

Using a Geographic/Land Information System for Land Price Estimation in Krakow, Poland

Dennis Robinson
Director of Education
Lincoln Institute of Land Policy

Copyright © 1994 by Dennis Robinson

Purpose

This paper reports on one part of a larger project on land price modeling in Krakow, Poland. Currently, the city assesses a property tax on land value based on the size of property. The purpose of building a land price estimation model for the city is to begin the process of converting to an ad valorem tax, which has the potential to increase tax revenue and equity.

The purpose of this part of the project is to explore ways in which the use of geographic/land information systems (GIS/LIS) could enhance CAMA approaches that emphasize nonspatial attribute data. A spatial approach to valuation is useful in Eastern European countries with emerging market economies for three reasons:

1. Attribute data on properties are more limited, and most governments do not have the resources to construct full-scale assessment databases;
2. Use of locational influences for determining taxable property value is more politically acceptable; and,
3. The property tax is generally based on land value rather than building value.

This paper outlines two phases of the project: in the first, a homogeneous submarket was defined using GIS, and a value influence center was identified using GIS; in the second phase, submarkets were defined through the inclusion of a variable on planned land use, and value districts were defined and added to the model.

STAGE I

Data

Spatial and attribute data, as well as initial land price estimations based on the nonspatial attribute data, were identified and obtained on a site visit to Krakow. The attribute data on 269 vacant land sale transactions were transferred to GIS from a statistical file used in the development of the nonspatial model.

Planar coordinates of the boundaries of administrative districts (obrebs) within the city were obtained and transferred into the GIS. Planar coordinates for each of the land sale points were included in the attribute data, though they needed transformation before being transferred into the GIS so that they would be correctly positioned with respect to the obreb boundaries. The original data were collected using maps with a different orientation, so each coordinate pair was transposed — x coordinate became y and vice versa — and each coordinate number was multiplied by -1. The corrected coordinates, with associated point identifiers,

were then output from the statistical program to the GIS to create a point map of the observations.

Point-in-polygon analysis in the GIS located each point with respect to its obreb. The obreb numbers contained in the original data file were not the real obreb numbers, but were derived from another numbering system used during initial data collection. The correct obreb number was transferred from the GIS to the statistical data file for later analysis.

An atlas of the city of Krakow and an overlay map with obreb boundaries were also obtained during the site visit. Although the atlas maps were somewhat out of date, with most having dates from the early-1970s to the early-1980s, they are useful in approximating the kind of overlay analysis of land use, demographic, and socioeconomic data that could be used in this type of analysis.

Model with Limited Spatial Analysis

The land price estimation model prior to spatial analysis was:

$$\begin{aligned} \text{PRICE} = & - 86438.14 + 12.605 \text{ AZMTH} - 307.228 \text{ AZGEO} \\ & - 1560.916 \text{ LOGDIST} - 15.603 \text{ DEPTH} + 808.563 \text{ ECOHLV} \\ & + 1652.096 \text{ NIEGHLV} + 5361.841 \text{ USELV} + 2871.819 \text{ ACESGAS} \\ & + 10.122 \text{ SEWERLV} + 26101.74 \text{ TVLV} + 16487.52 \text{ LOGAREA} \\ & - 2021.369 \text{ SRAREA} + 18.129 \text{ AREA} \end{aligned}$$

where

PRICE	=	transaction price ÷ 20500 (to convert approximately to dollar values)
AZMTH	=	derived azimuth measure from city center to transaction point, ranging from 35 to 381
AZGEO	=	derived azimuth categorical measure from city center to transaction point (1=N, 2=NE, 3=E, 4=SE, 5=S, 6=SW, 7=W, 8=NW)
LOGDIST	=	log of distance from city center to transaction point, where distance is measured in meters
DEPTH	=	property depth, in meters
ECOHLV	=	transformed from categorical measure, which rates the level of ecological hazards
NIEGHLV	=	transformed from categorical measure, which rates neighborhood quality
USELV	=	transformed from categorical measure, which classifies types of allowed uses (high-density housing, low-density housing, farm buildings)
ACESGAS	=	transformed from categorical measure, which rates accessibility of gas lines
SEWERLV	=	transformed from categorical measure, which rates accessibility of sewer line
TVLV	=	transformed from categorical measure, which rates accessibility of cable television line
LOGAREA	=	log of area measure
SRAREA	=	square root of area measure
AREA	=	area measure, in square meters

The statistical results of this regression are included in Appendix A, model 1, along with statistical information on the resulting sales ratios.

