residents. The coefficient of *VACANCY* (the vacancy rate in the block) is also expected to positive, but perhaps insignificant since abandoned neighborhoods are unlikely to attract much renovation activity.

Virtually every case study has shown that gentrified neighborhoods are likely to be proximate to the city's central business district. Clay [11] finds that most are within two miles of the CBD, half (49%) within one mile, and more than one third (38%) within one half mile. *Ceteris paribus*, sites that are closer to downtown workplaces and city-center amenities are more attractive than outlying neighborhoods. Therefore, the coefficient of *DISTCBD* (the distance in miles from the building to City Hall, at the heart of the Chicago "Loop") is expected to be negative. Chicago's elevated railroad system is an efficient and widely-used transit system. Since all nine branches of the six lines lead directly to the downtown Loop, a residence's proximity to one of the city's 134 "El" stations increases its accessibility to the CBD. Therefore, the coefficient of *ELTRAIN* (the distance in miles from the building to the nearest station) is expected to be negative. Additionally, a third variable, *NODRIVE* (the percentage of workers in each blockgroup who do not commute by car), is included to capture the proximity effect. Its coefficient is expected to be positive.[20]

Like the El train lines, four interstate highways radiate from the Chicago Loop. The convenience of living near a highway, however, may be outweighed by the accompanying noise, pollution, and traffic congestion. Particularly for inner-city residents, who are less likely to travel short distance in town by car, proximity to an interstate is a strong disamenity that is likely to discourage renovation activity. Therefore, the coefficient of the dummy variable *HIGHWAY* (which equals one for buildings located within one-tenth of a mile of an interstate) is expected to be negative.[21] Clay observes that housing is not the only

[20] Surprisingly, *NODRIVE* is only moderately correlated with *ELTRAIN* ($\rho = -0.376$) and *DISTCBD* ($\rho = -0.425$).

[21] A higher percentage of renovated buildings are located near a highway than unrenovated buildings (2.1% vs. 1.9%, respectively). Even though the difference is statistically significant, it is small in absolute value and counter-intuitive, so this hypothesized influence of *HIGHWAY* is maintained.

land use in nearly all (94%) of the gentrified neighborhoods in his study, but renovation is unlikely in areas with "nuisance uses" such as public housing. Proximity to one of Chicago's notorious housing projects is a disamenity that is likely to deter renovation, so the coefficient of *PUBHSNG* (the distance in miles from the nearest Chicago Housing Authority property) is expected to be positive.

Conversely, neighborhood amenities are likely to attract gentrification and stimulate renovation activity. Natural amenities, such as high elevations, attractive views, and proximity to water, are characteristic of many gentrified neighborhoods. Chicago's terrain (unlike, e.g., San Francisco's) is famously flat, so topographic elevation is not included as an explanatory variable. Lake Michigan is scenic and its shore (unlike, e.g., Manhattan's) has been well-developed for public recreation and pedestrian accessibility, so buildings near the lake are particularly desirable. Therefore, the coefficient of the dummy variable *LAKE* (which equals one for buildings located within one quarter mile of the lake shore) is expected to be positive. Man-made amenities that are attractive to renovators include parks, monuments, landmarks, and other atmospheric features that enhance a neighborhood's character. The dummy variable *PARK* (which equals one for buildings that face a city park) is expected to have a positive coefficient.[22]

The influence of *MEDVALUE* (the median value of owner-occupied houses in the block) is difficult to predict. It is unclear whether low property values stimulate or deter renovation activity: low-priced housing is more affordable and is usually more needy of renovation, but the returns to investments in improving such housing may be unpromising. Moreover, since each individual building's value is not included in the data, its influence cannot be separated from that of the median value in the neighborhood.

As noted earlier, the data do not distinguish between renovations performed by owner-occupiers and landlords, and the theoretical model developed in Section 2 considers the renovation decisions of owner-occupiers who are unconcerned with the property's rental value (real or imputed). However, it is reasonable to expect that the optimal level of capital of an absentee landlord's rental building is lower than that of an owner-occupier's house, since the landlord's marginal rent revenue from renovations is likely to be less than the homeowner's marginal utility. Mayer suggests several additional reasons,[23] and his empirical results support the expectation that the variable *OWNEROCC* (the percentage of housing units in each block that are owner-occupied) has a positive coefficient.

Gentrifiers often appreciate their neighborhoods' ethnic flavor and racial integration. Forty percent of the neighborhoods surveyed by Clay were considered to be ethnic communities before gentrification occurred; half were predominantly white and the other half, predominantly black. However, brisk renovation activity is unlikely in neighborhoods whose ethnic or racial composition is at either extreme. The same relationship probably also applies to income: gentrification occurs in stable moderate-income neighborhoods, not deeply impoverished areas or established upper-income enclaves. Furthermore, the

[22] As described in the next section and summarized in Table 1, *PARK* is the only explanatory variable whose mean does not differ significantly between the groups of renovated and unrenovated buildings; see Table 1.

