Improving the Accuracy of Downtown Land Assessment

Richard Ashley, Florenz Plassmann,
and Nicolaus Tideman

**Lincoln Institute of Land Policy
Working Paper**

Lincoln Institute Product Code: WP99RA1

Abstract

A tax that is levied on the value of land alone, irrespective of any improvements to land, does not cause any economic distortions. Levying such a tax requires an ability to estimate the value of land separately from improvements, which may be particularly difficult for downtown parcels, because almost all parcels are already improved, so there will be few sales of near-by vacant lots that would permit inferences about the value of land in the area. This paper provides an initial effort to assess the value of commercial land in downtown Portland, Oregon, by using a combination of a hedonic model of the value of improvements and a quadratic spatial smoothing technique for the value of land. Compared to the official assessments of Portland assessors, our assessments of all property values are only slightly less accurate, while our assessments of the value of vacant parcels are more accurate.

About the Authors

Richard Ashley is associate professor of economics at Virginia Polytechnic Institute and State University, where he has taught since 1981. He received his Ph.D. from the University of California at San Diego. His main areas of research are the econometrics of time series and non-linear models.

Contact Information:
Department of Economics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
tel: (540) 231-6220
email: ashleyr@vt.edu

Florenz Plassmann is assistant professor of economics at State University of New York at Binghamton, where he has taught since 1999. He received his Ph.D. from Virginia Polytechnic Institute and State University. His main areas of research are applied econometrics and the economics of land value taxation.

Contact Information:
Department of Economics
SUNY Binghamton
P.O. Box 6000
Binghamton, NY 13902-6000
tel: (604) 777-4353
email: fplass@binghamton.edu

Nicolaus Tideman is professor of economics at Virginia Polytechnic Institute and State University, where he has taught since 1974. He received his Ph.D. from the University of Chicago. His main areas of research are public choice, economic justice, and the economics of land value taxation.

Contact Information:
Department of Economics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
tel: (540) 231-7592
email: ntideman@vt.edu

Contents

Introduction	1
Location	2
Data	2
Map 1: General Location of the Downtown Portland Area Relative to the Rest of the City	2
Map 2: The Boundaries of Our Study Area	3
Analysis	4
Results	8
Conclusion	9
Reference	10
Map 3: Estimated Land Values in Our Study Area in 1986 Dollars	11
Map 4: Estimated Land Values in Our Study Area, Looking South from Glisan Street	12
Table 1: Coefficients of Variation within the Sample	13
Table 2: The Relative Inaccuracy of Estimates and Assessments of of Vacant Land within the Sample	14
Table 3: Out-of-Sample Accuracy	15
Figure A: Prices in Time Sequence	16
Figure B: The Log of Price in Time Sequence	17
Figure C: The Log of Real Price in Time Sequence	18
Figure D: The Log of Real Price Plotted Against the Log of Parcel Size and a Proportional Relationship Between Price and Size	19
Figure E: The Log of Real Price per Square Foot in Time Sequence	20
Figure F: The Log of Real Price per Square Foot and the Fitted Values of Estimate 1	21

Figure G: Ordered Residuals from Estimate 1 22

Figure H: The Log of Price per Square Foot and the Fitted Values of Estimate 2 23

Figure I: The Frequency of Residuals from Estimate 2 and a a Fitted Normal Distribution 24

Figure J: Residuals from Estimate 2 in Time Sequence 25

Figure K: Residuals from Estimate 2 Plotted Against Age of Structures 26

Figure L: The Frequency of Residuals from Estimate 3 and a Fitted Normal Distribution 27

Figure M: Residuals from Estimate 3 in Time Sequence 28

Appendix 1: The Data 29

Improving the Accuracy of Downtown Land Assessments

Introduction

This is a report on research, sponsored by the Lincoln Institute of Land Policy, to measure the accuracy with which downtown land can be assessed. Knowledge of the value of downtown land is required for successful implementation of a tax on land value. A tax on land value has the advantage that, unlike most other sources of public revenue, it is non-distorting and can therefore be employed without discouraging economic activity.