Removal of Azimuth Measures

While the use of azimuth measures (AZMTH and AZGEO) may, along with distance from city center, provide some approximation of location in the absence of spatial analysis, the spatial analysis to follow allows these measures to be removed from the model.

The reason the azimuth measures were considered for removal from the model is because any relation between these measures and transaction price would be relatively arbitrary, depending on the start point of the measure and land market patterns within the city. Additionally, when using a linear model, an azimuth measure would not provide information on the existence or location of multiple value influence centers, areas around which property values increase or decrease.

The removal of the two azimuth measures decreases somewhat the explanatory power of the model, with adjusted R^2 decreasing from 0.9675 to 0.9639, and the unexplained error increasing (Appendix A, model 2).

Homogeneous Submarket

Point maps were created of transaction price and price per square meter to identify areas within the city in which price or price per square meter are significantly higher. Since the data are skewed toward lower-priced properties, the maps are oriented to identify areas of high variation in transaction price or price per square meter to locate specific submarkets based on the spatial distribution of prices.

Map A shows the spatial distribution of three different ranges of the transaction price divided by a factor of 20500, which approximates conversion to dollars. The price ranges were established to locate spatially higher-priced properties. Ranges were determined using frequency distributions of transaction prices. The vast majority of properties (250 of 269) had prices at or below 10000.

Map B shows the spatial distribution of three ranges of price per square meter. The ranges are equal-interval ranges; once again, the majority of properties (190 of 269) fall into the lowest category and those in the two higher price per square meter ranges are more prominently displayed.

Map B shows one concentrated area, northwest of the city center, where price per square meter is generally greater than that for the city as a whole. Points to the east of the city are too widely dispersed and too scarce to provide enough data for a model of a submarket area. The area directly south of the city is a homogeneous submarket showing a relatively low number (4 of 79) of points with high prices.

The model was recalibrated for this last submarket area. Map C shows the sales ratios in the submarket for both the nonspatial model and the submarket model. Sales ratios are ranged in three groups, with the middle one (shown with small boxes) centered on 1.0, and ranging 0.1 above and below. Low sales ratios are shown with points, and high sales ratios are identified by large boxes. Appendix A, model 3, provides the statistical results of the submarket model and statistics on the sales ratios resulting from the recalibration. The explanatory power of the model increases from that of the original model. Despite the fact that it is expected that the adjusted R^2 will decrease due to decreased variance in the submarket compared to the entire city, the adjusted R^2 increases from 0.9675 to 0.9897. Additionally, the sum of squares of the errors decreases from 596,037,500 to 127,712,400.

Value Influence Centers

A response surface map of the residuals of the regression equation for the submarket was used to identify potential value influence centers. The left side of Map D shows a two-dimensional thematic map derived from a three-dimensional triangulated response surface of the residuals. The map shows the "peaks and valleys" of the residuals unexplained by variables in the submarket model. The darker the color, the higher the residual.

The submarket shows two large areas of high residuals, one in the northeast and one in the southwest. Examination of maps from the Krakow atlas shows that the high residual area in the southwestern obreb has characteristics that could explain the high residuals and might exert some influence on surrounding areas. First, one map shows that near this area is a concentration of service centers and repair shops. In fact, it is the biggest concentration of such entities in the submarket. Proximity to these services and shops may be a positive locational factor unexplained by the variables in the submarket model. Second, two other maps show that the land use of the area is low-density housing, green spaces, and parks. Finally, one map also shows that the area is designated as an area of special care for conservation of existing green space and land use.

