Housing Quality in the Mid-Hudson Valley

by Dimitrios A. Giannias

I. Introduction

This paper analyzes the real estate market of the Mid-Hudson Valley, New York, to estimate housing quality. Cross section data on real estate sales in the Mid-Hudson Valley are used to estimate the housing quality equation. This equation is used to provide a housing quality based ranking of several municipalities, offering a picture of the real estate market within various communities of the region. Unlike the previous work in the area, e.g., Blomquist et. al. (1985 and 1988), Roback (1982 and 1988), and Rosen (1979), the method employed in this paper allows researchers to investigate how a ranking will be affected by changes in the distribution of housing characteristics and/or the distribution of local amenities.

Previous work in this area (for example, see the papers mentioned in the previous paragraph) defines a quality index that is a linear function of local amenities and the work uses that index to rank urban areas; the contention is that the well-being of economic agents depends (among other factors) upon local amenities. The weights assigned to these amenities are linear functions of the amenities' implicit (non-market) prices from the housing and/or labor market. To derive these weights, the features of the hedonic housing price and wage (if needed) functions are empirically approximated using...
fitting criteria to derive them. This provides the flexibility of letting the data determine the price and/or wage equations at the cost of not being able to test whether the assumed functional forms are consistent among themselves and the underlying economic structure. In addition to that, this method does not provide the equilibrium price and/or wage equations. As a result, the changes in the implicit (non-market) prices of amenities that are implied by changes in exogenous parameters (e.g., the mean of the population density distribution) cannot be predicted. The latter implies that in these cases a researcher would not be able to find the changes in a ranking of urban areas. This paper follows a different method.

This study makes prior assumptions about the characteristics of the economic agents interacting to form the equilibrium, uses that to derive the form of the equilibrium housing function, and then estimates only that. Imposing these prior restrictions helps by providing the additional theoretical information that is essential in analyzing the housing market, estimating the quality index equation, and creating a housing quality based ranking of various communities in the Mid-Hudson Valley.

Section II introduces the theoretical model. This model assumes that the income distribution and the supply for housing characteristics are exogenous and that consumers utilize only one house. However, the model can be extended to relax these assumptions; see Giannias (1987). The model is estimated and tested in Section III, and a housing quality ranking is discussed in Section IV. Concluding remarks are given in Section V.

II. The Theoretical Model

I consider a competitive economy in which individuals consume one unit of the differentiated good housing, and the numeraire good, x. Consumer preferences are described by a utility function $U(h,x;a)$, where $h$ is the housing quality, $a = [a_1, a_2, \ldots, a_n]$, and $a_i$ is a utility parameter that differentiates consumers for all $i = 1, 2, \ldots, n$. It is assumed that consumers are able to move costlessly within a community and across communities. In equilibrium, there is a housing quality distribution for each community. Comparison of these distributions can provide a housing quality-based ranking of communities.

The differentiated good housing is described by a bundle of attributes $v = [v_1, v_2, \ldots, v_m]$, where $v_i$ is a housing characteristic for all $i = 1, 2, \ldots, m$. The elements of $v$ are assumed to be the following: SQFT
is the size of the house in square feet; BRMS is the number of bedrooms; BATHS is the number of bathrooms; AGE is the age of a house (1988 - year built); LOT is the size of the lot in acres; STORY is the number of floors; and POPDEN is the population density of a community (population per square mile). Specific distributional assumptions have been made about the supply for \( v \) and \( z = [a I] \), where \( I \) is consumer income\(^1\).

Consumers are assumed to maximize utility. The functional form of the utility function and the optimization problem that consumers solve are given in Appendix A. When consumers choose housing, they consider the whole bundle of attributes \( v \). This is mapped into an index that defines housing quality. Since utility depends on \( h \), it makes sense within our framework to assume that in equilibrium the price equation is a function of \( h \). The housing quality is assumed to be a function of the following form\(^2\):

\[
h = e_0 \text{SQFT} + e_1 \text{BRMS} + e_2 \text{BATHS} + e_3 \text{AGE} + e_4 \text{LOT} + e_5 \text{STORY} + e_6 \text{POPDEN}.
\]

Solving the utility maximization problem to obtain the demand for \( h \) and substituting it into the equilibrium condition, namely, Aggregate Demand for \( h \) = Aggregate Supply for \( h \), for all \( h \), it is obtained\(^3\) that, for the economy described above, the equilibrium price equation is linear in housing quality, that is,

\[
P(h) = q_0 + q_1 h.
\]

The exact relationship between the parameters of the price equation and the exogenous parameters of the model are given in Appendix B.

