DETERMINING FAIR MARKET VALUE

OF

REAL ESTATE

Sico Stevens A. Yao

Econometrics 560

Department of Economics
The University of Montana
Spring 2004

Abstract

This paper illustrates how data can be used to make estimations and predictions about the market value of real estate. To obtain these predictions, we propose a model that incorporates the period in time the property is acquired and different characteristics of the location. The study shows that a hedonic model which includes the elevation and the date of purchase of the property, the flood condition of the location, the county in which the lot is situated and its distance toward San Francisco can be formulated for prediction purposes using the city of Mountain View, CA data. We also discuss other substantive issues regarding any violations of the model assumptions.

Introduction

Determining a fair market value of a real estate is not an easy task. This is essentially due to the nature of the product which is characterized by a wide variability in price. In addition, the price itself is attributable to intangible factors which attributes are difficult to measure. This is the situation in which the city of Mountain View, CA was confronted to when it wanted to acquire 246.8 acres of land owned by Leslie Salt Company. The purpose of the city of Mountain View was to transform Leslie Salt property previously used for salt evaporation, into a city park. The parcel of land was diked for preventing the waters from the San Francisco Bay to flood the site.

Thus, the purpose of this paper is to demonstrate how regression analysis can be used to build a model from a set of time series data to assess the value of the Leslie Salt property.

We first begin our discussion by providing a description of our data used in the model suggested. In the next section, we suggest an estimation procedure. Finally, we conclude by discussing the results and implications.

Data

In our empirical application, we combine 31 observations on bayland properties that were sold during the previous 10 years. The raw data collected in 1968 can be found in the textbook *Analyzing Multivariate Data* by J. Lattin, J. D. Carroll, P. E. Green [2003] and is available in appendix B as well. It includes seven independent variables that might not all be candidate for the Leslie model. The variables COUNTY and FLOOD which both represent dummy variables indicate that the properties subject to tidal flooding are located in San Mateo. Table 1 summarizes the data used in our analysis. A description of the variables included in the data set, their means, standard deviations, minima and maxima are shown in the table.

Variables	Definitions	Mean	Standard Deviation	Mini.	Maxi.
PRICE	Sale price in $000 per acre	11.952	7.7147	1.7	37.2
COUNTY	San Mateo=0, Santa Clara=1	0.6129	0.49514	0	1
SIZE	Size of the property in acres	139.97	327.17	6.9	1695.2
ELEV	Average elevation in feet above sea level	4.6452	4.3554	0	20
SEWER	Distance in feet to nearest sewer connection	1981.3	2481.3		10000
DATE	Date of sale counting backward from current time (in months)	-58.645	24.527	-103	-4
FLOOD	Subject to flooding by tidal action=1; otherwise=0	0.16129	0.37388	0	1
DIST	Distance in miles from Leslie property (in almost all cases, this is toward San Francisco)	5.1323	4.5364	0	16.5
n	Nombre of observation, n=31				

Table 1 - Means and Standard Deviations

Model & Hypothesis

It is our goal to capture the variables that relevant in the value estimation of the Leslie Salt property. To do so, we suggest a classical linear regression model which supposes for the stochastic term to be normally distributed. We have adopted a log-lin functional form for the Leslie property model in order to correct for the assumption of normality not met by the errors term. A Jarque-Bera (JB) test was used to assess that the normal distribution assumption was valid after the log transformation in the model. We also assume no autocorrelated disturbances even though the Durbin-Watson test is inconclusive for this matter. We finally allow no multicollinearity which is supported by the correlation matrix whose coefficients are all less than .65.

We have reasons to believe that the variables COUNTY, SIZE and SEWER do not belong to the Leslie model. For instance, the information on the size has already been captured by the sale price measured per acre. Moreover, all the properties subject to tidal flooding are located in San Mateo which might suggest that this information is redundant. Likewise, SEWER might be irrelevant information as it is probably more meaningful to the properties subject to tidal flooding which are once again located in San Mateo. Thus, our hypothesis leads us to the following model:

$$\mathrm{Log}(\mathrm{PRICE}_i) = b_0 + b_1\mathrm{ELEV}_i + b_2\mathrm{DATE}_i + b_3\mathrm{FLOOD}_i + b_4\mathrm{DIST}_i + \mu_i \qquad (1)$$

However, a deeper analysis reveals the interaction of the two explanatory variables: COUNTY and ELEV. The graph below (figure 1) illustrates the interaction of the two explanatory variables. Indeed, if there were no interaction, the lines of San Mateo and Santa Clara on the plot of log (PRICE) against ELEVATION would be parallel.