This special care area is a potential value influence center. Since coordinates of the area are not known because that data layer has not been added at this stage of the analysis, the x-y coordinates of the approximate center point of the area of high residual were determined. Distance from each transaction point to this point was computed and added as a variable to the submarket regression.

Appendix A, model 4 provides the statistical results of the respecified model and statistics on the resulting sales ratios. Statistically, the model changes little overall from the results obtained in the former submarket regression. The adjusted R^2 and sum of squares of errors are at approximately the same level.

However, there is improvement in the residuals in the special care area. A response surface map of the residuals of the respecified model was created. The right side of Map D shows the two-dimensional thematic map derived from the three-dimensional response surface. It is clear from this map that the addition of the distance variable to the model has reduced the

level of residuals in that area to the level of the surrounding area (change from black to grey).

STAGE II

The analysis in Stage II was extended along two dimensions: (1) inclusion of a planned-use variable to define submarkets; and, (2) definition and use of value districts in the model. Both of these extensions resulted in more homogeneous groups of properties, which yield better price estimations. Furthermore, the definition and use of value districts provide a more rational and explainable estimation approach for an emerging land market in a city the size and complexity of Krakow.

Planned Land Use

The International Association of Assessing Officers' standard states that for vacant land sales, the coefficient of dispersion (COD) for the sales ratio (assessed value/transaction value) should be less than 20. One of the reasons that previous model results were not within this guideline was because the model was estimated and applied to all types of properties represented in the database. The database included transactions of land intended for low- and high-density residential uses as well as land for farm buildings. Most likely, each of these intended uses has a separate market.

The model was, therefore, estimated separately for low-density residential and farm buildings. The model is not estimated for high-density residential land use because the sample size of 17 observations is too small for the number of independent variables included in the model. Summary statistics from each model are included in Appendix A. The results for low-density residential land use (model 5.a) do not improve when the model is estimated for all transaction points throughout the city. There is, however, improvement in the COD of the sales ratios for land planned for farm building use (model 5.b) from 29.72 to 24.65.

Value Districts

Experience with the emerging land market in Krakow indicates that different parts of the city are different enough for residential land buyers so that submarket analysis needs to be part of the modeling process. Based on some firsthand information about different parts of Krakow as well as information on the major road and rail networks, a first attempt was made at defining contiguous geographic areas that exhibit more homogeneity, grouped spatially, than one can expect throughout the entirety of a city the size of Krakow.

Map E shows the network of major roads and rails in the city. Map F shows the 22 value districts that were established. These value districts were defined in the GIS and then a point-in-polygon analysis was run to locate each transaction point with respect to the value districts.

These data were transferred to the statistical database and established as binary variables. The binary variables were added to the specification of the current model and estimated for

all observations (model 6.a) and then for low-density residential transactions (model 6.b). The resulting CODs were lower than the model without these value district variables, but still above 27.

The next step was to estimate separate models for each of the value districts for which there were sufficient data. Two methods were followed for each of these. The first estimates the existing model for the specific value district. The second uses stepwise regression to respecify the model, with the results then estimated for the value district.

In value district 12 (see Map F for value districts), the CODs for both models (7.a and 7.b) were below 18. In value district 13, the CODs were 21.5 for the current model (8.a) and 21.9 for the model specification obtained from stepwise regression (8.b). Value districts 15 and 16 both provide mixed results, with the better results coming in both cases from the model specification obtained from the stepwise regressions (models 9.a-10.b).

Limitations of the Analysis

Some data and statistical anomalies in the nonspatial model should be noted in interpreting the results. First, the data were taken as given, without information on how data were gathered. Second, some of the observations (41 of 310) were removed in the development of the nonspatial model. Criteria for removing observations are not known and need to be examined. Third, the nonspatial model exhibits multicollinearity among the area variables. The variance inflation factors (VIF) for these variables in the original model were 101.36 for AREA, 122.99 for LOGAREA, and 401.32 for SRAREA. Variables with VIFs above 5 or 10 are often considered for removal from a model. While multicollinearity may be acceptable to capture some of the curvature of the relationship between area and price, further conditioning of these variables may improve the model. Finally, the number of observations in the database affects both the type and validity of the analyses that can be done, especially when the data are divided into submarkets and value districts. The estimations derived for value districts, especially, need to be retested as more data are gathered in these areas.

