Columbia University

Department of Economics
Discussion Paper Series

"Causes" of Homelessness: Understanding City-and Individual-Level Data

Brendan O'Flaherty

Discussion Paper #:0102-59

Department of Economics
Columbia University
New York, NY 10027

April 2002

“Causes” of Homelessness: Understanding City- and Individual-Level Data

Brendan O'Flaherty*

April 22, 2002

Abstract

Studies of homelessness that use city-level observations get systematically different results from studies that use individual-level data. I explain why. The findings are consistent with a model of homelessness as a condition requiring a conjunction of unfortunate circumstances.

*Department of Economics. Department of International and Public Affairs. Columbia University. New York, New York 10027. Email: bo2@columbia.edu. I am grateful to Edgar Olsen, Steve Cameron, and Eugene Smolensky for helpful advice and suggestions. An earlier version of this paper was presented at the Goldman School of the University of California at Berkeley.

Two different kinds of empirical cross-section studies of modern American homelessness have arrived at apparently contradictory conclusions. Studies that take as their unit of observation homelessness rates in different cities have generally found that housing market conditions have large effects, while population composition—the size of the mentally ill population outside of state psychiatric facilities, for instance, or the or the extent of poverty–uusally does not. By contrast, studies that take individuals as their unit of observation find effects for housing market conditions of the cities where the individuals find themselves, and strong effects for personal personal characteristics. The two types of studies seem to contrasting policy advice: city-level studies say reduce rents and increase vacancies, individual-level studies say work on pathology and poverty.

I will argue that the two sets of results are complementary, not contradictory, by showing a very simple model of homelessness that implies both kinds of results (and also implies that both varieties of regression are misspecified).

1 LITERATURE REVIEW

Since modern homelessness first rose in the early 1980s, over a dozen published empirical studies have attempted to determine what factors are responsible for its volume. Early debates, primarily outside economics, focused on the question of whether "individual failings"(for instance, mental illness or substance abuse) or "structural problems" (for instance, high rents) were "responsible" for homelessness (see, e.g., Burt (1992), Jencks (1994)), but such stark contrasts are no longer so pervasive in the theoretical literature. The majority of empirical studies of homelessness in economics have used cities (or counties or metropolitan areas) as their units of observation. These include Tucker (1989), Quigley (1990), Appelbaum, Dolny, Dreier and Gilderbloom (ADDG) (1991), Bohannon (1991), Elliott and Krivo (1991), Burt (1992), Filer and Honig (1993), Grimes and Chressanthis (GC) (1997), Troutman, Jackson, and Ekelund (TJE) (1999), Quigley, Raphael and Smolensky(QRS) (2001), and Early and Olsen (EO) (2001). Another group of empirical studies use individuals as their units of observation, and include individuals in different cities. These studies include Early (1998, 1999) and 1999) and Early and Olsen (1999). The contrast between these studies is the focus of this paper. (A third set of empirical studies uses data drawn entirely from within the same city, with same city, with variation being supplied by either the cross-section (for example, in Bassuk (1997) or the time series O'Flaherty (1999)). These within-city studies are not the subject of this paper.)

Different kinds of variables tend to be important in these two different kinds of studies. In the city-level studies, researchers generally find that housing market parameters, broadly understood, determine the volume of homelessness, and that indicators of personal characteristics have little or no influence. In particular, measures of rent almost always have significant coefficients in these studies, while measures of poverty almost always do not. Other housing market variables like vacancy rates, climate, and the presence of rent control are sometimes significant, while other measures of individual characteristics like race, gender, drug use, and mental illness are rarely significant.

These results are almost entirely reversed in the studies that use individual-level observations. Variables like poverty, gender, race, and mental illness are almost always significant in these regressions, while rents and vacancy rates are never significant. significant. Sometimes climate and rent control matter, but not in all studies. A naive observer who read both kinds of wanted to know whether individual failings or structural problems were responsible for homelessness would be very confused.

Table 1 below summarizes the results of most of these studies. The studies use different datasets, different definitions of homelessness, different measures of the variables of interest, different techniques, and different sets of explanatory variables that are not reported on. Some of the studies are much more sophisticated and carefully executed than some of the others. Thus there are many reasons to expect different results. Nevertheless, the pattern in table 1 is quite strong—personal characteristics matter in individual-level studies, housing market characteristics matter in city-level studies.