[23] Landlords may be less likely to renovate because of uncertainty about renters' tastes and willingness to pay higher rents; uncertainty about tenants' stewardship of the property; or uncertainty about future expectations due to unfamiliarity with the property or neighborhood.

effects of neighborhood income are clouded by the indistinguishability of gentrification activity from incumbent upgrading: gentrification-based renovation tends to occur in low- and middle-income neighborhoods, but frequent and continual renovation activity is also likely to occur in well-maintained indigenous upper-income areas. For these reasons, it is difficult to predict the signs of the coefficients of four demographic variables: *BLACK* and *OTHERMIN* (the percentages of the population in each block who are black or non-black, non-Hispanic minorities, respectively); *FOREIGN* (the percentage of the population in each blockgroup born outside the United States); and *MFI* (median family income in each blockgroup). Since the average income level and racial composition of a neighborhood often dramatically change as gentrification progresses, these demographic data have been drawn from the 1990 Census, at least five years prior to any of the renovations.

Most studies characterize the prototypical gentrifier as young, unmarried, and well-educated. Accordingly, the coefficients of *YOUNG*, *SINGLE*, and *COLLEGE* (the percentages who are in the 18-to-34 age bracket, have never been married, and have graduated from college, respectively) are expected to be positive, and the coefficient of *OLD* (the percentage over 65) is expected to be negative. Since gentrifiers tend to be childless singles or "empty nesters," the last block-level age variable, *KIDS* (the percentage under 18), is also expected to have a negative coefficient, though the summary statistics do not support this prediction.[24]

4. Results

Table 2 presents the results of the probit estimations, and Tables 3 and 4 present the results of the tobit estimations with *RENBLDG* and *RENUNIT*, respectively, as the dependent variables. For both sets of models, the variables' marginal effects are reported in addition to their untransformed coefficients.[25] In Table 5, the signs and significance of the

[24] Since the Census data are from 1990 and the renovation data begin 1995, the population age variables *KIDS*, *YOUNG*, and *OLD* are lagged by five years, e.g., *OLD* is actually the percentage of block residents who were older than 70 in 1995 (and didn't die or move out of the neighborhood between 1990 and 1995).

[25] For the tobit results, three marginal effects are calculated: (1) the change in the unconditional expected value of the "latent" dependent variable (optimal renovation investment); (2) the change in the expected value of the dependent variable, conditional on the observation being uncensored (i.e., the change in expected renovation expenditure for renovated buildings); and (3) the change in the probability that an observation will be uncensored (i.e., that a building will be renovated). In Tables 3 and 4, they are labeled as Δ*OPTINV*, Δ*RENBLDG* or Δ*RENUNIT*, and $\Delta\Pr[\mathit{OPTINV} > 0]$, respectively. McDonald and Moffitt [21] demonstrate that these marginal effects are related as follows:

$$\underbrace{\frac{\partial E[y_i]}{\partial x_i}}_{\Delta \mathit{OPTINV}} = \Pr\big[y_i^* > 0\big]\underbrace{\left(\frac{\partial E[y_i^* \mid y_i^* > 0]}{\partial x_i}\right)}_{\substack{\Delta \mathit{RENBLDG}\\ \text{or}\\ \Delta \mathit{RENUNIT}}} + E\big[y_i^* \mid y_i^* > 0\big]\underbrace{\left(\frac{\partial \Pr[y_i^* > 0]}{\partial x_i}\right)}_{\Delta \Pr[\mathit{OPTINV} > 0]}.$$

The third tobit marginal effect ($\Delta\Pr[\mathit{OPTINV} > 0]$) and the probit marginal effect can both be interpreted in the same way; both have identical signs, but their values are not systematically related to one another.

Table 2
Estimation results of the probit models (dependent variable is *RENACT*[a])