Conclusions and Possible Extensions

The goal of this inquiry was to identify ways in which spatial analysis could aid in land market analysis and modeling of land prices, and some promising areas exist and have been shown.

Stage I identified a homogeneous submarket, in this case based on spatial disaggregation. Other potential submarkets based on spatial analysis can be explored, especially as more data are collected.

The use of residual analysis to identify a value influence center and the statistical effect of a value influence center were also examined. This analysis utilized a point location for the value influence center, but these centers could also be lines (a major road, for example) or polygons (an industrial area, for example). This analysis used direct distance to the value

influence center as a variable, but other measures could also be explored as the influence may not be a direct function of distance.

Other methods of locating value influence centers could also be explored. These include additional sales ratio mapping and land value mapping to complement the analysis reported here. Analysis of the spatial patterns of development and local buyer preferences could also be examined with firsthand knowledge of the local land markets.

The results obtained in Stage II are promising, with three of four value districts tested having CODs well below the IAAO standards, and the fourth significantly better than results previously obtained. The analysis was done by segmenting the data and developing models for the value districts. Additional techniques could also be tested for adding value districts to the model for the city as a whole.

The value districts were defined without regard to where transaction data were obtained, and, in fact, some of the value districts have no transaction points in them, and many had too few observations to model. Value districts can be refined and analyzed based on the emerging land markets in the city and addition of sales points to the database.

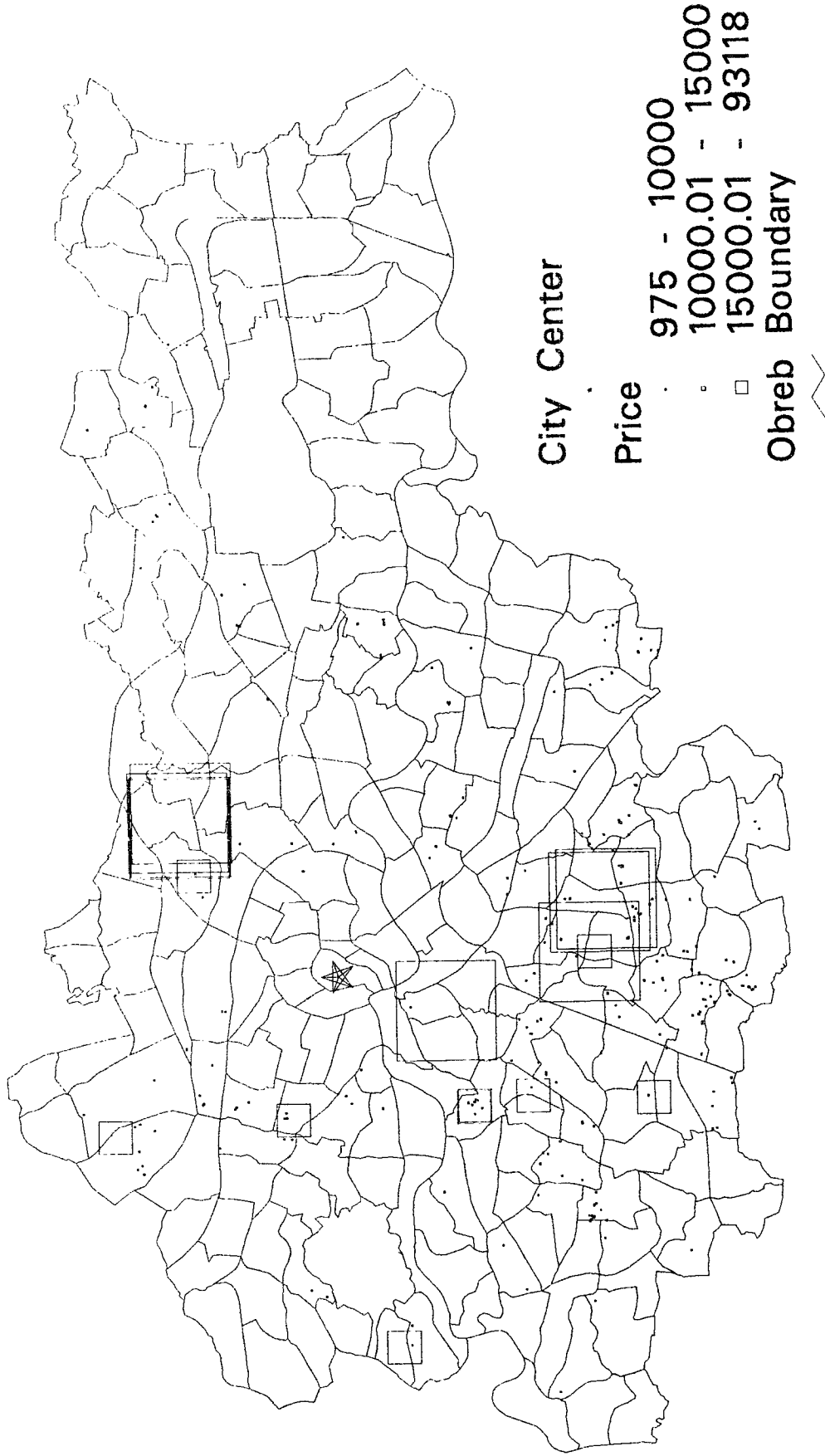
Appendix A: Summary of Statistical Results