The motivation for the research comes from the pessimism expressed by Edwin Mills regarding the possibility of adequate assessment of developed commercial land. He said:

> There is no prospect of a hedonic equation that would be adequate to assess site values of developed residential properties; much less a prospect of an equation that could assess site values of developed commercial property; and there is simply no other way to estimate site values of developed properties (1999:47).

Mills's conclusion seemed to us premature. The question of how accurately developed commercial land can be assessed is an empirical one, and we are not aware of previous published studies that address the question. This report describes the results of an initial effort to measure and improve the accuracy with which downtown commercial land is and can be assessed.

We developed a combination of a simple hedonic model of the value of improvements plus a quadratic smoothing technique that estimates property values only slightly less accurately than the assessor, on the 81 sales that were used to develop the model. Four of these sales were of vacant or nearly vacant land. On these sales, our model performed noticeably better than the assessor, though slightly worse than on improved sites. We found five sales that occurred in our sample area after the period for which we initially collected data. On these sales our model had better accuracy than during the sample period and slightly better accuracy than the assessor. We found no sales of vacant land in the post-sample period.

Our basic conclusions are as follows. There are relatively few sales of downtown commercial land in the city we studied. Nevertheless, what information is available can be used in interesting ways to estimate land values. Whether these techniques provide land value assessments that are sufficiently reliable is a question that can only be answered with further research. The degree of success that we achieved suggests that such research would be worthwhile.

The remainder of the report consists of a brief section on our reasons for choosing the location we chose (Portland, Oregon), a section describing the data we employed, a section that analyzes the data, and a brief concluding section.

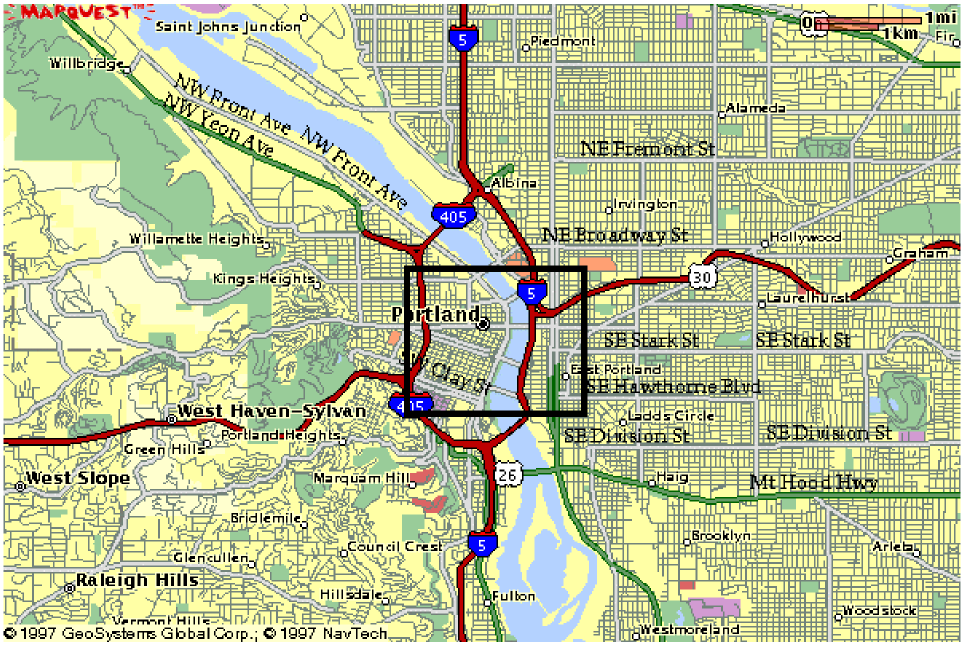
Location

We chose to study Portland, Oregon because it is a fairly large city (480,000 persons in the city; 1.75 million in the PMSA) with a substantially fully built-up downtown area. Our preliminary inquiries led us to believe that the data that we would need to undertake the study would be available. One of us (Nicolaus Tideman) had gone to college in Portland and knew friends there who would help gather information.