I omitted Tucker (1989) from this table because he employed no personal characteristics and his methods were unorthodox. I also omitted QRS (2001) because it was difficult to summarize—it involved four different data sets, and several regressions with each data set. Although QRS used many variables different from those in the other city-level studies and tested a particular model of the housing market, their results generally conformed to the results of the other city-level studies. They included measures of the mentally ill and ex-offender populations that were never significant, while many of their housing market measures were significant. Compared with other city-level studies, they found more significant results about poverty, but their use of poverty-related variables was in part based on theories about the effect of poverty on the housing market.

TABLE 1: Studies of Homelessness

	Personal characteristics					Housing market characteristics			
cities	Poverty	Gender	Race	Mental h'lth	Drugs	Rent	Vacancy	Rent c'tr'l	Climate
Quigley -1	0	-	-	-	-	**	0	0	**
ADDG	0	-	-	-	-	0	*	0	**
Burt-2	0	-	-	-	-	*	0	-	-
Elliott-Krivo	0	-	**	**	-	**	-	-	-
Bohanon-3	na	-	-	**	-	**	-	0	0
Filer-Honig	0	-	*	*	-	**	*	-	0
TJE-4	0	-	-	-	*	*	**	**	0
GC-4	0	-	-	0	-	**	-	**	0
EO 2001	0	0	0	0	-	**	0	0	0
individuals									
Early 98	**	**	**	**	0	0	0	-	*
Early 99	**	**	0	**	-	0	0	-	*
EO 99	**	**	**	**	0	0	0	*	**

Legend: * = coefficient significant at the 10% level and in the "right" direction.

** = coefficient significant at the 5% level and in the "right" direction.

0 = coefficient insignificant or not in the "right" direction.

- = explanatory variable not included in the study.

Notes:

1. Quigley's equation II
2. Equation 9-7, best model, all cities.
3. Dependent variable is ln (homeless/poor)
4. 1990 census data

Early and Olsen (2001) note that one important difference between the individual-level studies and the city-level studies is in the amount of variation they allow the econometrician to observe. Individual-level studies provide a great deal of variation in personal characteristics, and so allow the coefficients of those variables to be estimated very precisely, but they do not allow a great deal of variation in housing market characteristics. City-level studies are the reverse. I do not dispute this explanation, or deny that it has some relevance. The difference between individual-level studies and studies, however, is much more fundamental than a difference in precision of estimates, which would disappear in infinitely large samples. I argue below that even with infinitely large samples, the two types of studies will give different estimates.

2 A MODEL

Consider a set of cities $i-1,...C$,each with population n (this mitigates weighting problems). Each individual j in city i is either at risk of homelessness or not. "Risk of homelessness" depends on such individual-level characteristics as mental illness, poverty, substance abuse, maleness, minority status, weak family ties, tastes for independence, skills in living outdoors, living outdoors, and so on. We assume that "risk of homelessness" is a binary variable (either you are at risk of or you are not), perfectly observable to the econometrician, and denote $m_{ij} = 1$ if person j in city i is at risk; m$_{ij} = 0$ otherwise. Write $m_i = \sum_j m_{ij}$ as the size of the at-risk population in city i.

We model the housing market in each city even more simply. There are a fixed and immutable number of houses $(n - H_i)$ in city city i. We also assume that H_i is perfectly observable to the econometrician (who may have to use rent, vacancy, climate, and rent control variables in order to observe

Homelessness arises when at-risk individuals cannot find places to live.Specifically, there are two types of cities: those in which the housing shortfall is smaller than the at-risk population; and those in which it is larger. The former we call "housing-rich" cities, the latter we call "housing-short" cities. We denote the set of housing-rich cities as R

$$R = \{i \lfloor m_i \geq H_i\}$$

and the set of housing-short cities as S

$$S = \{i \lfloor m_i \leq H_i\}$$

People who are not at-risk never become homeless.

In housing-rich cities, the number of at-risk people who become homeless is H_i. At-risk people are at the end of the queue for housing, and the housing supply runs out before they can all be accommodated, but after some of them are accommodated. In housing-short cities, all at-risk people are homeless, but people who are not at-risk find some other way of coping with the housing shortage: they double-up with relatives, pay for illegal subletting and subdivision, take long vacations, move into hotels, and so on. (Alternatively, you can think of $(n - H_i)$ as the immutable number of affordable houses; then people who are not at risk live in houses that are not affordable in cities in housing shortage.)