Variable[b]		Probit I			Probit II[c]		
		Coefficient	Marg. effect[d]	$\|z\|$-value	Coefficient	Marg. effect[d]	$\|z\|$-value
Constant		−1.9655	–	−36.953***	−1.9944		−56.633***
Building							
AGE	(yrs)	0.0043	0.0006	28.110***	0.0043	0.0006	28.231***
DWLUNITS	(#)	−0.0013	−0.0002	4.100***	−0.0013	−0.0002	−4.138***
STORIES	(#)	0.0370	0.0049	11.730***	0.0373	0.0050	11.853***
VACANT	(0|1)	0.1040	0.0149	2.700***	0.1056	0.0151	2.743***
Demographic							
BLACK	(%)	0.2842	0.0377	18.620***	0.2972	0.0394	25.040***
OTHERMIN	(%)	0.1499	0.0199	3.240***	0.1387	0.0184	3.158***
COLLEGE	(%)	0.3070	0.0407	8.000***	0.3105	0.0412	8.288***
FOREIGN	(%)	−0.0310	−0.0041	1.000			
MFI	($K)	−0.0015	−0.0002	3.810***	−0.0015	−0.0002	−3.935***
KIDS	(%)	0.0337	0.0045	0.510			
YOUNG	(%)	−0.0798	−0.0106	1.270			
OLD	(%)	−0.1736	−0.0230	2.780***	−0.1440	−0.0191	−3.653***
SINGLE	(%)	0.0292	0.0039	0.520			
Neighborhood							
DISTCBD	(mi)	−0.1697	−0.0225	9.570***	−0.1716	−0.0228	−9.704***
ELTRAIN	(mi)	−0.0676	−0.0090	3.100***	−0.0639	−0.0085	−2.952***
HIGHWAY	(0|1)	0.0393	0.0054	1.910*	0.0391	0.0053	1.895**
LAKE	(0|1)	0.1562	0.0232	3.480***	0.1593	0.0237	3.562***
MEDVALUE	($K)	0.0004	0.0000	5.590***	0.0004	0.0000	5.562***
NEIGHAGE	(yrs)	0.0010	0.0001	2.200**	0.0009	0.0001	2.150**
NODRIVE	(%)	0.1981	0.0263	6.530***	0.1950	0.0259	6.548***
OWNEROCC	(%)	0.1241	0.0165	3.280***	0.1274	0.0169	3.471***
PARK	(0|1)	0.0301	0.0041	1.760			
PUBHSNG	(mi)	0.2399	0.0318	12.130***	0.2423	0.0322	12.304***
UNIT1	(%)	−0.1044	−0.0139	3.160***	−0.0986	−0.0131	−3.028***
UNIT2	(%)	−0.1993	−0.0265	8.730***	−0.1960	−0.0260	−8.694***
UNIT3OR4	(%)	0.1103	0.0146	4.020***	0.1099	0.0146	4.025***
VACANCY	(%)	0.3394	0.0450	7.400***	0.3477	0.0461	7.623***

[a] The dependent variable *RENACT* equals one if the building was renovated, zero otherwise.
[b] Units of measurement are in parentheses: (0|1) indicates a dummy variable; ($K) denotes thousands of dollars.
[c] Probit II includes variables from Probit I that are significant at the 95% level.
[d] The change in predicted probability when the explanatory variable is increased one unit from its mean, holding the values of all other variables constant at their means; e.g., a one-mile increase in distance from the CBD (*DISTCBD*) from its mean (i.e., from 1.5 to 2.5 miles) decreases the likelihood of renovation by 2.25%.
* Significant at 90% level.
** Significant at 95% level.
*** Significant at 99% level.

Table 3
Estimation results of the tobit models (dependent variable is *RENBLDG*[a])

Variable[b]		Tobit I		Tobit II[c]		Tobit II: marginal effects[d]		
		Coefficient	$\|t\text{-value}\|$	Coefficient	$\|t\text{-value}\|$	Δ*OPTINV*	Δ*RENBLDG*	$\Delta\Pr(OPTINV > 0)$
Constant		−161,931.10	36.202***	−157,104.90	48.257***			
Building								
AGE	(yrs)	320.98	25.322***	321.00	25.580***	23.45	49.30	0.000536
DWLUNITS	(#)	−47.58	2.095**	−47.26	82.091**	−3.44	−7.24	−0.000079
STORIES	(#)	4,116.13	17.193***	4,112.66	17.204***	300.04	630.97	0.006865
VACANT	(0|1)	21,032.75	6.965***	20,920.59	6.938***	1526.40	3209.97	0.034926
Demographic								
BLACK	(%)	20,273.62	15.853***	20,356.19	16.474***	1485.26	3123.46	0.033985
OTHERMIN	(%)	5,531.45	1.424					
COLLEGE	(%)	24,030.02	7.570***	22,663.09	8.766***	1653.44	3477.14	0.037833
FOREIGN	(%)	−6,176.37	2.373**	−4,655.40	1.918*	−339.39	−713.72	−0.007766
MFI	($K)	−18.92	0.605					
KIDS	(%)	10,657.75	1.919*					
YOUNG	(%)	−11,408.32	2.177**	−18,374.51	4.549***	−1340.13	−2818.26	−0.030664
OLD	(%)	−13,220.83	2.536**	−21,009.74	5.476***	−1531.96	−3221.66	−0.035053
SINGLE	(%)	−300.79	0.065					
Neighborhood								
DISTCBD	(mi)	−18,847.70	12.752***	−19,092.36	13.128***	−1392.65	−2928.70	−0.031866
ELTRAIN	(mi)	−3,650.23	1.986**	−3,673.60	2.013**	−268.01	−563.61	−0.006132
HIGHWAY	(0|1)	2,722.50	1.566					
LAKE	(0|1)	31,948.21	9.566***	31,953.67	9.728***	2331.24	4902.53	0.053342
MEDVALUE	($K)	43.63	8.319***	41.61	8.239***	3.04	6.38	0.000069
NEIGHAGE	(yrs)	73.02	1.994**	70.87	1.946*	5.16	10.85	0.000118
NODRIVE	(%)	24,284.05	9.622***	24,576.71	10.235***	1792.52	3769.63	0.041016
OWNEROCC	(%)	4,909.31	1.575					
PARK	(0|1)	7,550.20	5.438***	7,624.01	5.494***	556.32	1169.92	0.012729
PUBHSNG	(mi)	22,045.22	13.362***	22,345.11	13.877***	1630.49	3428.87	0.037308
UNIT1	(%)	−11,129.17	4.096***	−7,521.94	6.153***	−548.43	−1153.34	−0.012549
UNIT2	(%)	−24,447.35	12.920***	−22,918.77	13.879***	−1672.25	−3516.70	−0.038264
UNIT3OR4	(%)	−1,149.23	0.507					
VACANCY	(%)	37,770.92	10.060***	36,484.77	10.231***	2659.29	5592.40	0.060848