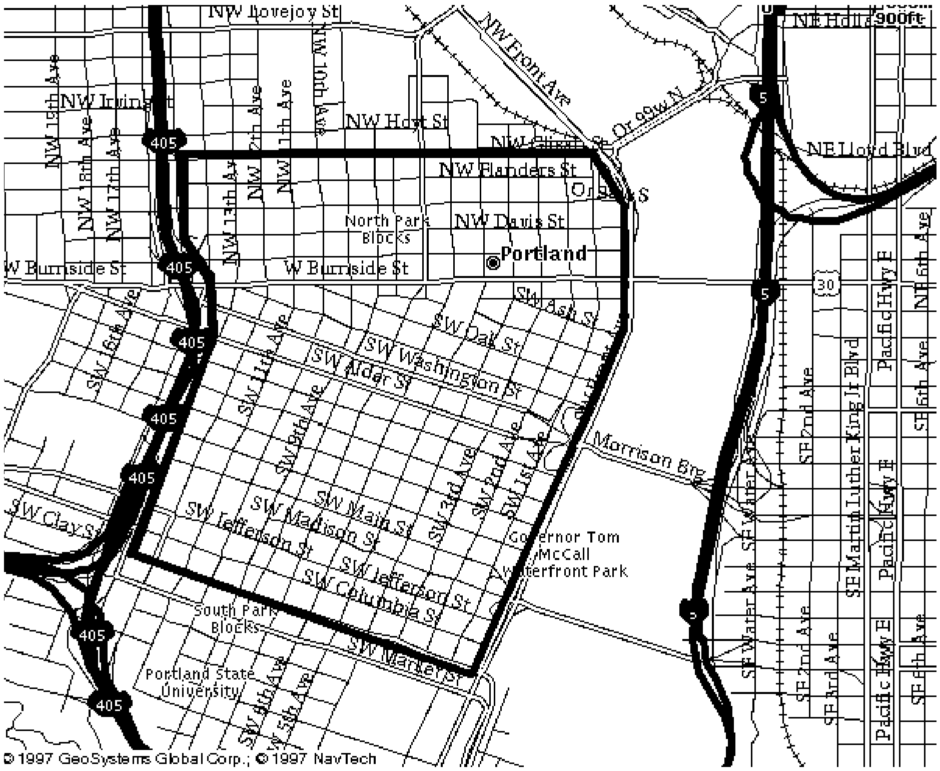
Data

There are at least six companies that sell the type of data that might be used for land value studies. We opened an account with one of these, Data Quick, to obtain data for Portland. Data Quick permits its customers to download information for all sales within any specified distance of any chosen address. We obtained data for all sales from February 1, 1987 to September 22, 1998, within about a mile of a point that we thought might be the center of downtown Portland (SW Fourth and Washington Streets). Later we restricted the data to a roughly trapezoidal area bounded by NW Glisan Street on the north, the Willamette River on the east, SW Clay Street on the south, and highway I-405 on the west. Map 1 shows the general location of the downtown Portland area relative to the rest of the city. Map 2 shows the boundaries of our initial study area.

Map 1: General Location of the Downtown Portland Area Relative to the Rest of the City


MAPQUEST
Saint Johns Junction
Piedmont
Willbridge
NW Front Ave
NW Yeon Ave
NW Front Ave
Alameda
NE Fremont St
Albina
Irvington
Willamette Heights
NE Broadway St
Hollywood
Graham
Kings Heights
Laurelhurst
Portland
SE Stark St
SE Stark St
SW Clay St
East Portland
SE Hawthorne Blvd
Ladds Circle
West Haven-Sylvan
Portland Heights
SE Division St
SE Division St
Green Hills
West Slope
Marquam Hill
Haig
Mt Hood Hwy
Bridlemile
Brooklyn
Council Crest
Arleta
Glencullen
Raleigh Hills
Hillsdale
Fulton
Woodstock
Westmoreland
1mi
1km
© 1997 GeoSystems Global Corp.; © 1997 NavTech

Map 2: The Boundaries of Our Study Area



The data from Data Quick provided us with much of what we needed—address, date of sale, selling price, land use, zoning, buyer's name, and a number of characteristics that we did not use.