Based on the preliminary observation, we propose model that takes into account the interaction term of the two explanatory variables COUNTY and ELEV. The improved model can be described as follows:

$$\mathrm{Log}(\mathrm{PRICE}_i) = b_0 + b_1\mathrm{ELEV}_i + b_2\mathrm{DATE}_i + b_3\mathrm{FLOOD}_i + b_4\mathrm{DIST}_i + b_5\mathrm{COUNT}_i + b_6(\mathrm{COUNTELE}_i^{1}) + \mu_i \qquad (2)$$

We anticipate a positive sign for the coefficient on elevation (ELEV), date (DATE), distance (DIST), and county (COUNTY). On the contrary, we predict the coefficient on flood (FLOOD) to be negative. Moreover, we suspect the change in the independent variables ELEV and FLOOD to have a significant impact on the price. In the same logic, we should

[1] COUNTELE = COUNT * ELEV

expect the parcels of land in Santa Clara to be more expensive that the ones subject to tidal flooding and located in San Mateo.

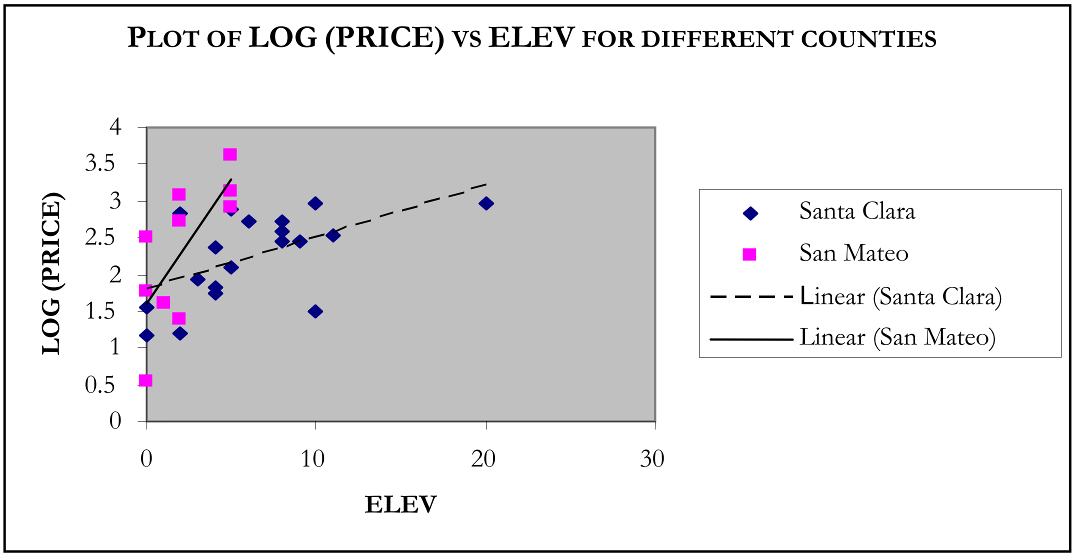

Figure 1 – Plot of log (PRICE) vs. ELEV for different counties

In order to capture the interaction effect, equation (2) can be simplified. Thus, equation (2) can be rewritten for San Mateo (i.e. COUNTY=0) as:

$$\text{Log}(\text{PRICE}_i) = b_0 + b_1\text{ELEV}_i + b_2\text{DATE}_i + b_3\text{FLOOD}_i + b_4\text{DIST}_i + \mu_i \qquad (3)$$

Likewise, equation (2) can be reduced for Santa Clara (i.e. COUNTY =1) as follows:

$$\text{Log}(\text{PRICE}_i) = (b_0 + b_5) + (b_1 + b_6)\text{ELEV}_i + b_2\text{DATE}_i + b_3\text{FLOOD}_i + b_4\text{DIST}_i + \mu_i \quad (4)$$

Empirical Results

Equation (2) was estimated using ordinary least squared (OLS) in SHAZAM v. 7.0. The results of the regression analysis are represented in table 2 below.

