I. Data

Neighborhood amenities used for housing/neighborhood quality parameters

An ideal dataset would contain a set of variables reflecting neighborhood *physical* amenities, such as quality of houses in the neighborhood, presence of parks and open spaces, distance to major landmarks and business centers, and distance to bus, train and subway stations, if applicable. Some housing quality variables are available from Census data, but the other variables may only be constructed, if at all, manually. Thus, the neighborhood characteristics I included in the regressions used to construct neighborhood quality parameters include average housing features, computed them based on Census data at the tract level, and distance with respect to the metro area geographic center, computed using Arc View.

The housing variables are the following:

- fraction of housing units in urban (as opposed to rural) areas
- fraction of housing units with a given number of rooms (1, 2, 3, 4, 5, 6, 7, 8, 9 or more)
- fraction of housing units categorized by units in the structure (1-detached, 1-attached, 2, 3 or 4, 5 to 9, 10 to 19, 20 to 49, 50 or more, mobile homes or trailers, others)
- fraction of housing units with a given number of bedrooms (0, 1, 2, 3, 4, 5 or more)
- fraction of houses of a given age (0-1, 2-5, 6-10, 11-20, 21-30, 31-40, 41-50, 50 or more years old)

Although the Census provides other housing characteristics as well, there is virtually no variation for them among the tracts in my sample.

II. Computation of Rental Values

Obtaining the average rental value

The dependent variable in the regression used to obtain the neighborhood quality parameters is the 1990 log of tract average rental value. The Census provides information to compute the average value for owner-occupied housing units and the average rent for renter-occupied housing units. Hence, I calculated the tract average rental value as follows:

(II.1) $ARV = p \times uc \times V + (1 - p)R$

where:
ARV = average rental value
p = proportion of specified owner-occupied housing units
uc = user cost rate
V = average value for specified owner-occupied housing units
R = average annual rent for specified renter-occupied housing units

The user cost rate equals:

(II.2) $uc = (1 - t_y)(i + t_p) + m + d - a$

where:
t_y = income tax rate
i = interest rate on mortgage loans
t_p = property tax rate
m = maintenance, as a fraction of the value of a housing unit
d = risk premium
a = expected price appreciation for the housing unit.

Data used to compute average rental value

The data used to compute the 1990 average rental value for a given geographic unit proceeds from the sources listed below.

1. Data on the proportion of specified owner-occupied housing units, the average value for specified owner-occupied housing units and the average rent for specified renter-occupied housing units comes from the Census data at the desired geographic level.

2. The income tax rate is the addition of federal and state rates. To compute the income tax rate, I take the income distribution for the metropolitan area for 1990 (for households with children enrolled in school) and compute the midpoint for each of the income brackets defined in the data. For each midpoint, I obtain the 1990 effective federal and state marginal rates from the TaxSim calculator (http://www.nber.org/~taxsim/taxsim-calc4/index.html).[1] Then I compute the weighted average of the marginal tax rates over all the midpoint income levels of the metropolitan area. This average is my measure for the income tax rate for the metropolitan area.

3. The (effective) interest rate of mortgage loans comes from the Federal Housing Finance Board. This rate corresponds to single-family mortgages in the metropolitan area in 1990.

4. The metropolitan area-average residential property tax rate came from Government of the District of Columbia (1990)).

5. Following Green (1996) and Green and Hendershott (1996), I set the maintenance rate to 0.03 and the risk premium to 0.04.

6. I computed the expected rate of housing price appreciation as the average annual rate of price change between 1979 and 1989 for the metropolitan area. I used the house price

[1] TaxSim operates on the basis of hypothetical individuals. I assume family households with two children.

indexes from the Office of Federal Housing Enterprise Oversight for 1979 through 1989. The OFHEO house price index applies to single-family detached properties, and uses a weighted-repeat sales methodology.

