UC Berkeley
Fisher Center Working Papers

Title
Statistical Analysis of Sales Data to Verify Appraisal Information

Permalink
https://escholarship.org/uc/item/5th178c6

Authors
Case, Bradford
Quigley, John M.

Publication Date
1988-10-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Institute of Business and Economic Research
University of California, Berkeley

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
WORKING PAPER SERIES

WORKING PAPER NO. 88-150

STATISTICAL ANALYSIS OF SALES DATA
TO VERIFY APPRAISAL INFORMATION

BY

BRADFORD CASE
JOHN M. QUIGLEY

These papers are preliminary in nature; their purpose is to stimulate discussion and comment. Therefore, they are not to be cited or quoted in any publication without the express permission of the author.

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
UNIVERSITY OF CALIFORNIA AT BERKELEY

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
J. W. Garbarino, Director

The Institute of Business and Economic Research is a department of the University of California with offices on the Berkeley campus. It exists for the purpose of stimulating and facilitating research into problems of economics and of business with emphasis on problems of particular importance to California and the Pacific Coast, but not to the exclusion of problems of wider import.

STATISTICAL ANALYSIS OF SALES DATA TO VERIFY APPRAISAL INFORMATION*

By

Bradford Case
and
John M. Quigley

October, 1988

Working Paper No. 88-150

I. Introduction

II. Advantages and Disadvantages of Statistical Appraisal Methods

III. Full Utilization of Available Information in Real Estate Appraisal Models

a. A Hybrid Model
b. Empirical Analysis

IV. Practical Application to Verification of Appraisals

References

---

* A previous version of this paper was presented at the Conference on Contemporary Problems in Assessment and Appraisal, Cambridge, MA, August 1988. The authors are grateful to the Lincoln Institute of Land Policy and the International Association of Assessing Officers for research support.

Abstract

Statistical models of real estate price determination can be an important tool in managing the appraisal function. The accuracy of these models depends crucially on the inclusion of all relevant information as well as the correct specification of the price function.

This paper proposes a hybrid model of the hedonic price relationship, combining information on single sales and multiple sales of properties. The paper also indicates the utility of such a model in the verification of appraisal data.

STATISTICAL ANALYSIS OF SALES DATA TO VERIFY APPRAISAL INFORMATION

By

Bradford Case
and
John M. Quigley

I. Introduction

Computer based statistical models of real estate markets have been employed for more than twenty years in the assessment, appraisal, and verification of market prices and in the forecasting of future real estate prices. Compared with less systematic methods of appraisal and assessment, computer based models provide a number of clear advantages. A more widespread diffusion of computer models into the markets for private and public sector real estate appraisals depends crucially on exploiting the advantages of such systematic methods while remaining conscious of their inherent limitations.

This paper considers the role of computerized appraisal techniques. In Section II below, we discuss some of the practical advantages and disadvantages of statistical valuation models. In Section III we compare a "standard" regression model of real estate prices with what has been termed a "weighted repeat sales" model. We then suggest a "hybrid" approach that incorporates the desirable properties

of each method. We also present and compare empirical estimates of a simple model of real estate prices using these three methods. In Section IV we describe one practical use of these models in the verification of private appraisals.

II. Advantages and Disadvantages of Statistical Appraisal Methods

Computer assisted real estate models offer a number of advantages over more traditional methods. Perhaps most important, these methods provide rigorous controls for the varying characteristics of properties. An important factor distinguishing the real estate market from most other economic markets is the wide heterogeneity of the product. This heterogeneity or product differentiation is particularly evident in the residential sector where the sale of a property involves a simultaneous transaction in the markets for land, for structures, for neighborhoods, for locations, and for access to the local public goods produced by governments. These latter markets, for neighborhoods, locations, and access to local public goods such as schools and police protection, include many trades in valuable attributes that are not even produced by landlords or developers.

The computer assisted models that have been developed to analyze real estate markets provide well-known methods to control for the joint influence of these important characteristics upon the observed transaction prices of

properties. In this way, they allow the investigation of price trends over time, or the comparison of properties standardized with respect to all but one physical or locational characteristic.

A second advantage of computer based models is that they permit the partition of the observed variation in real estate prices into individual determining factors. From the viewpoint of local officials and government planners, for example, statistical models can be used to isolate and forecast the price implications of public policies affecting property values, such as restrictions on the availability of land for development or improvements in accessibility to jobs and amenities. In the private sector, the same partitioning can be used to isolate and estimate the market return to investment in particular attributes or types of properties or in particular locations.