### III. Estimation of the Model

Given the results of the previous section and assuming an additive error term, I can substitute the housing quality equation into the equilibrium price equation to obtain that the equilibrium price equation\(^4\) is equivalent to:

\[
P = b_0 + b_1 \text{SQFT} + b_2 \text{BRMS} + b_3 \text{BATHS} + b_4 \text{AGE} + b_5 \text{LOT} + b_6 \text{STORY} + b_7 \text{POPDEN} + u
\]

where \( u \) is the econometric error; \( b_i \) is a parameter to be estimated; \( i = 0, \ldots, 6 \); and \( b_j = q_i e_j \) for \( j = 1, \ldots, 6 \). Note that the
relationships among the exogenous parameters of the model and the price equation parameters that are given in Appendix B imply that the following is satisfied:

\[ e_j = \frac{b_j}{b_1} \text{ for } j = 1, \ldots, 6 \]  

(2)

Given estimates of the price equation parameters, (2) can be used to obtain estimates of the parameters of the housing quality equation.

The equilibrium price equation is estimated by ordinary least squares using 1988 real estate data on housing prices, SQFT, BRMS, BATHS, AGE, LOT, STORY, and POPDEN for various communities across the Mid-Hudson Valley. The results are given in Table 1. Table 1 and equation (2) imply that the quality of life equation is the following:

\[ h = \text{SQFT} + 1271.6 \text{ BRMS} + 1093.1 \text{ BATHS} - 17.34 \text{ AGE} + 765.9 \text{ LOT} + 1695.5 \text{ STORY} + 0.147 \text{ POPDEN} \]  

(3)

To see if the model makes a significant contribution to explaining the data, the hypothesis that all the coefficients of equation (1) equal zero is tested. This hypothesis is rejected at the one percent significance level. To investigate whether the price equation is misspecified by the omission of some variables, a Ramsey test, see Ramsey (1969), is applied on (1). This test provides evidence that there are not any variables omitted from (1).

**TABLE 1**

**THE PRICE EQUATION**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COEFFICIENT</th>
<th>STANDARD ERROR</th>
<th>T-STATISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>-5701.572</td>
<td>17460.97</td>
<td>-0.3265323</td>
</tr>
<tr>
<td>SQFT</td>
<td>15.35109</td>
<td>8.886736</td>
<td>1.727416</td>
</tr>
<tr>
<td>BRMS</td>
<td>19520.55</td>
<td>6411.557</td>
<td>3.044588</td>
</tr>
<tr>
<td>BATHS</td>
<td>16779.94</td>
<td>6935.449</td>
<td>2.419446</td>
</tr>
<tr>
<td>AGE</td>
<td>-266.1433</td>
<td>156.9363</td>
<td>-1.695869</td>
</tr>
<tr>
<td>LOT</td>
<td>11757.55</td>
<td>1785.322</td>
<td>6.585675</td>
</tr>
<tr>
<td>STORY</td>
<td>26027.68</td>
<td>12215.67</td>
<td>2.130681</td>
</tr>
<tr>
<td>POPDEN</td>
<td>2.259910</td>
<td>1.574723</td>
<td>1.435116</td>
</tr>
</tbody>
</table>

NUMBER OF OBSERVATIONS = 104

R^2 = .72
IV. A Housing Quality-Based Ranking of Communities in the Mid-Hudson Valley, New York

In equilibrium, there is a housing quality distribution for each community. The results of the previous section can be used to obtain an estimate of the mean of this distribution for each of the communities in the Mid-Hudson Valley included in this study. The mean of the housing quality in each community and a housing quality based ranking are given in Table 2. To obtain the mean housing quality value of a community, I substitute in (3) the community's POPDEN value and its sample means of SQFT, BRMS, BATHS, AGE, LOT, and STORY. To facilitate comparisons, all rankings are scaled from 0 to 100.