The table below summarizes information and results for models estimated.

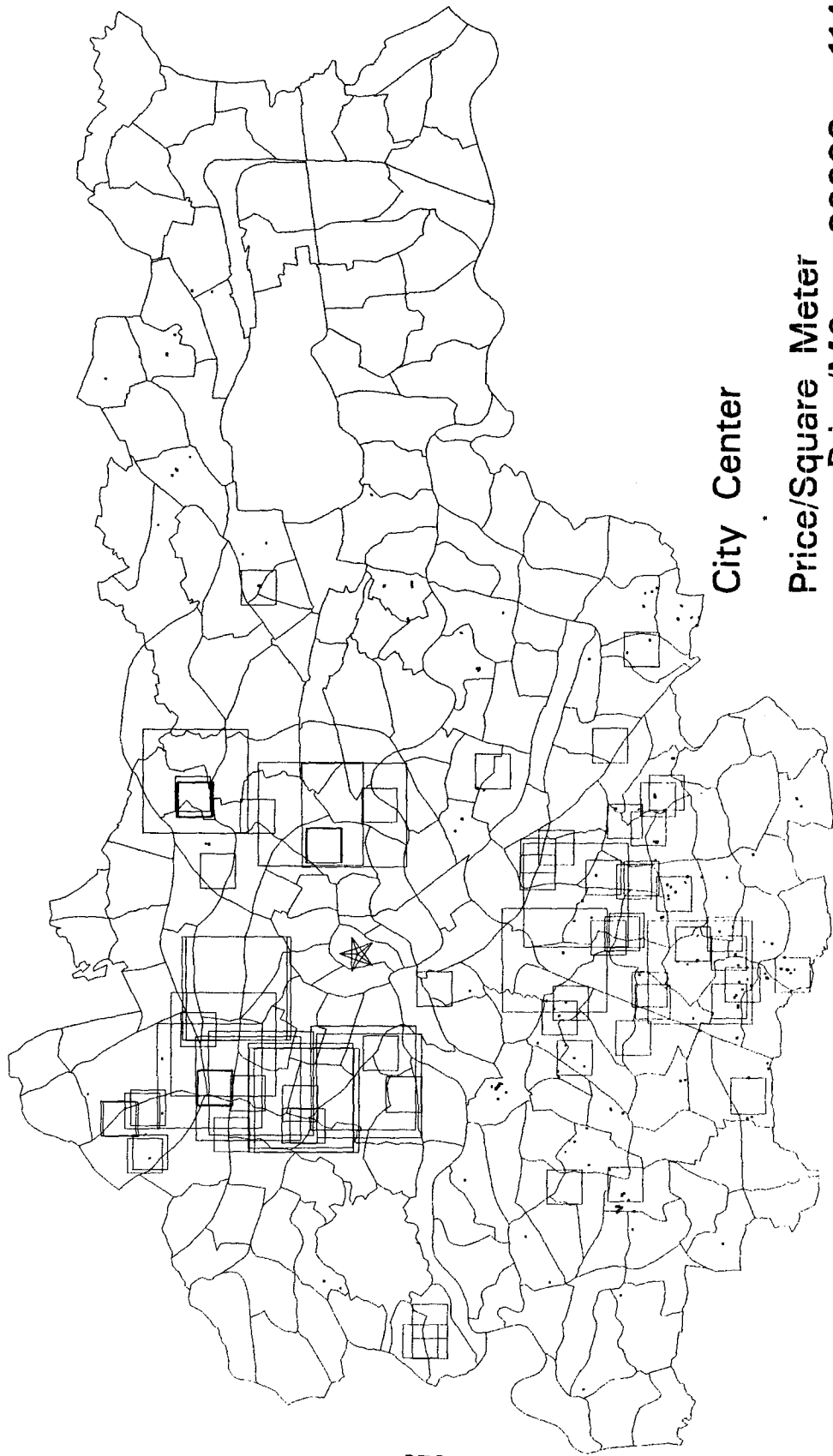
<u>Model</u>	<u># Obs</u>	<u>Model</u> <u>Adj R²</u>	<u>Error</u>	<u>Sales Ratio</u> <u>COD</u>	<u>Median</u>
1. Model with limited spatial analysis, estimated for all transaction points.	269	0.968	5.96 E08	29.72	1.020
2. Model 1, with azimuth measures removed, estimated for all transaction points.	269	0.964	6.66 E08	30.37	1.031
3. Model 2, estimated for submarket.	79	0.990	1.28 E08	31.11	1.003
4. Model 3, with distance to value influence center variable added, estimated for submarket.	79	0.90	1.27 E08	30.84	1.012
5. Model 2, estimated separately for (a) low-density residential, and (b) farm building transactions.					
a.	201	0.633	4.18 E08	32.07	1.044
b.	51	0.600	1.02 E08	24.65	1.097
6. Model 2 with value district binary variables added, estimated for (a) all transaction points and (b) low-density residential transaction points.					
a.	269	0.970	5.11 E08	27.67	1.004
b.	201	0.699	3.13 E08	27.59	1.022

7.a.	Model 2 estimated for low-density residential in value district 12.				
	42	0.842	3.53 E07	17.4	1.03
b.	Model respecified using stepwise regression for low-density residential observations in value district 12.				
	42	0.881	2.83 E07	17.2	1.03
8.a.	Model 2 estimated for all low-density residential in value district 13.				
	38	0.616	5.19 E07	21.5	1.022
b.	Model respecified using stepwise regression for low-density residential observations in value district 13.				
	38	0.727	4.10 E07	21.9	0.992
9.a.	Model 2 estimated for all low-density residential in value district 15.				
	35	0.372	3.38 E07	26.6	1.042
b.	Model respecified using stepwise regression for low-density residential observations in value district 15.				
	35	0.751	1.18 E07	15.2	0.979
10.a.	Model 2 estimated for all low-density residential in value district 16.				
	29	0.221	4.57 E07	28.9	0.759
b.	Model respecified using stepwise regression for low-density residential observations in value district 16.				
	29	0.804	1.15 E07	20.0	0.975