Finally, computerized appraisal models are of considerable potential value in guiding investors in their portfolio allocations. For example, all sophisticated theories of individual investment conclude that the allocation of an investment portfolio across securities should depend upon the mean return and the variability of returns (or risk) to the various kinds of securities. For most investments, historical data permit the returns to be calculated quite easily from a lively spot market in the commodity in question.

In real estate markets, however, trades are relatively infrequent and the markets are rather thin. Thus the estimation of an investment return series for real estate depends crucially on methods of imputation--that is, upon methods that use an observed sample of transactions to infer the market value of the unsold stock of property. Virtually all investment return series in real estate are calculated by some form of imputation of market prices based upon appraisal techniques. These appraisal techniques are ideally suited to computer based statistical modelling to represent expected property values and price over time.

On the other hand, it must be recognized that computer assisted real estate valuation models have rather well-defined limitations. Accurate and reliable statistical analysis depends crucially upon the identification of the variables relevant to the analysis, as well as the correct specification of the form in which these variables should be included. If one or more factors influencing property values are excluded from the statistical model, the conclusions may be systematically incorrect. Similarly, even if all relevant factors are included, if they are specified improperly the results may be misleading.

Of course, if extraneous information is included in a model it will not ordinarily affect the results, and statistical methods can provide assistance in eliminating

extraneous information. But statistical analysis alone cannot establish whether the relevant information has been captured in any model, nor can statistical results alone verify that the form in which this information has been utilized is "correct" in any theoretical sense. The models and the methods themselves are no substitute for more fundamental inquiry into the operation of the real estate market. Stated another way, guidance in the choice of empirical techniques, variables, and sources of data, as well as the functional form for relationships, depends upon a satisfactory theory of the real estate market.

More generally, if consumers in the market differ by the cost of acquiring information and suppliers recognize these differences, then there may exist an equilibrium *distribution* of market prices rather than a *single* market price for comparable real estate. The function of the appraiser, then, is to extract the "signal" from the "noise" associated with each observation on real estate transactions. Presumably, as information is revealed over time to participants in the market, the equilibrium price distribution will tend to converge to a single market price. However, the appraisal techniques used may inhibit or encourage this convergence of equilibrium prices. From this perspective, computer assisted appraisal techniques may represent the most rigorous systematic method of assimilating information on market behavior.

The technological changes of the past decade have accentuated both the advantages and the disadvantages of computerized appraisal. Reductions in the cost of computation and in the availability of computers have made it extremely inexpensive to gain access to basic information on real estate sales and real estate characteristics. The wide availability of statistical software packages has made it relatively simple for those with only rudimentary training in quantitative methods to estimate quite sophisticated statistical pricing models. However, the very simplicity of estimation has meant that decision makers can easily be inundated with superficially similar statistical models of appraisal. Models with similar statistical properties but utilizing different information (or simply utilizing information differently) may yield quite different forecasts, especially for time periods or properties outside the range of the sample used in estimation itself. Only rigorously specified models utilizing all available information can be expected to generate similar conclusions and forecasts.

III. Full Utilization of Available Information in Real Estate Appraisal Models

Given the inherent inability of computer appraisal techniques themselves to establish the "correct" representation of the determinants of real estate value, and given the tendency of such models to generate incorrect

conclusions when relevant information is excluded, it is imperative that statistical methods begin by utilizing *all* available information in estimating market prices. Unfortunately, most current models of real estate prices typically fail to utilize fully the available information on property values, especially when the same properties are sold several times.

a. A Hybrid Model

Consider, for example, a regression model of the relationship between the value of a property at time t, V_t, its physical and locational characteristics, X, and some representation of time:

(1) $V_t = f(X,t)$.

Interpretation of this relationship depends crucially upon the inclusion of the correct set of property characteristics, X, and the correct functional form, f(.), for the regression. Conditional upon these factors, however, the statistical model can be used to forecast the market price for a standardized or "quality adjusted" property over time. As an appraisal tool, this *standard* hedonic regression facilitates an appraisal of the value, $\tilde{V}_t$, of a property with characteristics $\tilde{X}$ at time $\tilde{t}$ through the computation of $\tilde{V}_t = f(\tilde{X},\tilde{t})$.