Let $h_{ij} = 1$ if person j in city i is homeless; $h_{ij} = 0$ otherwise, and $\sum_j h_{ij} = h_i$. Then on a city level we have

(1)

$$\begin{aligned} h_i &= \min[H_i, m_i] \\ &= H_i \qquad i \in R \\ &= m_i \qquad i \in S. \end{aligned}$$

On an individual level we have

$$\begin{aligned} \Pr(h_{ij} &= 1) = 0 \qquad if \quad m_{ij} = 0 \\ &= H_i/m_i \qquad if \quad i \in Randm_{ij} = 1 \\ &= 1 \qquad\qquad if \quad i \in Sandm_{ij} = 1 \end{aligned}$$

or

(2)

$$\Pr(h_{ij} = 1) = m_{ij} \min(H_i/m_i, 1).$$

Like all models, this one is a gross over-simplification. The key assumption is that being homeless is not just a matter of being *either* the wrong kind of person *or* in the wrong kind of place; rather, it depends on being *both* the wrong kind of person and in the wrong kind of place. Elsewhere (1995), for instance, I have developed a much more detailed model of homelessness, and Park (1997) has developed a model that includes both homelessness and positive vacancy rates. Those models can be understood as elaborate discussions of what determines H_i in H_i in each city. In the appendix, I show a simple equilibrium model with optimizing behavior in which (1) and (2)

3 ECONOMETRIC RESULTS

Equations (1) and (2) describe homelessness on a city level and on an individual level, respectively, under our model, but these are not the equations that have been estimated in the literature. In this section, we study simplified versions of the equations that have actually been estimated. We show that the estimates derived from these equations tell us little about homelessness, but much about the relationship between m_i and H_i, about which most people have no intuition or a great deal of interest (other than learning which is smaller in a particular city). We also show that the apparently divergent results of the city-level and individual-level estimations can be reconciled with our model and the additional premise that most cities are in set S.

For tractability, we confine our attention to OLS estimates.

A. City-Level Estimates

The usual equation estimated in studies with city-level observations can be thought of as

$$h_i = \alpha^c m_i + \beta^c H_i + \epsilon_i^c, \tag{3}$$

where the last term are i.i.d. mean zero errors. We omit an intercept in the belief that a city without either at-risk individuals or any kind of housing shortage would have no homelessness.

To understand the estimates for (3), it is helpful to think about two auxiliary regressions. These are the two linear relationships between m_i and H_i :

(4) $$m_i = B^m H_i + \varepsilon_i^m$$

(5) $$H_i = B^H m_i + \varepsilon_i^H.$$

Assume we have estimated these two equations by OLS, obtaining coefficient estimates $\widehat{B}^m$ and $\widehat{B}^H$ and residuals (e_i^m) and (e_i^H) in the process. It is easy to show that both coefficient estimates will be positive, and that $\widehat{B}^m \widehat{B}^H \leq 1$, with equality only if m_i and H_i are perfectly correlated, which we rule out for simplicity.

Then after much calculation we can write the OLS estimates of the coefficients in (3) as

(6)
$$\begin{pmatrix} \widehat{\alpha}^c \\ \widehat{\beta}^c \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{1 - \widehat{B}^m \widehat{B}^H} \begin{pmatrix} \frac{\sum_R (m_i - H_i) e_i^m}{\sum m_i^2} \\ \frac{\sum_S (H_i - m_i) e_i^H}{\sum H_i^2} \end{pmatrix}$$

where summations without arguments are over the entire set of cities. It is easy to prove that both coefficient estimates are nonnegative. Either coefficient estimate may be greater than one, although this is unlikely, and we can prove that at most one of them is greater than one:

Proposition 1: It is impossible for both $\widehat{\alpha}^c \succ 1$ and $\widehat{\beta}^c \succ 1$.

Proof of proposition 1: Suppose $\widehat{\alpha}^c \succ 1$.