Variables	Model 1	Model 2
ELEV	0.0746 (4.435)	0.3199 (4.394)
DATE	0.018570 (6.902)	0.018069 (6.536)
FLOOD	-0.77886 (-3.868)	-0.30327 (-1.373)
DIST	.0591 (3.575)	.1212 (4.919)
COUNTY	N/A	1.2926 (3.284)
COUNTELE	N/A	-0.2658 (-3.484)
CONST	2.8244 (13.22)	1.4881 (3.566)
R^2	0.7806	0.8565
R^2 adj.	0.7468	0.8206
Standard error	0.36068	0.30357
SSE	3.3824	2.2118

Table 2 – OLS Estimates[2] (Estimated t-statistics in parentheses)

To begin with, we first concentrate on providing a substantive interpretation of the results. Thus, our analysis reveals that the coefficient on elevation, ELEV is statistically significant and positive. In San Mateo, a foot increase in elevation implies a $3,823.27 per acre in price at the means, others things the same. In Santa Clara, a foot increase in elevation implies a $646.5 per acre in price at the means, others things the same. Therefore, in Santa Clara, a 1 foot increase in elevation corresponds to a 31.99% increase in price per acre compared to

[2] The dependent variable is the natural log of the net price.

only 5.41% for Santa Clara, all else constant. This analysis reveals that the coefficient on elevation has a greater marginal impact for San Mateo which should not be surprising. After all, we could expect the price of lands in San Mateo (subject to tidal flooding) to increase dramatically further we are above sea level.

As regard to the coefficient on date, it is statistically significant. It is positive as expected. In fact, the more recent the purchase of the land is, the more expensive it will be because of inflation.

On the contrary, the coefficient on flood is negative and not statistically significant. As a result, being subject to flooding by tidal action implies a 26.16%[3] decrease in price per acre, all else constant. This result is in fact compatible with our predictions.

As opposed to the coefficient on flood, the coefficient on distance is positive and statistically significant. We can also see that a mile increase from Leslie property (toward San Francisco) leads to a 12.12% increase in price per acre, all else the same. Likewise, the coefficient on county is positive and statistically significant. An analysis of this coefficient results to the following interpretation. Being in Santa Clara implies a 264% increase in price per acre, all else constant. It is obvious that the proprieties in Santa Clara are much more expensive. An explanation is that they offer more investment opportunities because they are not subject to flooding by tidal action. A t-test indicates that the interaction term COUNTELE contribute significantly to the explanation of the variation in the dependent variable Log (PRICE).

We tested the significance of the overall model by looking to see if the variance accounted for by the model is reasonably large. We obtain a value of 23.88 for the F-statistics. The critical value at level .05 being 2.99 suggests that our model is significant. We also identify

[3] Note that for dummy variables, the coefficient used for interpretation is found by computing $[\exp(b_i)-1]*100$

observation 2 as an outlier according to the R-Student coefficient. However, we decided to not to remove the observation from the data set.

The overall fit of our model is pretty good with a R^2 of .8565. Likewise, if we compare model (1) and model (2), it is easy to notice that model (2) fits the data better with a R^2adj. of .8206 compared to .7468 for the model (1).

At this time, we are interested in forecasting a fair market value for the Leslie property which is our initial motivation. The Leslie property is approximately 247, located in Santa Clara (COUNTY=1), at sea level (ELEV=0), not subject to tidal flooding (FLOOD=0), located relatively far from San Francisco (DIST=0) and to be sold at current time (DATE=0). From equation (2), we obtain Log (PRICE) = 2.781 with a confidence interval of [2.16, 3.40][4] . Using the fact the independent variable price is Lognormal distributed, our prediction for the price is about $16,896. Moreover, we are 95% confident that our value will fall in the price interval of [9080, 31377].

Conclusion

The purpose of our modeling work is to uncover the factors that might influence market valuation of the Leslie Salt property in order to estimate the value of the parcel of land. Our methods lead us to the solution of our problems. In fact, we found that there is a strong evidence of linear association between our independent variable and the retained independent variables.

There are, however, several limitations of this research as normality assumption might be suspected. Even though the JB test is conclusive, this test is best suited for larger samples.

[4] At level .05

Furthermore, a deeper analysis is required for providing the means of extrapolating predictions beyond the range of the data used in this analysis.

Despite these limitations, the empirical results reported here are extremely robust. Accordingly, we hope that the questions addressed in our empirical results will provide an incentive for further research on similar issues.