From the web site for the City of Portland we obtained a zoning map. We decided to restrict our analysis to downtown property that was zoned commercial (Cxd) and was neither a historical building nor in a historical district. We found 93 such sales in our data base. One of these was a sale of a building without the land beneath it, and we excluded this observation. The remaining 92 observations are shown in Table 4.

An important datum about parcels that was not available from the commercial data base was the number of square feet of the structure on the parcel. This information was available from the Portland Department of Buildings, which retrieved the data we needed for a very reasonable fee.

A second important piece of information that we worked hard to obtain was the exact dimensions and location of each parcel. To obtain this information, we purchased assessment maps from Data Quick. These show the location and dimensions of all parcels. From the Yahoo internet map service we learned the approximate location of each address, and from various clues (number of square feet, lot number within a block, etc.) deduced which particular parcel must be the one that was sold in any particular sale.

To give each parcel coordinates, we spliced together the assessment maps, printed coordinate lines on transparencies, and then photocopied the spliced map with the transparencies on top. This permitted us to assign coordinates to the centroid of each sale, with an accuracy of about 10 feet. These coordinates are shown in Table 4.

In three instances in our final data set, the same buyer bought two properties on the same date. In these cases, the reported price was generally the same for each parcel. Therefore we could rely on such numbers to be true sale prices only by aggregating the parcels into a single sale. We omitted an observation if one or more of the characteristics of the parcel or the sale that we used was not available, or if the observation consisted of an aggregation of parcels at widely separated locations.

Great care is needed to insure that the data used are appropriate. Sometimes the number of square feet in a parcel is rounded off to the nearest hundred. Sometimes it is reported to the nearest square foot. Sometimes the assessment data make it clear that the reported data are wrong. In a few cases, Yahoo put an address in a location that was inconsistent with nearly all of the other locations. Some of the data were supposedly single-family houses in the middle of downtown Portland. But they had parcel numbers that were inconsistent with the surrounding addresses, and in fact there were no such structures at the supposed addresses. Friends in Portland helped us to verify facts that were ambiguous or seemed anomalous in the data base, helping us ensure that nonsensical data were not incorporated into the analysis.

Analysis

Figure A shows the prices that were paid for properties in the order in which the sales occurred. The high degree of skewness in theses numbers makes it difficult to imagine seeing any meaningful patterns. Therefore we begin by taking the logarithms of the selling prices, as shown in Figure B.

Economic theory says that prices cannot be compared meaningfully until they have been adjusted for changes in the overall price level. To adjust prices of commercial property in Portland, we used the index of the price of shelter in Washington and Oregon, provided by the Bureau of Labor Statistics on their web site. The logarithm of real price, that is, the logarithm of the ratio of selling price to the price index for shelter in Washington and Oregon is shown in Figure C.

Although a property that is twice as large should sell for twice as much in general, it will not always be the case. If parcels have historically been subdivided into parcels of a size that is currently uneconomically small, there will be a premium for parcels that have been assembled into a more economical size. Nevertheless, we expect a reasonably proportional relationship between parcel size and price. Figure D shows a plot of the logarithm of real price against the logarithm of parcel size (in square feet), and a fitted proportional relationship between real price and parcel size. The fitted proportional relationship is obtained by an equation that specifies that the estimated log of real price is the log of parcel size plus 3.93 (which is the difference between the mean of the log of real price and the log of parcel size). The rough proportionality shown in this figure justifies taking the log of real price per square foot of land as the variable to be explained. Figure E shows a plot of the log of real price per square foot of land in time sequence.