[a] The dependent variable *RENBLDG* is the renovation expenditure per building.
[b] Units of measurement are in parentheses: (0|1) indicates a dummy variable; ($K) denotes thousands of dollars.
[c] Tobit II includes variables from Tobit I that are significant at the 95% level.
[d] Marginal effect (of a one-unit change in the explanatory variable) on the expected value of the "latent" variable *OPTINV*, and its McDonald–Moffitt decomposition into: (1) marginal effect on uncensored observations (Δ*RENBLDG*); and (2) marginal effect on probability that observation is uncensored ($\Delta\Pr[OPTINV > 0]$). See explanation in footnote 25.
* Significant at 90% level.
** Significant at 95% level.
*** Significant at 99% level.

Table 4
Estimation results of the tobit models (dependent variable is *RENUNIT*[a])

Variable[b]		Tobit III		Tobit IV[c]		Tobit IV: marginal effects[d]		
		Coefficient	|*t*-value|	Coefficient	|*t*-value|	Δ*OPTINV*	Δ*RENUNIT*	ΔPr(*OPTINV* > 0)
Constant		−57,233.97	36.959***	−55,334.21	58.236***			
Building								
AGE	(yrs)	135.53	30.676***	136.16	31.844***	9.95	20.91	0.000656
DWLUNITS	(#)	−46.77	5.232***	−46.96	5.261***	−3.42	−7.20	−0.000226
STORIES	(#)	864.36	9.634***	856.44	9.643***	62.47	131.38	0.004121
VACANT	(0|1)	5,612.12	5.237***	5,527.41	5.165***	403.30	848.13	0.026600
Demographic								
BLACK	(%)	6,268.26	14.238***	6,133.33	14.289***	447.56	941.20	0.029519
OTHERMIN	(%)	891.29	0.661					
COLLEGE	(%)	13,107.58	12.025***	13,211.31	12.839***	963.66	2026.56	0.063559
FOREIGN	(%)	−3,287.94	3.673***	−3,105.50	3.708***	−226.46	−476.24	−0.014936
MFI	($K)	−53.33	4.925***	−52.38	5.058***	−3.82	−8.03	−0.000252
KIDS	(%)	2,371.55	1.238					
YOUNG	(%)	−4,218.68	2.338**	−5,751.28	4.290***	−419.56	−882.32	−0.027672
OLD	(%)	−5,623.31	3.128***	−7,283.63	5.575***	−530.99	−1116.66	−0.035022
SINGLE	(%)	332.89	0.208					
Neighborhood								
DISTCBD	(mi)	−5,909.34	11.545***	−5,849.69	11.683***	−426.68	−897.30	−0.028142
ELTRAIN	(mi)	−1,524.86	2.424**	−1,625.89	2.613***	−118.57	−249.35	−0.007821
HIGHWAY	(0|1)	1,213.01	2.037*					
LAKE	(0|1)	3,749.60	2.973***	3,828.28	3.073***	279.26	587.28	0.018419
MEDVALUE	($K)	21.99	12.076***	21.71	12.024***	1.58	3.33	0.000104
NEIGHAGE	(yrs)	16.82	1.329					
NODRIVE	(%)	6,716.27	7.701***	6,571.79	7.720***	479.22	1007.79	0.031607
OWNEROCC	(%)	2,088.71	1.926*					
PARK	(0|1)	1,205.12	2.468**	1,205.98	2.471**	88.04	185.14	0.005806
PUBHSNG	(mi)	7,912.90	13.884***	7,845.97	14.123***	572.54	1204.05	0.037762
UNIT1	(%)	−1,354.69	1.431					
UNIT2	(%)	−5,773.08	8.786***	−5,146.21	10.383***	−375.68	−790.05	−0.024778
UNIT3OR4	(%)	2,728.67	3.472***	2,806.98	4.065***	204.58	430.23	0.013493
VACANCY	(%)	11,406.20	8.690***	10,585.97	8.655***	771.26	1621.93	0.050869

[a] The dependent variable *RENUNIT* is the renovation expenditure per dwelling unit.
[b] Units of measurement are in parentheses: (0|1) indicates a dummy variable; ($K) denotes thousands of dollars.
[c] Tobit IV includes variables from Tobit III that are significant at the 95% level.
[d] Marginal effect (of a one-unit change in the explanatory variable) on the expected value of the "latent" variable *OPTINV*, and its McDonald–Moffitt decomposition into: (1) marginal effect on uncensored observations (Δ*RENUNIT*); and (2) marginal effect on probability that observation is uncensored (ΔPr[*OPTINV*>0]). See explanation in footnote 25.
* Significant at 90% level.
** Significant at 95% level.
*** Significant at 99% level.