Even when the correct set of property characteristics and the correct functional form are employed, however, the standard method may fail to utilize fully the available information. The value of any specific property that was sold twice during a given time period presumably varies only with respect to time, since the physical and location characteristics of that property would not have changed. But the standard model expressed in Equation 1 treats multiple sales of the same property in the same way as multiple individual sales of very similar (but not identical) properties. In short, with multiple sales the standard model does not utilize completely the information on property characteristics.

Because of this problem, it has been suggested that the characteristics of each property be completely standardized with reference only to themselves, by confining the statistical analysis to properties that have been sold more than once during a given time period (Bailey *et al*, 1963):

$$(2) \quad V_t/V_{\hat{t}} = g(t,\hat{t}) \quad .$$

This formulation, termed the *weighted repeated sales* method, avoids the problems of the standard model by relating a change in the selling price of a property between time $\hat{t}$ and time t only to the timing of the two transactions, or perhaps to the time interval $(t-\hat{t})$ between sales. Price indices for single family houses have been computed using this technique by Mark

and Goldberg [1984], Palmquist [1980], and, more recently, by Karl Case and Shiller [1987].

Although this latter approach avoids the difficulty of specifying and measuring the various quality characteristics of real properties, it does so at considerable cost. By confining the analysis to properties sold more than once, it can be extremely wasteful of sales information. In any market run, the fraction of properties sold more than once is bound to be small. The estimation strategy implicit in Equation (2) simply ignores all information on the sale prices and the characteristics of single transactions.

Also, the weighted repeat sales technique is inappropriate when any of the characteristics of a property have been changed between sale dates. More importantly, success in the identification of properties with changed physical or locational characteristics entails the exclusion of still more information on market values of such properties. It is a curious research strategy indeed that excludes so much available information from the analysis.

Table 1 summarizes one measure of the sacrifice in information made by the recent studies utilizing the weighted repeat sales method. The table indicates that, despite the long time periods included (ranging from 14.5 years to 22 years), only a small fraction of the observed sales in the

TABLE 1

Samples of Sales and Repeat Sales
in Studies Using Weighted Repeat Sales Methods
for Single Family Housing

Study and Housing Markets Analyzed	Analysis Time Period	Sample Sizes Total Sales	Repeat Sales
Palmquist [1980]			
King County, WA	14.5 yrs	4,785	1,613
Mark and Goldberg [1984]			
Fraser, Vancouver	22.0 yrs	4,376	1,695
Kerrisdale, Vancouver	22.0 yrs	1,398	660
Case and Shiller [1987]			
Atlanta, GA	16.5 yrs	221,876	8,945
Chicago, IL	16.5 yrs	397,183	15,530
Dallas, TX	16.5 yrs	211,638	6,669
San Francisco, CA	16.5 yrs	121,909	8,066

Sources:

Raymond B. Palmquist, "Alternative Techniques for Developing Real Estate Price Indices," *Review of Economics and Statistics*, Vol 62, August 1980, pp. 442-448.

Jonathan H. Mark and Michael A. Goldberg, "Alternative Housing Price Indices: An Evaluation," *AREUEA Journal*, Vol 12, #1, 1984, pp. 31-49.

Karl E. Case and Robert J. Shiller, "Prices of Single Family Homes since 1970," *New England Economic Review*, Sept/Oct 1987, pp. 44-56.

seven housing markets analyzed were repeat transactions. Thus, by confining the statistical analysis to repeat sales, almost two thirds of Palmquist's sample of single family housing transactions was eliminated; in Case and Shiller's analysis more than 97 percent of the observed transactions were ignored.

Consider now an alternative research strategy that combines the desirable features of both Equations (1) and (2): it utilizes information on all transactions, whether single or multiple transactions; it capitalizes on the additional information available when multiple transactions are observed on any property. We call this the *hybrid* method.