We expect that the most important determinant of the log of real price per square foot will be the amount of building per square foot of land. Accordingly, we define the "coverage" of a parcel as the ratio of square feet of building to square feet of land. Figure F shows the log of real price per square foot and a fitted value plotted against coverage. The fitted value is a kinked line that is fitted by maximum likelihood, assuming a gamma distribution of residuals with a thick tail below the mean, corresponding to the pattern that can be seen in Figure F. That is, the four parameters of the kinked line (two intercepts and two slopes) and the two parameters of the gamma distribution are chosen to maximize the sum of the likelihoods of all observations. The likelihood of each observation is computed as $\Gamma[b - (A - E)]$, where b is the mean of a gamma distribution, A is the actual real price per square foot for the observation, E is the expected real price per square foot, as specified by the kinked line, and Γ is the gamma density function. (We allowed the fitted shape to take the form of a hyperbola, but it degenerated into the kinked line.) Maximization of the likelihood then produces Estimate 1:

First intercept:	2.506
First slope:	0.656
Second intercept:	3.438
Second slope:	0.191
Mean of the gamma distribution of residuals (upper limit on positive residuals)	2.707
Variance of the gamma distribution of residuals:	0.984
Log of the Likelihood:	-125.840

In other words, our best fit of the log of real price per square foot as a function of coverage (*Cov*) is Min(2.506 + 0.656 *Cov*, 3.438 + 0.191 *Cov*), and our estimate of the distribution of the residuals from this relationship is that (2.707 – Residual) is distributed gamma, with a mean of 2.707 and a variance of 0.984.

One implication of this result is that the average value of the log of real price per square foot that we attribute to unimproved land is 2.506, which implies a median real price per square foot of unimproved land of exp(2.506), or $12.26 per square foot. Note that the units of this result are "real dollars," that is, dollars of 1986, when the price index was

1.00. The price index in 1998 was 1.912, implying a median 1998 price per square foot of $23.44 for unimproved commercial land in downtown Portland.

Figure G shows the residuals from the above estimation, ordered by size. This distribution appears to have two sharp discontinuities, from –1.5 to –1 and from 1.5 to 2. This suggests that the nine outliers below –1.5 and the two above 2 may be generated by phenomena that are outside the scope of our model. For example, a property might be sold in serious disrepair, or sold for non-cash consideration in addition to the cash price that is reported. It is customary in analyses of property values to disregard such outliers. There is a good reason for doing so. Ultimately, we wish to explain the spatial variation in prices, and estimates of the price of land at particular places would be highly influenced by the inclusion of such outliers, because there is a relatively small number of properties that are sold in the vicinity of any given property. Therefore we decided to exclude these 11 outliers from the rest of our analysis, recognizing that if such a sale occurs among the out-of-sample sales on which we later measure the predictive accuracy of our estimating equation, we will estimate the price of that sale very badly. But we are willing to accept that cost for the sake of the improved estimation of the preponderance of out-of-sample sales.

Once the outliers are removed, we expect that the residuals will have a distribution that can be characterized as normal. Figure H shows the log of real price per square foot and a fitted value plotted against coverage for the sample with the 11 outliers removed. The fitted value is again a kinked line that is fitted by maximum likelihood, assuming this time that the residuals have a normal distribution. This produces Estimate 2:

First intercept:	2.911
First slope:	0.494
Second intercept:	3.498
Second slope:	0.219
Variance of the distribution of residuals:	0.338
Log of the likelihood:	-70.963

In other words, with the outliers removed, our best fit of the log of real price per square foot as a function of coverage is Min(2.911 + 0.494 *Cov*, 3.498 + 0.219 *Cov*), and we estimate that the residuals from this relationship have a normal distribution with a mean of zero and a variance of 0.338.

This result implies that the average value of the log of real price per square foot that we attribute to unimproved land is 2.911, which implies a median real price per square foot of unimproved commercial land in downtown Portland of exp(2.911), which is $18.37 per square foot in 1986 dollars, or $35.13 in 1998 dollars.

Figure I shows the distribution of the residuals from Figure H and densities of a corresponding normal distribution. The relationship between the two indicates that the assumption of a normal distribution of residuals is acceptable.

Figure J shows the residuals from Figure H in time sequence. The absence of obvious patterns related to time provides assurance that the influence of time is adequately dealt with by the process of dividing prices by an index of the cost of shelter.