We will show $\widehat{\beta}^c \prec 1$. Since $(1 - \widehat{B}^m \widehat{B}^H) \prec 1, \widehat{\alpha}^c \succ 1$ implies that the sum over R,

$$\sum_R (m_i - H_i) e_i^m \prec 0.$$

Since$(m_i - H_i) \geq 0$ for all $i \in R$, this can happen only if $e_i^m \prec 0$ for some city in R. Since

$$e_i^m = m_i - \widehat{B}^m H_i$$

and

$$m_i - H_i \geq 0$$

we see that $e_i^m \prec 0$ is for some i is possible only if $\widehat{B}^m > 1$. This implies $\widehat{B}^H < 1$. Since $\widehat{B}^H < 1$,

$$e_i^H = H_i - \widehat{B}^H m_i > H_i - m_i > 0$$

for all $i \in S$. Hence the summation over S,

$$\sum_S (H_i - m_i) e_i^H > 0.$$

From (6), this implies $\widehat{\beta}^c < 1$.

Similar reasoning leads to the conclusion that $\widehat{\beta}^c > 1$ implies $\widehat{\alpha}^c < 1. QED$

To understand (6), first think about what happens if the set S is empty: if $H_i < m_i$ in all cities. Then

$$\begin{pmatrix} \widehat{\alpha}^c \\ \widehat{\beta}^c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

all that matters in determining homelessness rates in a city are housing market conditions. Homelessness in each city is constrained not by the supply of individuals at risk, but by the size of the housing shortfall.

Symmetrically, if the set R is empty

$$\begin{pmatrix} \widehat{\alpha}^c \\ \widehat{\beta}^c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

all that matters in determining homelessness are population characteristics. Homelessness is constrained by the supply of at-risk individuals, not the housing shortfall.

In general, then, (6) shows that if most cities are in set R, and the difference $(m_i - H_i)$ is great, then the coefficient on housing market characteristics will be large and the coefficient on population characteristics will be small. Since most studies of this type reach this conclusion, what they are telling us is that most cities are in set R.

Such a conclusion does not strain intuition. Most estimates on the national level of the number of severely mentally ill people, the number of substance abusers, the number of extremely poor people, the number of male individuals, and the number of members of racial minority groups all place their numbers well in excess of the number of homeless people. We should not be surprised if the same were true for the most part on the local level as well. Alternatively, if we interpret people who have actually been homeless at any time in the past several years as having the individual characteristics that place them at risk of homelessness, then the standard results about turnover also imply that the number of people at-risk is substantially greater than the homeless population at any moment.

B. Individual-Level Estimates

The usual equation estimated in studies with individual-level observations can be thought of as

(7) $$h_{ij} = \alpha^I m_{ij} + \beta^I H_i + \epsilon_{ij}^I,$$

where the last term are i.i.d. mean zero errors.

As with the city-level estimates, understanding these estimates is helped by focussing first on three auxiliary equations:

(8) $$\frac{H_i}{m_i} = A^R + B^R H_i + E_i^R i \in R$$

(9) $$1 = A^S + B^S H_i + E_i^S i \in S$$

(10) $$\frac{h_i}{m_i} = A + BH_i + E_i i \in R \smile S$$

where (10) is clearly the pooled version of (8) and (9).

Starting with the unpooled equations is more convenient. We estimate all these equations by weighted least squares, with the weight on each city being the square root of its at-risk population.

Equation (9) is trivial; the estimates are obvious:

$$\begin{pmatrix} A^S \\ B^S \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

In set S, the rate of homelessness among the at-risk population is independent of the housing market; all at-risk people are homeless.

Equation (8) is considerably more complex. It has no obvious solution, and I am aware of no theory that predicts values for the coefficients. Two simple atheoretical stories give opposite predictions: if the same proportion ρ of at-risk people are homeless in every city in R,

$$\begin{pmatrix} A^R \\ B^R \end{pmatrix} = \begin{pmatrix} \rho \\ 0 \end{pmatrix};$$

while if the number of at-risk people are the same (say, m) in every city in R,

$$\begin{pmatrix} A^R \\ B^R \end{pmatrix} = \begin{pmatrix} 0 \\ 1/m \end{pmatrix}.$$

In general, we have (11)

$$\begin{aligned}
\begin{pmatrix} A^R \\ B^R \end{pmatrix} &= \begin{pmatrix} \sum m_i & \sum m_i H_i \\ \sum m_i H_i & \sum m_i H_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum H_i \\ \sum H_i^2 \end{pmatrix} \\
&= \frac{1}{\sum m_i \sum m_i H_i^2 - \left(\sum m_i H_i\right)^2} \begin{pmatrix} \sum H_i \sum m_i H_i^2 - \sum H_i^2 \sum m_i H_i \\ \sum m_i \sum H_i^2 - \sum H_i \sum m_i H_i \end{pmatrix}
\end{aligned}$$

where all summations are over the set R.