Table 5
Signs and significance of coefficients

Variable	Expected	Probit	Tobit		Variable	Expected	Probit	Tobit	
			RENBLDG	*RENUNIT*				*RENBLDG*	*RENUNIT*
Building					*Neighborhood*				
AGE	+	+	+	+	*DISTCBD*				
DWLUNITS	−	−	−	−	*ELTRAIN*				
STORIES	+	+	+	+	*HIGHWAY*	−			
VACANT	+	+	+	+	*LAKE*	+	+	+	+
Demographic					*MEDVALUE*	?	+	+	+
BLACK	?	+	+	+	*NEIGHAGE*	+	+		
OTHERMIN	?	+			*NODRIVE*	+	+	+	+
COLLEGE	+	+	+	+	*OWNEROCC*	+	+		
FOREIGN	?				*PARK*	+		+	+
MFI	?	−		−	*PUBHSNG*	+	+	+	+
KIDS	−				*UNIT1*				
YOUNG	+		−	−	*UNIT2*	−	−	−	−
OLD	−	−	−	−	*UNIT3OR4*	+	+		+
SINGLE	+				*VACANCY*	+	+	+	+

Note. Symbols denote the (expected) sign and significance (at 95% level) of each variable, as follows: +: positive and significant, −: negative and significant, ?: expected sign and significance are uncertain, [blank]: insignificant.

explanatory variables' coefficients in all of the models are compared with the hypothesized results.

The pseudo-R^2 statistics of all of the models are very low; in fact, none is higher than 0.0219. However, this weak predictive power is to be expected, since the data include only the improvements made within a five-year span. Therefore, many buildings whose characteristics make them likely candidates for renovation may have been renovated before or after this time period. Despite their low goodness-of-fit measures, all of the models are highly significant overall: F-tests and χ^2-tests confirm that the probability is virtually zero that all of the coefficients are equal to zero.

The results consistently demonstrate that the building characteristics *AGE*, *DWLUNITS*, and *STORIES* are significant predictors of renovation. In every model, the positive coefficient of *AGE* is the most strongly significant of all, which supports the findings of other empirical studies: *ceteris paribus*, renovation is more likely and more extensive and/or expensive for older buildings. As expected, the coefficient of *DWLUNITS* is negative and the coefficient of *STORIES* is positive. While these signs are seemingly counterintuitive, they support the hypothesis that renovators prefer low-density buildings with large living spaces, as explained earlier.

The housing age and density variables are less statistically significant at the neighborhood level, but even many of the insignificant results support the hypothesized effects. The coefficient of *NEIGHAGE* is positive in the probit model but insignificant in both tobit models, which implies that a building in a historic neighborhood is more likely to be renovated, but that the level of renovation expenditure is no different from the level that would be expected if the same building was located in a newer neighborhood. To interpret the density variables *UNIT1*, *UNIT2*, and *UNIT3OR4*, it should be noted that their coefficients indicate their influence relative to the omitted categorical variable, which is the percentage of buildings with five or more dwelling units. The coefficient of *UNIT1* is negative in the probit model, which supports the expectation that renovation activity is less likely in neighborhoods in which there is a high percentage of single-family homes. Its coefficient is also negative in the tobit model in which *RENBLDG* is the dependent variable, but it is insignificant in the tobit model in which *RENUNIT* is the dependent variable. Though renovation expenditures should be influenced only by a building's own structural density (i.e., *DWLUNITS* and *STORIES*), not by the housing density in its neighborhood, both results make sense on the individual building level: since single-unit buildings are smaller than multiple-dwelling buildings, they should cost less to renovate; the average cost per dwelling unit, however, should not be significantly different. The same interpretations apply to the negative coefficient of *UNIT2*. The apparent explanation (again, on the building level) for its coefficient in the *RENUNIT* tobit is that many duplexes are attached or semi-attached buildings, so renovation costs per unit may be lower than they are in larger detached buildings. The probit coefficient for *UNIT3OR4* indicates that renovation is more likely in these moderate-density neighborhoods, and its *RENBLDG* tobit coefficient suggests that expenditures per building are not significantly different than they are in higher-density neighborhoods. The only conceivable interpretation for *UNIT3OR4*'s positive coefficient in the *RENUNIT* tobit is that (on the building level) per-unit renovation costs are lower for buildings with more dwelling units due to economies of scale.

As expected, the positive coefficients of *VACANT* indicate that vacant buildings are more likely to be renovated than occupied buildings, and that expenditures are likely to be higher. The coefficient of *VACANCY* is also positive in all three models, though the probit result is somewhat surprising, since brisk renovation activity seems improbable in neighborhoods with very high vacancy rates.