To focus on the essentials, consider a very simplified model of real estate values in which initial property values, V_0, vary with the qualitative and quantitative aspects of properties, X, according to a simple exponential relation:

$$(3) \quad \log V_0 = \log A + a \log X \quad ,$$

where A and a are parameters. Moreover, the value of a property at a given time, V_t, varies continuously over time t:

$$(4a) \quad \log V_t = \log V_0 + bt$$

or, equivalently,

$$(4b) \quad \log V_t = \log A + a \log X + bt \quad ,$$

where b is a parameter. This simple model implies that if we observe a single transaction at time t, the selling price of the property can be estimated from (4b), which describes the *standard* model. Suppose, however, we observe two sales of a property, at t and $\hat{t}$, whose characteristics are unchanged during the interval from $\hat{t}$ to t. From (4a), the estimated selling price is

$$(5) \quad \log V_t = \log V_{\hat{t}} + b(t-\hat{t}) \quad ,$$

which describes the *weighted repeat sales* model.

Finally, suppose we observe two sales of a property at t and $\hat{t}$, but, in this case, suppose the characteristics of the property are changed from X to X* at t*, $\hat{t}<t*<t$. In this case it can be shown that

$$(6) \quad \log V_t = \log V_{\hat{t}} + a \log(X*/X) + b\,(t-\hat{t}) \quad ,$$

where the property characteristics included in the set X are continuously measured.

Equations (4),(5), and (6) provide alternative methods for estimating the parameters of the appraisal model for three kinds of samples. Consistent estimates of the parameters can be obtained from any of the three samples, at least as long as the samples are random.

Clearly if information is available for two or more of these samples, as in the studies reported in Table 1, then *all* the information should be used to estimate the parameters of the model. This is accomplished by estimating the appropriate equations jointly in the model of real estate prices. The estimates themselves may be computed by generalized least squares (see Case and Quigley [1988] for details).

b. Empirical Analysis

The model presented in Equations (4), (5), and (6) is estimated from observations on the sales of single detached housing from the Kahala neighborhood of Honolulu, Hawaii during the period October 1980 to October 1987. The neighborhood consists of about 1100 residential parcels and is bounded by Kahala Beach, the Waialae Golf Course, and the Kalanianaole Highway. The sample for this analysis consists of every sale in this neighborhood during the seven year period: 418 residential transactions involving 310 separate properties.

The 1980's were a period of rapidly rising prices in Hawaii, from an already high base; this neighborhood is no exception. The median sale price of these dwellings was $372,000. In 1980, the average transaction was for $361,000. In 1987, unadjusted sales prices averaged $845,000. More detail on the average characteristics of the sampled dwellings is reported in the appendix.

Table 2 summarizes the regression estimates based upon this body of data. It presents estimates based upon the *standard* method (Equation 4b), the *weighted repeat sales* method (Equation 5), and the *hybrid* method (Equations 4b, 5, and 6). For completeness, it also presents estimates of Equation 6 for those properties whose characteristics have been changed, as well as estimates of Equation 4b for the subsample of properties sold only once.

As the table indicates, an analysis based upon the weighted repeat sales method is extremely wasteful of the information contained in this sample. Of the 418 transactions observed during the seven year period, only 108 involved multiple transactions on the same property, and of these only 47 involved multiple transactions on properties with unchanged characteristics. An analysis based on the weighted repeat sales method uses only 11 percent of the data.

The table also reports the coefficients estimated for the time parameter b, as well as its standard error, and the explained variation in sale prices implied by each model. As compared to the results of the so-called *standard* model, the results of the *weighted repeat sales* model are suspect--despite the high proportion of the variation in sale prices accounted for by the model, the standard error associated with b is much

TABLE 2

Estimates of Real Estate Price Model*
All House Sales in Kahala, Honolulu
Oct 1980 - Oct 1987

$$\log V_t = \log A + a \log X + bt$$

Specification and Sample	Number of Observations	Estimated Coefficient, b	Standard Error of Coefficient, b	Explained Variation in Model
Standard Model				
All Sales	418	0.217	0.022	0.804
Single Sales only	310	0.179	0.022	0.796
Multiple Sales Models				
Weighted Repeat Sales	47	-0.136	0.097	0.955
Sales of Changed Properties	61	0.059	0.099	0.897
Hybrid Model	418	0.069	0.017	0.887

* Each regression also includes a set of X variables which measure land area, distance to the shore, living area, other covered area, and the age of structure as well as the number of bedrooms and bathrooms, and dummy variables for the presence of a swimming pool and for the conveyance of title in fee simple. Complete regression results are reported in Appendix Table A1.

larger and the estimated coefficient itself, contrary to expectations, is negative.