Usually, since (H_i/m_i) is positive and less than one in set R, while H_i is a large positive number (considerably greater than one, that is), you would expect that $A^R > 0$ and $B^R < 1$, but this result is difficult to prove with complete generality. We can prove two somewhat weaker propositions: namely, that if population characteristics matter at all, then the coefficient on housing markets is very small; and that if the sample includes enough people at risk, then the coefficient on housing markets is less than one. This is important because for cities in R, in city-level regressions the coefficient on the housing market is always one.

Proposition 2: If $A^R \geq 0$, then $B^R < \frac{1}{\overline{H}_R}$, where $\overline{H}_R$ is the weighted average of H_i in set R.

Proof: Let over-bars denote weighted means in R. Since a regression line goes through the sample's weighted centroid

$$\overline{\left(\frac{H}{m}\right)} = A^R + B^R \overline{H}_R$$

and so

$$B^R = \frac{\overline{\left(\frac{H}{m}\right)} - A^R}{\overline{H}_R} < \frac{1}{\overline{H}_R},$$

since (H_i/m_i) is less than one in set $R.QED$

Proposition 3: Let μ denote the smallest m_i in R. If μ is sufficiently large, then $B^R < 1$.

Proof: Consider the expression

$$Z = \sum m_i H_i (H_i - \overline{H}_R)$$

where all summations are over $i \in R$. Since the (weighted) variance of H_i, which is

$$\begin{aligned} \sum m_i (H_i - \overline{H}_R)^2 &= \sum m_i [H_i (H_i - \overline{H}_R) - \overline{H}_R (H_i - \overline{H}_R) \\ &= Z - \sum m_i \overline{H}_R (H_i - \overline{H}_R) = Z, \end{aligned}$$

is positive, Z is positive.

Let R_1 denote the subset of R where $H_i \geq \overline{H}_R$ and R_2 its complement. Let $Z_k, k = 1, 2$ denote the corresponding partial sums, where $Z_1 > 0, Z_2 < 0, Z_1 + Z_2 = Z$. Thus

$$\frac{Z_1}{\mid Z_2 \mid} > 1.$$

Let m_1 denote the smallest value of m_i in R_1 and m_2 denote the largest value of m_i in R_2. Note that $\mu \preceq m_k, k = 1, 2$. Then

$$\begin{aligned} Z_1 &= \sum_1 H_i (H_i - \overline{H}_R) m_i = \sum_1 H_i (H_i - \overline{H}_R)(m_i - 1) \frac{m_i}{m_i - 1} \\ &< \sum_1 H_i (H_i - \overline{H}_R)(m_i - 1) \frac{m_1}{m_1 - 1} = X_1 \frac{m_1}{m_1 - 1}. \end{aligned}$$

Similarly

$$\begin{aligned} \mid \quad & Z_2 \mid = \sum_2 H_i (H_i - \overline{H}_R) m_i = \sum_2 H_i (H_i - \overline{H}_R)(m_i - 1) \frac{m_i}{m_i - 1} \\ > & \sum_1 H_i (H_i - \overline{H}_R)(m_i - 1) \frac{m_2}{m_2 - 1} = X_2 \frac{m_2}{m_2 - 1}. \end{aligned}$$

Combining these we obtain

$$\frac{Z_1}{\mid Z_2 \mid} < \frac{X_1}{\mid X_2 \mid} \frac{m_2}{m_2 - 1} \frac{m_1}{m_1 - 1} < \frac{X_1}{\mid X_2 \mid} \left(\frac{\mu}{\mu - 1} \right)^2$$

or

$$\left(\frac{\mu-1}{\mu}\right)^2 \frac{Z_1}{\mid Z_2 \mid} < \frac{X_1}{\mid X_2 \mid}.$$

Thus for any $Z_k, k = 1, 2$, for μ sufficiently large, $X_1 >\mid X_2 \mid$.