Estimate 2 uses only coverage. A building's age is an additional piece of information that is relevant to the value of a building and is reasonably easy to obtain. Figure K shows the residuals from Estimate 2 plotted against the age of buildings. This figure suggests that the value of a building declines exponentially to some standard fraction of its value when new. That is, a building whose age is *Age* has a value equal to the fraction $[a_1 + (1-a_1)\exp(-a_2\,Age)]$ of the value it would have if it were new, where a_1 is the fraction of new value that old buildings approach and a_2 is the rate of exponential decay per year. Introducing a_1 and a_2 into our estimation, we obtain Estimate 3:

First intercept:	2.895
First slope:	0.688
Second intercept:	4.006
Second slope:	0.221
a_1	0.714
a_2	0.033
Variance of the distribution of residuals:	0.322
Log of the likelihood:	-68.980

In other words, we now estimate that the median value of commercially zoned land in Portland is exp(2.895), which is \$18.09 in 1986 dollars, or \$34.59 in 1998 dollars. We estimate that the log of the real value per square foot of any commercially zoned property in Portland (land and building together) is

$$2.895 + [0.714 + (0.286)\exp(-0.033Age)][\mathrm{Min}(2.895 + 0.688Cov, 4.006 + 0.221Cov) - 2.895].$$

Figure L shows the distribution of the residuals from Estimate 3 and densities of a corresponding normal distribution, permitting a judgement of the adequacy of the assumption that the residuals are distributed normally. Figure M shows the residuals from Estimate 3 in time sequence, permitting a judgement of the adequacy of the assumption that there is no time dependence in these residuals.

All of the estimates so far make the implicit assumption that the value of commercial land in downtown Portland is uniform. This is not plausible. To identify spatial patterns in sales, we do some quadratic smoothing. That is, we take the sum of the first intercept

(2.895) and the residual as an initial estimate of the log of land value. Then for each sale in the reduced sample, i, we run a weighted least squares regression in which the dependent variable is the estimated log of land value for all sales other than i and the independent variables are the spatial coordinates of those sales and quadratic terms in those coordinates. That is, we run 81 weighted regressions of the form

$$L = a_{0i} + a_{1i}X + a_{2i}Y + a_{3i}X^2 + a_{4i}Y^2 + a_{5i}XY + \varepsilon, \tag{1}$$

where L is the initial estimate of the log of land value for sales $j \neq i$, and X and Y are spatial coordinates of those sales. For regression i we weight sale j by $\exp(-kD^p)$, where D is the distance from sale i to the sale j (the square root of $[(X_i - X_j)^2 + (Y_i - Y_j)^2]$), and k and p are parameters to be optimized. After the coefficients of the regression i have been computed, we use them to calculate a revised estimate of the log of land value for sale i. The difference between our initial estimate and the revised estimate of the log of land value for sale i is a "basic residual." To optimize the parameters k and p we assumed that the basic residuals are normally distributed and identified the combination of k and p with maximum likelihood. We found that the optimal values of the parameters k and p are $k = 1.1716$ and $p = 0.2338$. In other words, if one seeks to explain the apparent land price of sale i (the sum of the first intercept and the residual from the explanation of the price of sale i as a function of coverage and age) by fitting the apparent land prices of surrounding sales to a quadratic function of their coordinates, with weights of the form $\exp(-kD^p)$, then values of $k = 1.1716$ and $p = 0.2338$ provide weights that estimate land prices most accurately.

Once the optimal values of k and p have been identified, one can estimate land value at any point (X, Y) by using all 81 observations to estimate coefficients of equation (1) for (X, Y), employing the optimized values of k and p to determine the weights of all sales. Map 3 was produced in this way. For each set of coordinates, it shows an estimate of the value of land in 1986 dollars. Map 4 shows a three-dimensional picture of our estimates of land value in 1986 dollars.

The final column of Table 4 shows our final estimated values of the 81 properties in our reduced sample. Each of these estimates is expressed in terms of the dollars of the year before the sale, as if we were reporting assessed values on January 1 of the year of the sale, which would then be compared with selling prices in the subsequent year.