In comparison with either of these models, the *hybrid* model proposed in this section performs better as measured by statistical significance, and better than the standard model as measured by explained variation. Since the hybrid model utilizes the information recorded for *all* sales *and* capitalizes on the additional information provided by repeat sales, this result is to be expected. For this sample, however, the gains in statistical significance appear to be quite large.

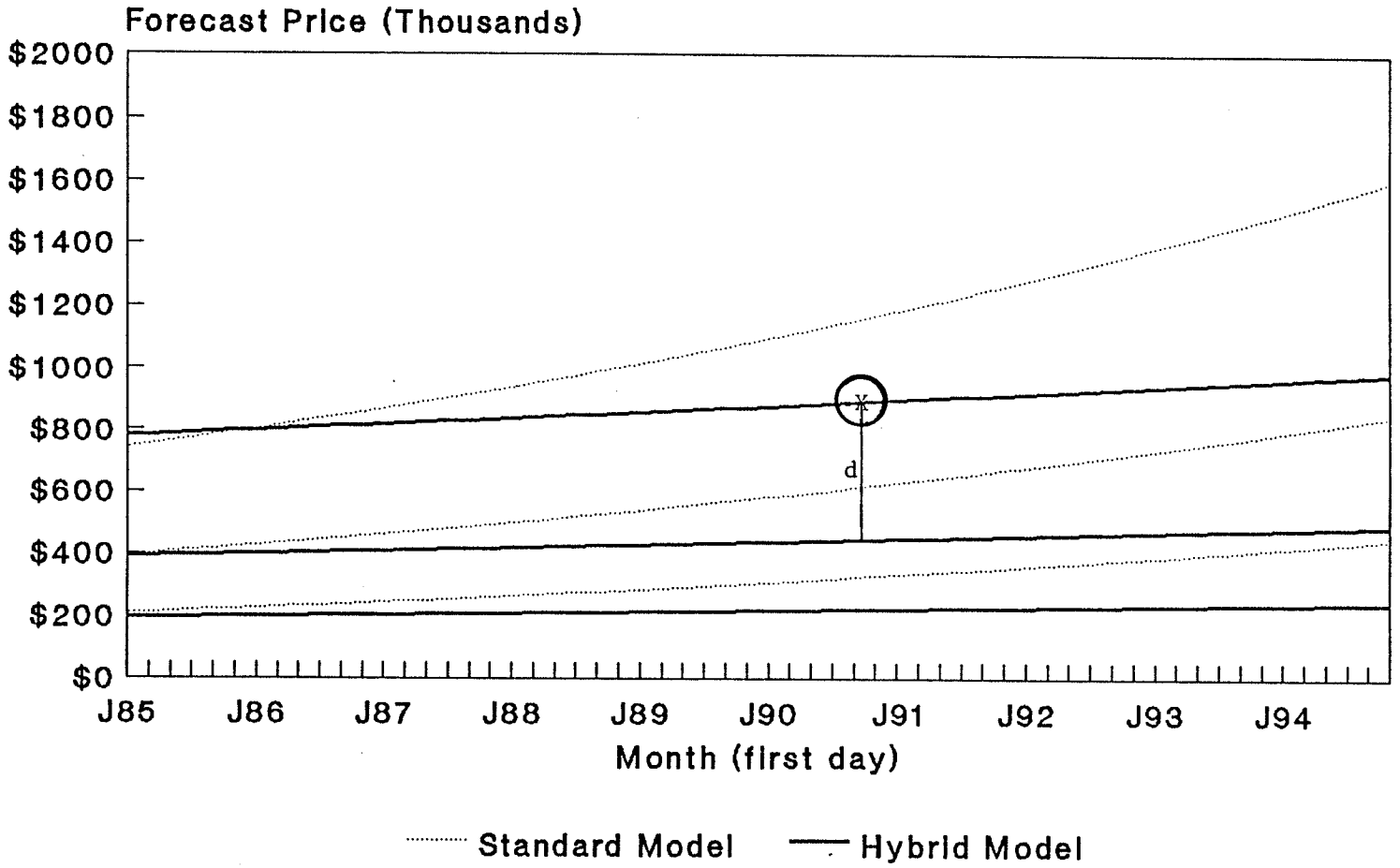
IV. Practical Application to Verification of Appraisals

The statistical model developed in the last section could have widespread applicability in the management and evaluation of real property appraisals. The statistical model could be used as a complement to traditional appraisal methods, or to verify the accuracy and reliability of traditional appraisals.