III. Estimation Strategy

Estimating variances using the first-stage residuals

The first-stage residuals from equation (8) in the text are used to compute $\hat{\sigma}_j^2$ as follows (recall that *j* and *k* denote variables, such as income and housing value, and *i* denotes an observation or district). Take the residual $e_{ji} = y_{ji} - \hat{y}_{ji} \approx y_{ji} - E(y_{ji})$. Then, $E(e_{ji}^2) \approx V(y_{ji}) = \frac{\sigma_j^2}{n_i}$, which implies: $E\left(\sum_{i=1}^{D} e_{ji}^2\right) = \left(\frac{1}{n_1} + \frac{1}{n_2} + ... + \frac{1}{n_D}\right)\sigma_j^2$. A natural estimator of σ_j^2 is $\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{D} e_{ji}^2}{\sum_{i=1}^{D} \frac{1}{n_i}}$, which yields $\hat{\sigma}_{ji}^2 = \frac{\hat{\sigma}_j^2}{n_i}$.

In a similar fashion, $e_{ji}e_{ki} = \left(y_{ji} - \hat{y}_{ji}\right)\left(y_{ki} - \hat{y}_{ki}\right) \approx \left(y_{ji} - E(y_{ji})\right)\left(y_{ki} - E(y_{ki})\right)$. Then, $E(e_{ji}e_{ki}) \approx C(y_{jki}) = \frac{\sigma_{jk}}{n_i}$, where *C* denotes "covariance." This implies $E\left(\sum_{i=1}^{D} e_{ji}e_{ki}\right) = \left(\frac{1}{n_1} + \frac{1}{n_2} + ... + \frac{1}{n_D}\right)\sigma_{kj}$. A natural estimator of σ_{jk} is $\hat{\sigma}_{jk} = \frac{\sum_{i=1}^{D} e_{ji}e_{ki}}{\sum_{i=1}^{D} \frac{1}{n_i}}$, which yields $\hat{\sigma}_{jki} = \frac{\hat{\sigma}_{jk}}{n_i}$.

Standard Errors

I compute standard errors according to the usual expression for Feasible Generalized Non-Linear Least Squares:

(III.1) $V(\hat{\theta}) = \left(Z' \sum Z\right)^{-1}$

The construction of the Z and $\sum$ matrices is explained below.

Z matrix

First, *Z* is the following $4D \times 4$ matrix:

(III.2) $Z = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{pmatrix}$

where Z_1, Z_2 , Z_3 and Z_4 are each of size $D \times 4$. In particular, $Z'_{ji} = \dfrac{\partial \hat{y}_{ji}(\theta)}{\partial \theta}$ is the 4×1 vector of derivatives of $\hat{y}_{ji}$ with respect to the parameters.

$\sum$ Matrix

Second, $\sum$ is the following symmetric $4D \times 4D$ matrix:

(III.3) $\sum = \begin{pmatrix} \sigma_{11}\bar{I} & \sigma_{12}\bar{I} & \sigma_{13}\bar{I} & \sigma_{14}\bar{I} \\ & \sigma_{22}\bar{I} & \sigma_{23}\bar{I} & \sigma_{24}\bar{I} \\ & & \sigma_{33}\bar{I} & \sigma_{34}\bar{I} \\ & & & \sigma_{44}\bar{I} \end{pmatrix}$

In this matrix, the σ_{jk}'s are estimated as described above, and $\bar{I}$ is a diagonal matrix whose (i,i) element is equal to ($1/n_i$).

References

Government of the District of Columbia, Department of Finance and Revenue, Office of Economic and Tax Policy (1990). *Tax Rates and Tax Burdens in DC: A Nationwide Comparison*. Washington, DC.

Green, R. (1996). "Should the Stagnant Homeownership Rate Be a Source of Concern?". *Regional Science and Urban Economics*, vol. 26, 337-368.

Green, T. and P. Hendershott (1996). "Age, Housing Demand, and Real House Prices". *Regional Science and Urban Economic*s, vol. 26, 465-480.