Map A: Transaction Price Ranges



Map B: Price/Square Meter Ranges

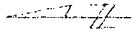


City Center

Price/Square Meter

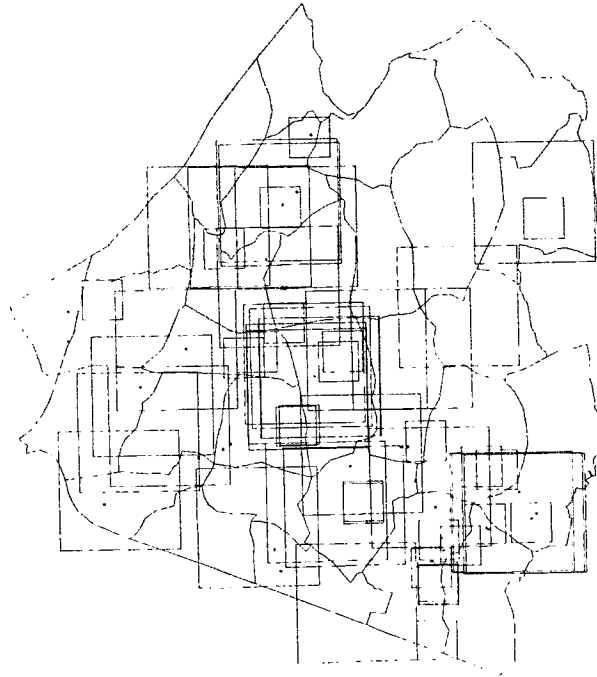
- Price/M2 = 30908 - 114461
- Price/M2 = 114461 - 198013
- ◻ Price/M2 = 198013 - 281566

Obreb Boundary



Map C: Submarket Sales Ratios

Nonspatial Model Submarket Model



Sales Ratio: Nonspatial Model

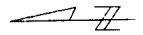
- 0.4289564 - 0.8999999
- 0.90 - 1.10
- ◻ 1.1000001 - 2.6171540

Obreb Boundary

Sales Ratio: Submarket Model

- 0.4617245 - 0.8999999
- 0.90 - 1.10
- ◻ 1.100001 - 2.6078380

Obreb Boundary



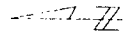
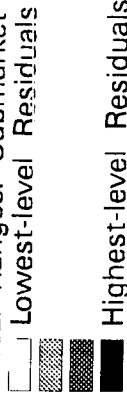
Map D: Residual Ranges (from surface analysis)