Results

To evaluate our accuracy in assigning values to property, we first compare our accuracy within the sample with that of the assessor. Table 1 shows the result. For each year, our coefficient of variation is computed as

$$\frac{\sqrt{\frac{\sum_{i=1}^{n}(r_i-\bar{r})^2}{n-1}}}{\bar{r}}, \tag{2}$$

where each r_i is the ratio of our final estimated value of a property to the selling price of that property and $\bar{r}$ is the average of these ratios for the year. (The division by $n - 1$ rather than n is the standard adjustment for the estimation of the variance in a small sample.) The assessor's coefficient of variation is computed the same way, except that the assessed value assigned by the assessor is used in place of our estimated value. We did not do quite as well as the assessor. We achieved an average coefficient of variation of 0.653, while the assessor achieved an average of 0.553. Thus the assessor was slightly more accurate than our model. However, the year-to-year variations in results make it clear that either model could turn out to be better in a larger sample.

Table 2 shows the model's performance with respect to vacant lots in the sample. There were three completely vacant lots and one (sale number 26) with just 101 square feet of structure, which we chose to classify as vacant. For each vacant lot, we computed the relative inaccuracy of our estimate of its value as

$$\frac{n}{n-1}\frac{(r-\bar{r})^2}{\sigma^2}, \tag{3}$$

where n is the number of sales in the year, r is the ratio of the estimated price to the actual price, $\bar{r}$ is the year's average of this ratio, and σ is the standard deviation of the r's. This measure would be 0 if we estimated a price perfectly, and it would be 1 if we estimated vacant land prices as well as we estimate the price of the average improved property. The square root of the average relative inaccuracy is a measure for which "twice as bad" has its ordinary meaning. Thus our square root of average relative inaccuracy of 1.057 means that the inaccuracy of our estimates of vacant land prices was about 5.7% worse than our average inaccuracy of prices of all property. The assessor, on the other hand, had a relative accuracy that was about 80% worse for vacant land than for all properties. The difference in performance is large enough to suggest that the proposed methodology may be a better way to assess land. However, the number of observations is so small that no definite conclusion can be reached.

The most important test of a model's power is its ability to predict outside of the sample that was used to generate it. After we developed out model, we went back to Data Quick and obtained information about sales that had occurred since our initial data-gathering effort. We found five additional sales of downtown commercial property in Portland. None of them were sales of vacant land, so we can report only on the out-of-sample predictive power of our model for improved downtown commercial land. Table 3 shows the results. For the five out-of-sample observations, we had a coefficient of variation of 0.488, compared to the assessor's 0.542. Thus we did better than in the sample period and somewhat better than the assessor, but the number of observations is so small that the difference is not meaningful.