From algebra, it is easy to derive that

$$B^R = 1 - \frac{\sum m_1}{\Delta}\left[X_1 + X_2\right],$$

where

$$\Delta = \sum m_i \sum m_i H_i^2 - \left(\sum m_i H_i\right)^2 > 0.$$

Thus $B^R < 1.QED.$

Now consider equation (10), which is the same as equations (8) and (9) except that the two sets of cities are pooled. Let (A^x, B^x) denote the intercept and the slope respectively of the line between the (weighted) centroids of the cities in R and the cities in S. Then

$$\begin{aligned} B^x &= \frac{1 - \overline{(H/m)}_R}{\overline{H}_S - \overline{H}_R} \\ A^x &= \frac{1}{\overline{H}_S - \overline{H}_R}\left[\overline{H}_S \overline{(H/m)}_R - \overline{H}_R\right]. \end{aligned}$$

The numerator of B^x is positive, but the signs of the other terms are indeterminate. If $\overline{H}_S - \overline{H}_R > 0$, then $B^x > 0, A^x < 1$;in the opposite case, $B^x < 0, A^x > 1$. Most importantly, as long as either $\overline{H}_S - \overline{H}_R > 1$ or $\overline{H}_S - \overline{H}_R < 0$, we have $B^x < 1$. With large numbers, one or the other of these cases is virtually certain.

The slope of the pooled equation (10) is a weighted average of the slopes of the two sets, B^R and B^S, and the slope "between the sets," B^x. The weights are proportional to contributions to total (weighted) variance in H_i of the two sets and the difference between the means of the two sets. Specifically, let

$$V = \sum m_i (H_i - \overline{H})^2$$

where the summation is over all cities, denote the total variance. Then the weights are

$$\begin{aligned} w_k &= \frac{1}{V}\sum_k m_i(H_i - \overline{H}_k)^2 k = R, S \\ w_x &= \frac{1}{V}\frac{\sum_R m_i + \sum_S m_i}{\sum m_i}(\overline{H}_R - \overline{H}_S)^2 \end{aligned}$$

where it is easy to see that the sum of weights is one. Then the slope in the pooled equation is

$$B = w_R B^R + w_S B^S + w_x B^x.$$

Since $B^S = 0$, and for large populations, $B^R < 1, B^x < 1$, we have for large populations, $B < 1$. In general, most extant theories say nothing about (A, B).

Finally, an alternative way of estimating (10) is directly, through the usual formulas. This gives us

(12)
$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum m_i & \sum m_i H_i \\ \sum m_i H_i & \sum m_i H_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_R H_i + \sum_S m_i \\ \sum_R H_i^2 + \sum_S m_i H_i \end{pmatrix}$$

where summations without arguments are over all cities.

Having examined the auxiliary equations, we turn to (7), the individual-level equation we are interested in. The usual OLS formula gives

(13)
$$\begin{pmatrix} \widehat{\alpha}^I \\ \widehat{\beta}^I \end{pmatrix} = \begin{pmatrix} \sum m_i & \sum m_i H_i \\ \sum m_i H_i & \sum n H_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_R H_i + \sum_S m_i \\ \sum_R H_i^2 + \sum_S m_i H_i \end{pmatrix}.$$

Comparing this expression with (12), we see that the only difference is in the lower right corner of the first matrix. In (13), this is $\sum nH_i^2$; in (12), it is $\sum m_i H_i^2$. Since $m_i \leq n$ for all i,

$\sum nH_i^2 \geq \sum m_i H_i^2,$

with strict inequality if at least one person is not at risk of homelessness.

The intuition behind this correspondence is that the weighted city-level equation (10) recovers the same coefficients as the individual-level equation (7) with the sample restricted to the at-risk population (with the intercept in (10) being the coefficient on at-risk status in restricted (7)). The dependent variable in (10) is the probability of being homeless if you are at risk; the independent variable is the housing gap. The same is true for equation (7) restricted to the at-risk population. Unrestricted equation (7) just adds in the population that is not at risk, but none of this population is homeless.

Thus, since the lower right component of the first matrix affects only the denominator in the expression for $\widehat{\beta}^I$, we see unambiguously that

$$| \widehat{\beta}^I | < | B |$$

with the extent of the difference being greater the greater the number of not-at-risk people in the sample (weighted by H_i^2). Since in the large sample case, $B < 1$,

$$\widehat{\beta}^I < 1$$

even if all cities are in R (and provided that $B \succeq 0$ if some cities are not in R).

The reason for the attenuation of housing market effects in the unrestricted sample is also simple to understand. The housing market has no effect on people who are not at risk; they are never homeless. Thus pooling the at-risk population with the population not at risk reduces the average effect of the housing market.