Appendix

- Appendix A: *Note on Normal & Lognormal distribution*

If Log(X) ~ N (μ, δ2), then X~Lognormal (μ, δ2). The expression from the expected value for all moments is given by E (X^n) = exp (nμ + $n^2\delta^2/2$).

In particular, for n=1 we have:

$$E(X) = \exp(\mu + \delta^2/2).$$

- Appendix B: *Leslie Salt Data*

PRICE	COUNTY	SIZE	ELEVATION	SEWER	DATE	FLOOD	DISTANCE
4.50	1	138.40	10	3000	-103	0	0.30
10.60	1	52.00	4	0	-103	0	2.50
1.70	0	16.10	0	2640	-98	1	10.30
5.00	0	1695.20	1	3500	-93	0	14.00
5.00	0	845.00	1	1000	-92	1	14.00
3.30	1	6.90	2	10000	-86	0	0.00
5.70	1	105.90	4	0	-68	0	0.00
6.20	1	56.60	4	0	-64	0	0.00
19.40	1	51.40	20	1300	-63	0	1.20
3.20	1	22.10	0	6000	-62	0	0.00
4.70	1	22.10	0	6000	-61	0	0.00
6.90	1	27.70	3	4500	-60	0	0.00
8.10	1	18.60	5	5000	-59	0	0.50
11.60	1	69.90	8	0	-59	0	4.40
19.30	1	145.70	10	0	-59	0	4.20
11.70	1	77.20	9	0	-59	0	4.50
13.30	1	26.20	8	0	-59	0	4.70
15.10	1	102.30	6	0	-59	0	4.90
12.40	1	49.50	11	0	-59	0	4.60
15.30	1	12.20	8	0	-59	0	5.00
12.20	0	320.60	0	4000	-54	0	16.50
18.10	1	9.90	5	0	-54	0	5.20
16.80	1	15.30	2	0	-53	0	5.50
5.90	0	55.20	0	1320	-49	1	11.90
4.00	0	116.20	2	900	-45	1	5.50
37.20	0	15.00	5	0	-39	0	7.20
18.20	0	23.40	5	4420	-39	0	5.50
15.10	0	132.80	2	2640	-35	0	10.20
22.90	0	12.00	5	3400	-16	0	5.50
15.20	0	67.00	2	900	-5	1	5.50
21.90	0	30.80	2	900	-4	0	5.50

- Appendix C: *Shazam program codes*

```
set noscan
delete / all
sample 1 31
read (leslie_salt.csv) price county size elev sewer date flood dist /  skiplines=1
stat price / means=mprice
stat / all
ols price county size elev sewer date flood dist / gf pcor resid=e1
gen lprice=log(price)
ols lprice county size elev sewer date flood dist / gf resid=e2
test dist=0
gen1 a=.025
distrib a / type=t df=23 inverse
ols lprice date flood elev sewer dist / pcor resid=e3
*pc date flood elev dist
*plot e1 lprice
*plot e2 lprice
gen ee1=e1*e1
gen e=abs(e1)
*plot ee1 price
*plot e1 price
gen ee3=e3*e3
plot ee3 lprice
ols lprice elev date flood dist
gen countele=county*elev
ols lprice elev date flood dist county countele / coef=b rstat influence loglin
diagnos / jackknife
*test
*test countele=0
*end
*distrib a / type=f df1=6 df2=24 inverse
test
test elev=0
test date=0
test flood=0
test dist=0
test county=0
test countele=0
end
distrib a / type=f df1=6 df2=24 inverse
pc elev date flood dist county countele
gen1 marg1=b:1*mprice
gen1 marg2=(b:1 + b:6)*mprice
print marg1
print marg2
```

References

- DeGroot, Morris H. and Mark J. Schervish [2002], *Probability and Statistics*, Third Edition (Adison Wesley)
- Gujarati, Damodar N. [2003], *Basic Econometrics*, Fourth Edition (McGraw Hill)
- Gujarati, Damodar N. [1999], *Essentials of Econometrics*, Second Edition (McGraw Hill)
- Kennedy, Peter [2003], *A Guide to Econometrics*, Fifth Edition (The MIT Press)
- Lattin, James J., Douglas Carroll and Paul E. Green [2003], *Analyzing Multivariate Data* (Thomson)