Submarket Model

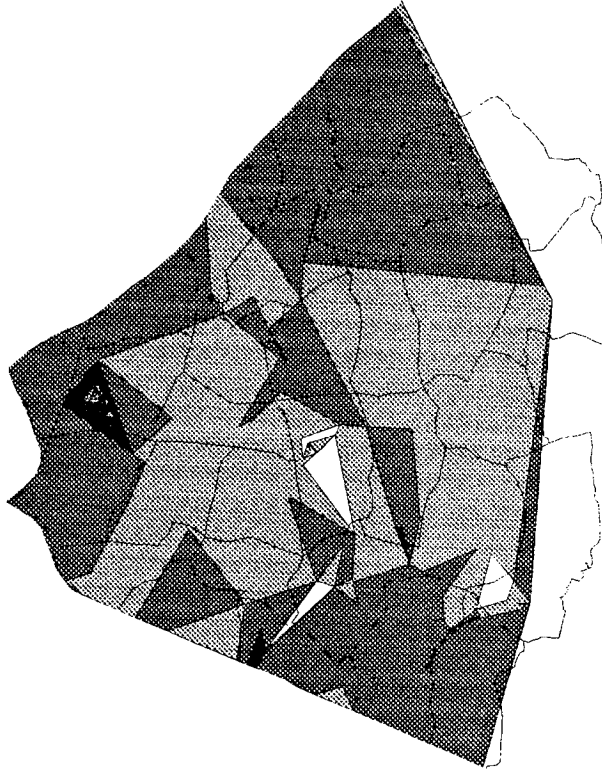


Obreb Boundary

Residual Ranges: Submarket Model

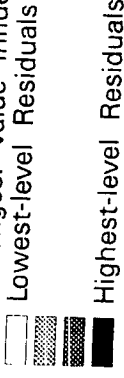


Value Influence Model

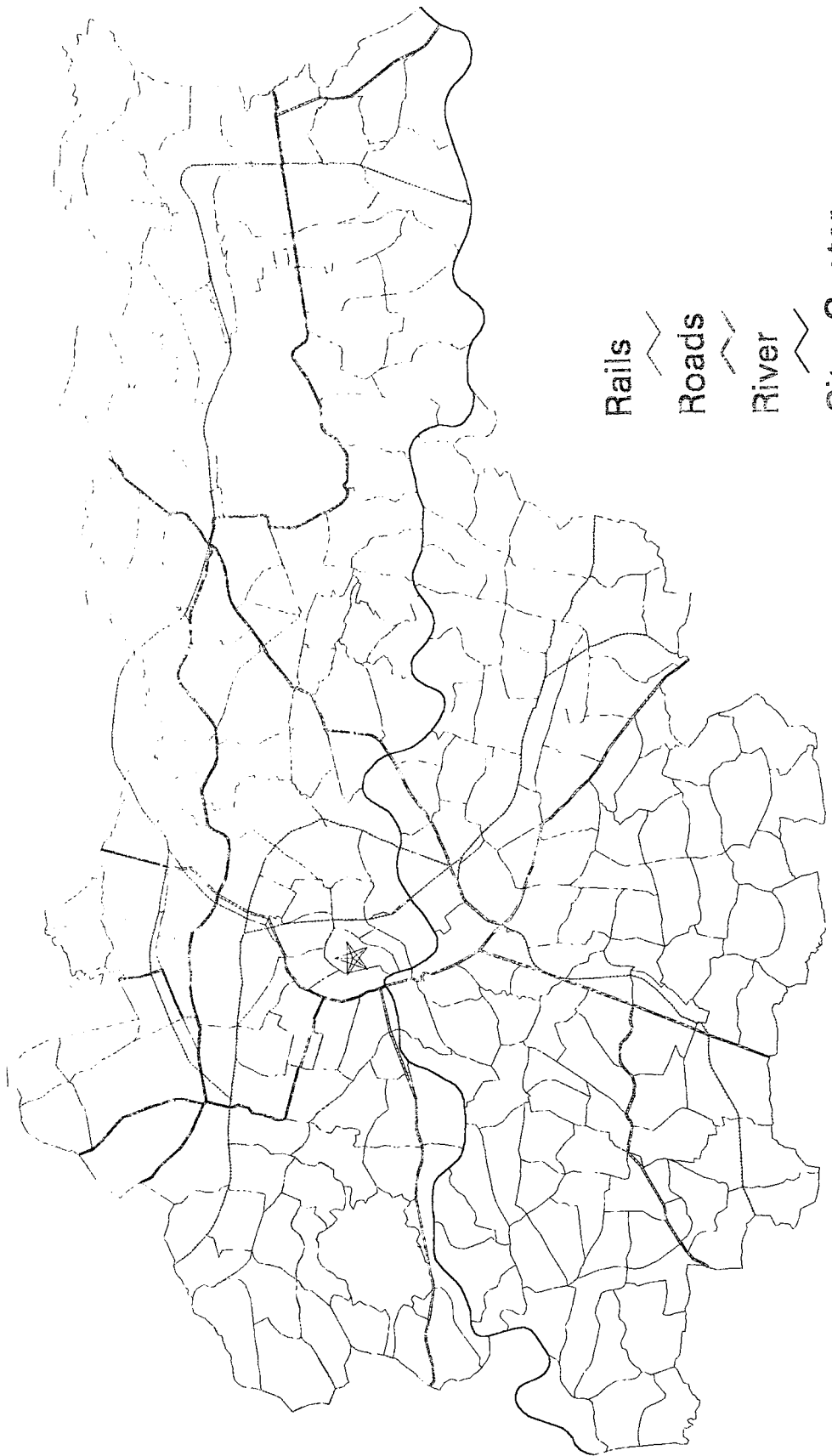


Obreb Boundary

Residual Ranges: Value Influence Model



Map E: Major Roads and Rails



Rails
Roads
River
City Center
Obregon Boundary



Map F: Value Districts