As for the coefficient $\widehat{\alpha}^I$ of individual characteristics, the sign of the difference with A is ambiguous, since the lower right component enters into both the numerator and the denominator. If no cities are in set S (which makes the coefficient on population characteristics disappear in the city-level regressions), $\widehat{\alpha}^I$ may still be positive, since if (H_i/m_i) is approximately constant in set R, $A^R = A$ will be positive. Since both the numerator and the denominator of $\widehat{\alpha}^I$ are bigger than the numerator and the denominator respectively of A, if $A = A^R$ is positive, $\widehat{\alpha}^I$ will be positive, too.

4 CONCLUSION

Thus the results in the literature–with individual-level observations, personal characteristics matter and housing markets don't; the opposite with city-level observations–are consistent with a world in the simple-minded theoretical model holds, most cities are in set R, and in those cities the housing gap is roughly proportional to the size of the at-risk population.

This is not, however, the major conclusion I wish to draw from this exercise. Rather, the major conclusion for researchers is that they should think more carefully about the interaction between individual and market characteristics, and not just let the sets of variables "fight it out for themselves." For policy, the conclusion is similar: interaction matters.

The obvious question for future research is whether other phenomena, not just homelessness, work in this fashion. I suspect that they do. Consider child abuse, for instance. Paxson and Waldfogel (1999a, 1999b, 2000) use state level data and find that higher AFDC benefits and lower unemployment rates reduce child abuse, but Berger (2002), using individual data, is unable to replicate the strong relationships that Paxson and Waldfogel found. It is possible to think of child abuse occurring only when a conjunction of unfortunate circumstances occurs–when parents somehow predisposed to abusing their children find themselves operating in an an environment where child abuse is not sufficiently discouraged. The need for a conjunction is what drove the theoretical results in section 2, and so child abuse may be like homelessness in this respect.

REFERENCES

Richard P. Appelbaum, Michael Dolny, Peter Dreier, and John I. Gilderbloom, 1991, "Scapegoating Rent Control: Masking the Causes of Homelessness," Journal of the American Planning Association 57, 153-64.

E. Bassuk, J. Buckner, L. Weinreb, A. Browne, S. Bassuk, R. Dawson, and J. Perloff, 1997, "Homelessness in Female-Headed Families: Childhood and Adult Risk and Protective Factors," American Journal of Public Health 87, 241-48.

Lawrence M. Berger, 2002, Economic Analysis of Child Abuse and Neglect. PhD dissertation, Columbia University School of Social Work.

Cecil Bohanon, 1991, "The Economic Correlates of Homelessness in Sixty Cities," Social Science Quarterly 72, 153-64.

Martha Burt, 1992, Over the Edge: The Growth of Homelessness in the 1980s. New York: Russell Sage Foundation.

Michael Cragg and Brendan O'Flaherty, 1999, "Do Homeless Shelter Conditions Determine Shelter Population? The Case of the Dinkins Deluge," Journal of Urban Economics 46, 377-415.

Dirk W. Early, 1998, "The Role of Subsidized Housing in Reducing Homelessness: An Empirical Investigation Using Micro-Data," Journal of Policy Analysis and Management 17, 687-96.

Dirk W. Early, 1999, "A Microeconomic Analysis of Homelessness: An Empirical Investigation Using Choice-Based Sampling," Journal of Housing Economics 8, 312-27.

Dirk W. Early and Edgar O. Olsen, 1998, "Rent Control and Homelessness," Regional Science and Urban Economics 28, 797-816.

Dirk W. Early and Edgar O. Olsen, 2001, "Subsidized Housing, Emergency Shelters, and Homelessness: An Empirical Investigation Using Data from the 1990 Census," unpublished manuscript, University of Virginia.

Marta Elliott and Lauren Krivo, 1991, "Structural Determinants of Homelessness in the United States," Social Problems 38, 113-31.

Paul Grimes and George Chressanthis, 1997, "Assessing the Effect of Rent Control on Homelessness," Journal of Urban Economics 41, 23-37.

Marjorie Honig and Randall Filer, 1993, "Causes of Intercity Variation in Homelessness," American Economic Review 83, 248-55.

Christopher Jencks, 1994, The Homeless. Cambridge, MA: Harvard University Press.

Brendan O'Flaherty, 1995, "An Economic Theory of Homelessness and Housing," Journal of Housing Economics 4, 13-49.

June Park, 1997, Homelessness in America and Housing Vacancies. PhD dissertation, Columbia University, Department of Economics.