To illustrate this, Figure 1 shows a graphical representation of some of the results reported in Table 2. Specifically, it presents the 80 percent confidence interval associated with forecasts of the market price over time of a property with average physical and locational characteristics as estimated using the *standard* and the *hybrid* models discussed earlier. For reasons already noted, the confidence interval is considerably narrower for the hybrid model.

FIGURE 1

Forecast Price of Average Property
January 1985 through December 1994



Suppose, for illustration, that a lender has estimated the *hybrid* model indicated by the solid line and the associated confidence interval, and that an appraiser has subsequently presented an estimate of market value for some specific property with given values of the characteristics x included in the model. Let the symbol ⊗ on the diagram represent the appraised value.

The distance d on the diagram is a measure ex ante of the confidence of the lender in the report submitted by the appraiser. The larger the distance d, the less likely it is that both the statistical model and the appraisal are valid. As the diagram is drawn, this can be given an explicit probabilistic interpretation: if the model is correct, then the probability that the actual value of the property is equal to or greater than the appraised value ⊗ is only 0.2, corresponding to the distance d.

This interpretation suggests that the probability associated with an appraisal can be used as a management tool by lenders, insurers, public agencies, or other participants in real estate markets. If, conditional upon the statistical model, the value reported by the appraiser is relatively unlikely, then additional resources should be devoted to verification of the appraisal value. The particular decision rule adopted by a manager will depend upon the cost of

verifying an appraisal and the cost associated with appraisal error.

Presumably the costs associated with appraisal error are asymmetric. In the housing market, for example, an appraisal that is "too low" may result in lost business for a lender or insurer, while an appraisal that is "too high" will generally increase the default risk to the lender or insurer. Similarly, a low appraisal would result in lost property tax revenues for a public agency, while a high appraisal might be considered confiscatory.

Clearly it can be of significant benefit for a lender or other participant in the real estate market to evaluate the accuracy and reliability of independent property appraisals. The evidence based on this sample suggests that the hybrid model could permit substantial improvements over other more conventional approaches in the evaluation of appraised property values.

TABLE A-1

Summary Data and Regression Results
(Standard Deviation/t Ratio of Estimates in Parentheses)

Variable	Mean Value	Standard Model	Weighted Repeat Sales Model	Hybrid Model
Intercept: A	-	7.479 (10.46)**		7.532 (7.72)**
Land Area (sq.ft.):a_1	12,186 (5,266)	0.378 (5.70)**		0.301 (3.33)**
Distance to Shore (ft.):a_2	1,664 (1,079)	-0.166 (15.00)**		-0.166 (10.09)**
Living Area (sq.ft.):a_3	2,257 (757)	0.370 (6.36)**		0.455 (6.58)**
Other Covered Area (sq.ft.):a_4	557 (617)	-0.013 (2.09)*		0.019 (3.78)**
Age (years):a_5	25.4 (16.0)	-0.057 (6.64)**		-.072 (6.91)**
Bedrooms:a_6	3.57 (0.88)	-0.123 (1.85)		-0.082 (0.99)
Bathrooms:a_7	2.63 (0.79)	0.085 (1.53)		0.061 (0.93)
Fee Simple:a_8	0.69 (0.46)	0.288 (10.04)**		0.332 (9.35)**
Swimming Pool:a_9	0.51 (0.50)	0.059 (2.17)*		0.094 (2.76)**
Time (thousands of days):b	1.51 (0.78)	0.217 (9.84)**	-0.136 (1.40)	0.069 (3.96)**

REFERENCES

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, Vol 58, #304, December 1963, pp. 933-942.

Case, Bradford and John M. Quigley, "The Dynamics of Real Estate Prices," Center for Real Estate and Urban Economics Working Paper, University of California, Berkeley, October 1988.

Case, Karl E. and Robert J. Shiller, "Prices of Single Family Homes Since 1970," *New England Economic Review*, Sept/Oct 1987, pp. 44-56.

Mark, Jonathan H. and Michael A. Goldberg, "Alternative Housing Price Indices: An Evaluation," *AREUEA Journal*, Vol 12, #1, 1984, pp. 31-49.

Palmquist, Raymond B., "Alternative Techniques for Developing Real Estate Price Indices," *Review of Economics and Statistics*, Vol 62, August 1980, pp. 442-448.