In all of the models, *MEDVALUE* has a positive coefficient, which suggests that returns to improvements increase with the neighborhood's median housing value. Since data on individual housing values were not available, it cannot be determined whether this effect applies uniformly to all buildings or differs by relative value. The coefficients of *OWNEROCC* indicate that owner-occupiers are more likely than absentee landlords to renovate their properties, but that they spend neither more nor less than landlords when they do.

The negative coefficients of *DISTCBD* and *ELTRAIN* confirm that accessibility to the city center matters to renovators: the further a building is from the central business district and the further it is from an El station, the less likely it is to be renovated and the lower expenditures are likely to be. This tendency is also indirectly reflected by the positive coefficients of *NODRIVE*. Though *HIGHWAY* was expected to have a negative coefficient, it is insignificant in all of the models. This result implies that the noise, pollution, and congestion near an interstate highway is not a strong enough disamenity to deter renovation, or that a highway's disamenity value is offset by its convenience value.

As expected, public housing is a strong disamenity: the positive coefficients of *PUBHSNG* indicate that the likelihood and level of renovation activity increases with distance from a housing project. The influence of the amenity variables *LAKE* and *PARK* (in the tobit models) is consistent with the expectation that renovation is more likely and more extensive if a building is near Lake Michigan or a city park.

The effect of the neighborhood income variable *MFI*, which was difficult to predict, is negative in two of the three estimations. This finding implies that upper-income neighborhoods, in which buildings are expensive to acquire and have been consistently well-maintained, are less likely to experience extensive renovation activity.

Sociological studies of gentrification often assert that renovators are attracted by racially mixed and ethnic neighborhoods, but most empirical studies have found that minority areas experience less renovation activity. In this paper, the strongly and consistently positive coefficient of *BLACK*, as well as the positive probit coefficient of *OTHERMIN*, supports the former hypothesis, though it is difficult to posit a compelling explanation for why renovation is more likely or more extensive in areas with a high minority concentration. As hypothesized in the previous section, the influence of neighborhood racial composition is difficult to determine. The coefficient of *FOREIGN* is insignificant in the probit and *RENBLDG* tobit, but negative in the *RENUNIT* tobit. As with the race variables, it is probably unrealistic to conclude that the relationship between a neighborhood's concentration of immigrants and renovation activity is simply linear.[26]

The positive coefficient of *COLLEGE* in all three models supports the popular characterization of gentrifiers as well-educated. However, in light of this result and the correlation between *COLLEGE* and *MFI* ($\rho = 0.6614$), the negative coefficient of *MFI* is somewhat puzzling. The insignificant probit coefficient of *YOUNG* and the consistently insignificant coefficients of *SINGLE* and *KIDS* raise considerable doubt about the accuracy of the "life stage" hypothesis that much of the inner-city renovation activity is performed by unmarried, childless young adults. In fact, the coefficient of *YOUNG* is negative in both tobit models, which implies that neighborhoods with a high proportion of twenty- and thirty-somethings experience lower levels of renovation when it does occur. Since many of these young adults are probably renters, this result is understandable. At the other end of the age continuum, *OLD* has a consistently negative coefficient, as expected, which indicates that seniors' housing capital levels tend to be close to optimal.

Two blockgroup-level thematic maps are presented in Figs. 1 and 2. The first map illustrates the difference between the average predicted probability and the actual proportion of buildings that were renovated in each blockgroup. Each shade includes one quintile of the range of block-groups. Darker shades depict areas in which the model (Probit II) "under-predicts" renovation activity, i.e., block-groups in which the actual rate exceeds the predicted rate. The map in Fig. 2 is identical, except that only the most under-predicted quintile of block-groups is shaded.

The under-predicted areas appear to be concentrated in three distinct "clusters," which are circled on the second map. The northernmost cluster includes Lakeview and Uptown, two neighborhoods in the vicinity of Wrigley Field, into which gentrification has rapidly spread in recent years. Virtually all of Lincoln Park and the Near North Side, the neighborhoods between Lakeview and the downtown Loop, has extensively gentrified over the past two decades; accordingly, renovation there is over-predicted. The circled cluster to the west of the Loop includes several gentrifying residential areas (such as Ukrainian

[26] Including quadratic terms for the income and race variables failed to capture nonlinearities in these demographic variables' influences.

Fig. 1. Difference between predicted and actual renovation.

Village) and the West Loop, a formerly industrial area in which many warehouses are being converted into loft apartments. The cluster to the south includes Hyde Park and other residential neighborhoods in the vicinity of the University of Chicago and the University of Illinois at Chicago campuses.