### TABLE 2
**HOUSING QUALITY-BASED RANKING**

<table>
<thead>
<tr>
<th>RANK</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLEASANT VALLEY</td>
<td>1</td>
</tr>
<tr>
<td>LAGRANGE</td>
<td>2</td>
</tr>
<tr>
<td>GARDINER</td>
<td>3</td>
</tr>
<tr>
<td>EAST FISHKILL</td>
<td>4</td>
</tr>
<tr>
<td>WAPPINGER</td>
<td>5</td>
</tr>
<tr>
<td>HYDE PARK</td>
<td>6</td>
</tr>
<tr>
<td>STANFORD</td>
<td>7</td>
</tr>
<tr>
<td>POUGHKEEPSIE</td>
<td>8</td>
</tr>
<tr>
<td>FISHKILL</td>
<td>9</td>
</tr>
<tr>
<td>RHINEBECK</td>
<td>10</td>
</tr>
<tr>
<td>CLINTON</td>
<td>11</td>
</tr>
<tr>
<td>BEACON</td>
<td>12</td>
</tr>
<tr>
<td>NEW PALTZ</td>
<td>13</td>
</tr>
<tr>
<td>PLATTEKILL</td>
<td>14</td>
</tr>
<tr>
<td>UNION VALE</td>
<td>15</td>
</tr>
<tr>
<td>PINE PLAINS</td>
<td>16</td>
</tr>
<tr>
<td>RED HOOK</td>
<td>17</td>
</tr>
</tbody>
</table>

V. Conclusions

The empirical results of Section IV show that Pleasant Valley, LaGrange, and Gardiner occupy the top three positions of the rank-
ing and Union Vale, Pine Plains, and Red Hook the bottom. Table 2 also shows that the housing quality differences between New Paltz and Plattekill, Fishkill and Rhinebeck, Stanford and Poughkeepsie, and East Fishkill and Wappinger are relatively small.

**Appendix A**

The utility function is assumed to be a quadratic of the following form:

\[
U(h, x; a) = k_0 + (k_1 + k_2 a') h + 0.5 k_3 h^2 + k_4 x h
\]

where \( k_i \) is a utility parameter; \( i = 0, 1, 3, 4; k_2 \) is a \((1 \times n)\) vector of utility parameters; and \( a' \) is the transpose of \( a \).

A consumer solves the following optimization problem:

\[
\max U(h, x; a)
\]

with respect \( h, x \)

subject to \( I = P(h) + x \) and

\( P(h) = q_0 + q_1 h \)

where \( P(h) \) is the equilibrium price equation (giving the market value of a house as a function of its quality).

**Appendix B**

Solving the utility maximization problem to obtain the demand for \( h \) and substituting it into the equilibrium condition, namely, Aggregate Demand for \( h = \) Aggregate Supply for \( h \), for all \( h \), it is obtained that for the economy described above the equilibrium price equation is:

\[
P(h) = q_0 + q_1 h
\]

where

\[
q_i = [k_1 + r m(z)' - (2 k_3 q_1 - k_4) m(h)]/k_4
\]

\[
m(h) = m(SQFT) + e_1 m(BRMS) + e_2 m(BATHS) + e_3 m(AGE) + e_4 m(LOT) + e_5 m(STORY) + e_6 m(POPDEN)
\]

\[
V(h) = e V(v) e'
\]

\[
A = [r V(z) r' / V(h)]^{0.5}
\]

\[
r = [k_2 k_3]
\]

\[
e = [e_0 \ldots e_6]
\]

\[
v = [SQFT \ BRMS \ BATHS \ AGE \ LOT \ STORY \ POPDEN],
\]

\( m(t) \) is the mean of a variable \( t \), for all \( t \), and \( V(s) \) is the variance-covariance matrix of a vector of variables \( s \), for all \( s \).
The $q_0$ and $q_1$ parameters of the equilibrium price equation are unique. There are two solutions that satisfy the equilibrium condition. However, one of them is rejected because it does not satisfy the second order condition for the utility maximization problem.

Notes

1. The supply for $v$ is assumed to follow an exogenous multinormal distribution. $z = [a\ I]$ is also assumed to follow an exogenous multinormal distribution.

2. Without loss of generality, the housing quality is normalized by setting $c_n = 1$.

3. The proof is identical to the proof of Proposition 1 of Giannias (1987). The general strategy of the proof was introduced by Tinbergen (1959) and extended by Epple (1984).

4. The distributional assumptions about $v$ imply that the equilibrium price distribution is normal.

5. In addition to this test, the joint normality of prices and of the characteristics that describe a house is tested. D'Agostino and Pearson (1973) proposed an omnibus test that uses a statistic that is equal to the sum of standardized normal equivalents to the sample skewness, $c_1$, and kurtosis, $c_2$. To test the null hypothesis that the error term of (1) is normally distributed the following composite test statistic is used: $(N/6)(C_1)^2 + (N/24)(C_2 - 3)^2$, where $N$ is the number of observations. This composite test statistic is distributed as a $\chi^2$ with two degrees of freedom and provides evidence in favor of the null hypothesis. This normality test also implies that the price equation, given $v$, is linear in $v$.

References