Chris Paxson and Jane Waldfogel, 1999a, "Parental Resources and Child Abuse and Neglect," American Economic Review 89, 239-44.

Chris Paxson and Jane Waldfogel, 1999b, "Work, Welfare, and Child Maltreatment," NBER working paper 7343.

Chris Paxson and Jane Waldfogel, 2000, "Welfare Reforms, Family Resources, and Child Maltreatment." Paper presented at the Joint Poverty Center Conference on Incentive Effects of Tax and Transfer Programs, December 7-8, Washington, D.C.

John Quigley, 1990, "Does Rent Control Cause Homelessness? Taking the Claim Seriously," Journal of Policy Analysis and Management 9, 89-93.

John Quigley, Steven Raphael, and Eugene Smolensky, 2001, "Homeless in America, Homeless in California," Review of Economics and Statistics 83, 37-51.

William Troutman, John D. Jackson, and Robert Ekelund, 1999, "Public Policy, Perverse Incentives, and the Homeless Problem," Public Choice 98, 195-212.

William Tucker, 1989, "America's Homeless: Victims of Rent Control," Backgrounder 685. Washington, DC: The Heritage Foundation.

APPENDIX: A SIMPLE EQUILIBRIUM MODEL OF HOMELESSNESS IN A CITY

Consider a city with n people, of whom m suffer from pathologies. All have identical income, which we normalize to unity, and have identical utility functions if housed. Specifically, if housed, an individual's utility depends on the the quantity k of housing consumed and the quantity x of a numeraire non-housing good:

$$\begin{aligned} u(k,x) &= \ln x + \ln k, \\ s.t, rk + x &\leq 1, \end{aligned}$$

where r is the price of housing. The utility of an individual who is not housed but consumes x of the non-housing good is

$$\ln C + \ln x,$$

where $C = c > 0$ if the individual suffers from pathologies, and $C = 0$ otherwise. Thus those who do not suffer from pathologies are never homeless. The value of c may depend on climate, for instance, or shelter provision (if the latter is considered exogenous).

Let R denote the value of rent at which individuals who suffer from pathology are indifferent between being homeless and being housed. It is easy to show that

$$R = \frac{1}{4c}.$$

Then we can calculate the demand for housing $D(r)$ as a function of the price of housing:

$$\begin{aligned} D(r) &= \frac{n}{2r}, \text{ if } r < R \\ &= z, \qquad 2(n-m)c \leq z \leq 2nc, \text{ if } r = R \\ &= \frac{n-m}{2r}, \text{ if } r > R. \end{aligned}$$

We take the supply of housing $S(r)$ as a linear function of the price of housing:

$$S(r) = sr, s > 0.$$

The supply parameter s may vary from city to city, with institutions and geography.

In this model there are equilibria with no homelessness, and two kinds of equilibria with homelessness. We ignore the equilibria with no homelessness.

In one kind of equilibria with homelessness,

$$r = R,$$

and some individuals with pathologies are homeless but not all. The demand curve has a flat plateau, and these equilibria occur when the supply curve cuts the demand curve on this plateau. The quantity of housing required to house all the people who are homeless in an equilibrium like this is

$$sR - 2nc$$

and each housed person consumes

$$\frac{1}{2R}$$

quantity of housing. Thus the number of homeless individuals is

$$\frac{sR - 2nc}{1/2R} = \frac{s}{8c^2} - n.$$

Write

$$H = \frac{s}{8c^2} - n.$$

Notice that H depends only on housing market variables–the variable m does not appear–and is positively correlated with the observed rent R. The necessary condition for an equilibrium of this type to obtain is that the intersection of demand and supply be on the flat plateau, or

$$H \leq m.$$

Cities with this type of equilibrium would be called housing-rich in the text.

In the other kind of equilibria with homelessness, the demand curve intersects the supply curve to the left of and above the flat plateau. Equilibrium rent is above R, and as a result, all individuals with pathologies are homeless. Homelessness is simply the number of individuals with pathologies, and does not respond to small changes in housing market variables. Cities with this type of equilibrium would be called housing-short in the text. Thus, for instance, the difference between housing-rich and housing-short cities could lie in the steepness of the supply curve, with housing-rich cities having a more elastic supply of housing.

Thus the volume of homelessness in this very simple equilibrium model is described by (1) in the text. Thus (1) does not depend on the existence of irrational or non-equilibrium behavior.