This paper's model cannot explain why these three gentrifying areas are experiencing more renovation activity than other neighborhoods with comparable characteristics. One possibility is that the data omit some attributes or amenities that account for their attractiveness to renovators. Another possibility is that renovation activity has been catalyzed by governmental or non-profit redevelopment projects or subsidies, though this explanation is much more likely to account for the smaller clusters of under-predicted renovation in the South Side, e.g., HOPE VI redevelopment of public housing. A third—and the most interesting—explanation is that housing renovation exhibits

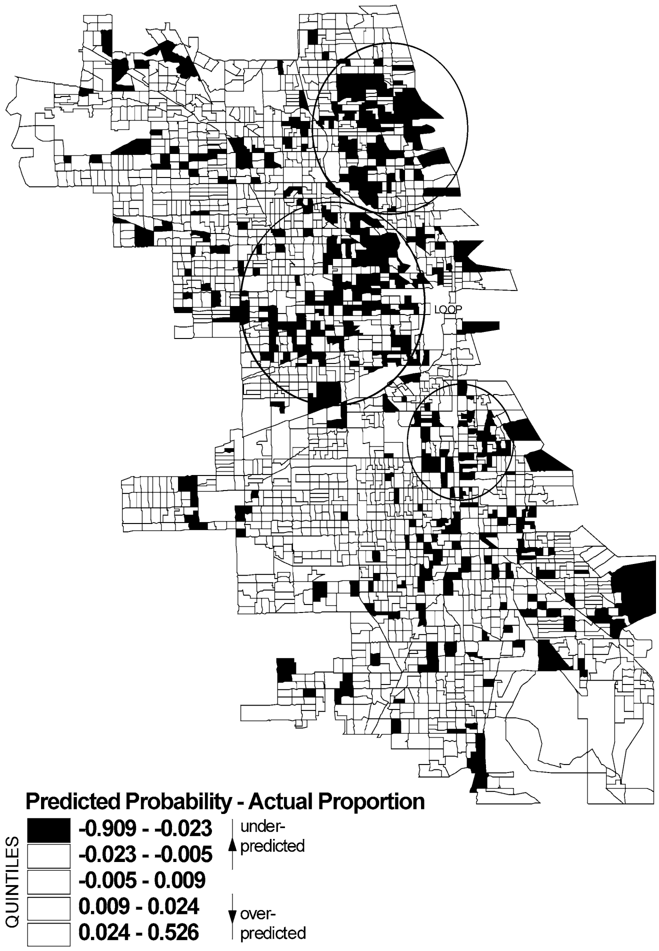

Fig. 2. Under-predicted areas (possible “clustering”).

spatial dependence, which standard econometric techniques cannot capture. For example, an endogenous neighborhood “spillover” or “feedback” effect may cause renovations performed on a building to increase the likelihood that other nearby buildings will be renovated.

5. Conclusion

Since virtually every existing empirical study of the determinants of inner-city housing renovation has yielded mostly inconclusive results, this paper closes a wide gap in the literature. Moreover, the unprecedented extent and detail of the microdata used in these estimations instill confidence in the results.

By and large, the results confirm intuitive expectations and support anecdotal accounts about the determinants of renovation, particularly as it occurs in the context of gentrification. Older, low-density houses in older, moderate-density neighborhoods are most likely to be renovated. Accessibility to the CBD matters: improvement is more likely in areas that are close to downtown and well-served by mass transit. Housing vacancy does not deter renovation, but nearby public housing projects do. Neighborhood amenities, including city parks and bodies of water (Lake Michigan in this case), encourage renovation activity.

The influences of some neighborhood attributes were difficult to predict, so these results are especially enlightening. Even more thought-provoking are the empirical results that are at odds with initial expectations or the conclusions of other studies. Housing that is near the busy interstate highways that traverse Chicago is no less likely to be improved. A high median housing value increases the likelihood of renovation activity, but a high median income level decreases the likelihood. Rehabilitation is more likely in areas where the population is well-educated, but less likely in areas with a high proportion of young adults, and neither more nor less likely in areas with a high concentration of singles or children. Particularly surprising is the finding that renovation seems to be more likely and more extensive in neighborhoods with a high population of blacks or other minorities.

While the data are rich in information about the buildings and neighborhoods, they lack any personal characteristics of the property owners. Consequently, owner-occupiers are inseparable from landlords. The block-level variable *OWNEROCC* was included to compensate for the omission of a building-level ownership variable, but the insignificance of its coefficient in both tobit models suggests that while these two groups of agents undoubtedly have different objectives, the determinants of renovation activity apply similarly to both. Another consequence is the indistinguishability of gentrification-based renovation from incumbent upgrading. As noted earlier, however, housing renovation is only one part of gentrification. By considering this paper's results in conjunction with studies of inner-city resettlement, the "back-to-the-city" movement can be more fully understood.

Other omissions include land-use and zoning variables, crime statistics, detailed information about architectural styles and construction materials, and the presence of governmental rehabilitation incentives and redevelopment programs. In particular, the latter could provide insight into the effectiveness of existing urban revitalization efforts. Even in its absence, however, the results of this paper offer important guidance to public policy makers.

Finally, this paper and other renovation studies do not consider the spatial phenomena known as "neighborhood effects" (discussed in the previous section).[27] These effects could be empirically examined in a time-series framework, in which the aggregate level of recent renovation activity in each building's block is included as a time-lagged explanatory variable. Alternatively, spatial econometric methods could be applied by using a "spatial lag" model to capture contemporaneous feedbacks among renovations in the same neighborhood.[28]

---

[27] Current work by the author empirically examines these effects, using a "spatial lag" model, as described.