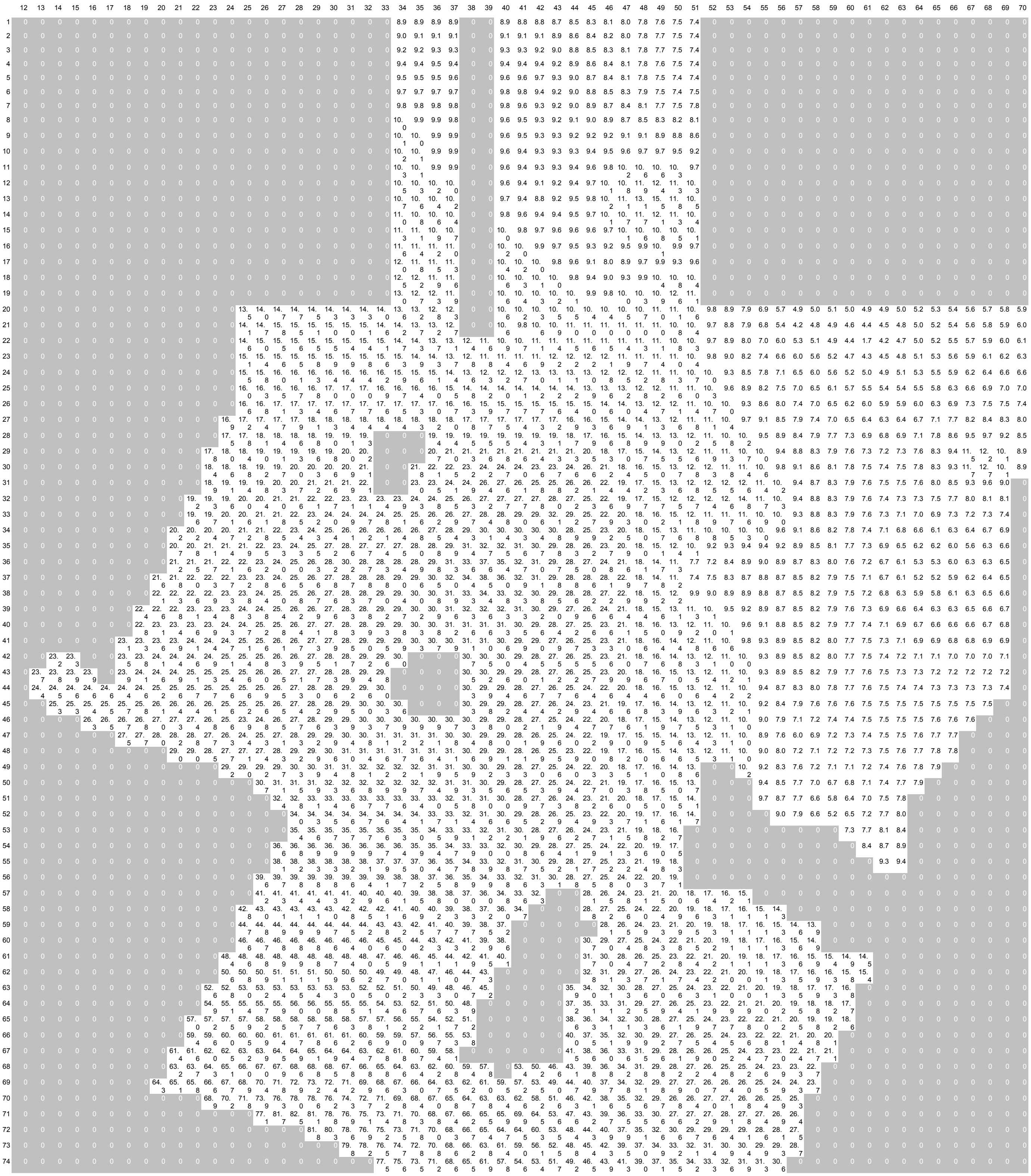
Conclusion

From this initial inquiry into the possibility of accurate assessment of downtown commercial land, one can conclude only that more work on this issue needs to be done and that it is worth doing. The amount of information that one can get from any one city at a modest cost is too little to reach a conclusion about the efficacy of the assessment method proposed here. The performance of the method was good enough to suggest that further study is warranted, but not good enough to establish that the proposed method is adequate.

Reference

Mills, Edwin S., “The Economic Consequences of a Land Tax,” in Dick Netzer, ed., *Land Value Taxation: Can It and Will It Work Today?*, Cambridge, Mass.: The Lincoln Institute of Land Policy (1998).

Map 3: Estimated Land Values in Our Study Area in 1986 Dollars



75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71.6	68.7	65.1	60.7	54.7	48.6	51.8	51.6	49.9	47.8	45.4	43.1	41.0	39.2	37.6	36.3	35.2	34.3	33.5	32.8	32.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	65.5	61.3	56.3	52.8	52.9	52.4	51.1	49.2	47.1	45.0	43.0	41.2	39.6	38.3	37.1	36.1	35.1	34.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59.5	56.8	55.4	54.1	52.6	50.7	48.8	46.8	44.9	43.2	41.6	40.2	38.9	37.8	36.8	35.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56.1	54.3	52.5	50.6	48.7	46.8	45.1	43.5	42.1	40.8	39.6	38.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54.3	52.4	50.5	48.7	47.0	45.4	44.0	42.6	41.3	40.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

replace with Map 4 (originally map 5)

replace with Table 1, etc.