[28] Anselin [2] provides comprehensive coverage of spatial econometric modeling.

Acknowledgments

This paper is based upon dissertation research that was completed at the University of Illinois and funded in part by a grant from the US Department of Housing and Urban Development, Office of Policy Development and Research. I thank my parents, Julie and Samuel Helms, and my undergraduate advisor, Clifford Kern, for their inspiration and encouragement, and my dissertation committee members and two anonymous referees for their insightful comments. I am especially grateful to my advisor and mentor, Jan Brueckner, for his limitless generosity and enduring support.

References

[1] W. Alonso, Location and Land Use: Toward a General Theory of Land Rent, Harvard Univ. Press, Cambridge, MA, 1964.
[2] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic, Dordrecht, 1988.
[3] R. Arnott, R. Davidson, D. Pines, Housing quality, maintenance, and rehabilitation, Review of Economic Studies 50 (1983) 467–494.
[4] R. Arnott, R. Davidson, D. Pines, Spatial aspects of housing quality, density, and maintenance, Journal of Urban Economics 19 (1986) 190–217.
[5] B. Berry, R. Bednarz, The disbenefits of neighborhood and environment to urban property, in: D. Segal (Ed.), The Economics of Neighborhood, 1979, pp. 219–246.
[6] A. Bogdon, Homeowner renovation and repair: the decision to hire someone else to do the project, Journal of Urban Economics 5 (1996) 323–350.
[7] J. Brueckner, J. Thisse, Y. Zenou, Why is central Paris rich and downtown Detroit poor? An amenity-based theory, European Economic Review 43 (1999) 91–107.
[8] Chicago, City of, Department of Buildings, Building permits data, Chicago, 2000.
[9] Chicago Property Information Project, Harris file, Chicago, 2000.
[10] P. Chinloy, The effect of maintenance expenditures on the measurement of depreciation in housing, Journal of Urban Economics 8 (1980) 86–107.
[11] P. Clay, Neighborhood Renewal: Middle-class Resettlement and Incumbent Upgrading in American Neighborhoods, Lexington Books, Lexington, MA, 1979.
[12] L. Dildine, F. Massey, Dynamic model of private incentives to housing maintenance, Southern Economic Journal 40 (1974) 631–639.
[13] D. Gale, Neighborhood Revitalization and the Postindustrial City, Lexington Books, Lexington, MA, 1984.
[14] G. Galster, Homeowners and Neighborhood Reinvestment, Duke Univ. Press, Durham, NC, 1987.
[15] C. Kern, Private residential renewal and the supply of neighborhoods, in: D. Segal (Ed.), The Economics of Neighborhood, 1979, pp. 121–146.
[16] C. Kern, Upper-income renaissance in the city: its sources and implications for the city's future, Journal of Urban Economics 9 (1981) 106–124.
[17] C. Kern, Upper income residential revival in the city, in: R. Ebel, J. Henderson (Eds.), Research in Urban Economics, Vol. 4, JAI Press, Greenwich, CT, 1984.
[18] S. Laska, D. Spain (Eds.), Back to the City: Issues in Neighborhood Renovation, Pergamon, Elmsford, NY, 1980.
[19] S. LeRoy, J. Sonstelie, Paradise lost and regained: transportation innovation, income, and residential location, Journal of Urban Economics 13 (1983) 67–89.
[20] N. Mayer, Rehabilitation decisions in rental housing: an empirical analysis, Journal of Urban Economics 10 (1981) 76–94.
[21] J. McDonald, R. Moffitt, The Uses of Tobit Analysis, Review of Economics and Statistics 62 (1980) 318–321.
[22] D. Melchert, J. Naroff, Central-city revitalization: a predictive model, AREUEA Journal 15 (1987) 664–683.

[23] R. Mendelsohn, Empirical evidence on home improvements, Journal of Urban Economics 4 (1977) 459–468.
[24] E. Mills, An aggregative model of resource allocation in a metropolitan area, American Economic Review 57 (1967) 197–210.
[25] C. Montgomery, Explaining home improvement in the context of household investment in residential housing, Journal of Urban Economics 32 (1992) 326–350.
[26] R. Muth, Cities and Housing, Univ. of Chicago Press, Chicago, 1969.
[27] K. Nelson, Gentrification and Distressed Cities, Univ. of Wisconsin Press, Madison, 1988.
[28] W. Shear, Urban housing rehabilitation and move decisions, Southern Economics Journal 49 (1983) 1030–1052.
[29] J. Sweeney, A commodity hierarchy model of the rental housing market, Journal of Urban Economics 1 (1973) 288–323.
[30] US Census Bureau, 1990 Census of Population and Housing, Summary tape files 1B and 3A, Washington, DC, 1990.
[31] US Census Bureau, Annual new privately-owned residential building permits, Washington, DC, 2000.
[32] US Census Bureau, Census 2000 Supplementary Survey Summary Tables, Washington, DC, 2000.
[33] W. Wheaton, Income and urban residence: an analysis of consumer demand for location, American Economic Review 67 (1977) 620–631.