

UC Berkeley

Fisher Center Working Papers

Title

The Dynamics of Real Estate Prices

Permalink

<https://escholarship.org/uc/item/1k074729>

Authors

Case, Bradford
Quigley, John M.

Publication Date

1988-10-01

Peer reviewed



Institute of
Business and
Economic Research

University of
California at
Berkeley

FISHER CENTER FOR REAL ESTATE AND URBAN ECONOMICS

WORKING PAPER SERIES

WORKING PAPER NO. 88-151

THE DYNAMICS OF REAL ESTATE PRICES

By

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.

**BRADFORD CASE
JOHN M. QUIGLEY**

WALTER A. HAAS SCHOOL OF BUSINESS

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

**Kenneth T. Rosen, Chair
Robert H. Edelstein, Co-Chair
Dwight M. Jaffee, Co-Chair**

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
Carl Shapiro, Director**

The Institute of Business and Economic Research is an organized research unit of the University of California at Berkeley. It exists to promote research in business and economics by University faculty. These working papers are issued to disseminate research results to other scholars. The authors welcome comments; inquiries may be directed to the author in care of the Center.

THE DYNAMICS OF REAL ESTATE PRICES*

by

Bradford Case

and

John M. Quigley
University of California
Berkeley

October 1988

Working Paper No. 88-151

- I. Introduction
- II. A Simple Model
- III. Empirical Analysis
- IV. Implications For Forecasting Market Prices
- V. References

* A previous version of this paper was presented at the Conference on Contemporary Problems in Tax Assessment and Appraisal, Harvard Law School, August 1988. Financial support has been provided by the Lincoln Institute of Land Policy, the International Association of Assessing Officers, and the Center for Real Estate and Urban Economics, University of California, Berkeley.

Abstract

Due to uncertainties about the specification and functional form of hedonic models of housing prices, several recent studies of housing price trends recommend confining statistical analysis to repeat sales of properties whose characteristics have not been changed (*i.e.* Case and Shiller, 1987a, 1987b).

This paper presents a methodology which combines information on the repeat sales of unchanged properties, repeat sales of improved properties, and information on single sales, all in one joint estimation.

Empirical evidence, based upon a sample containing some 418 transactions on single family houses, indicates the clear advantages of the proposed methodology, at least in one typical application.

THE DYNAMICS OF REAL ESTATE PRICES

by

Bradford Case

and

John M. Quigley

University of California

Berkeley

I. INTRODUCTION

Reports of increases in the price of real estate make front page news, but the techniques used for measuring these price changes are quite crude. For residential properties, the most widely reported price trends are those compiled by the National Association of Realtors. This information is confined to the median value of existing single family housing, as reported by the transactions of member realtors in a number of metropolitan areas.

These residential sale prices are not standardized for any characteristics of the dwellings bought and sold. For commercial properties, standardization is quite minimal; sale or rental prices are reported on a per square foot basis from survey data compiled by financial service institutions and brokerage firms.

It has been widely recognized that it is appropriate to control statistically for the varying characteristics of properties in inferring price trends (see Greenlees [1978] for a discussion), and during the past several decades a variety of hedonic techniques have been proposed to account for the important non-temporal determinants of price variation (Kain and Quigley [1970], Griliches [1971]). These techniques ultimately result in the estimation of some regression relationship between the sale price, V_t , of properties (or perhaps their rent per square foot) at time t , their physical and locational characteristics x , and some representation of time, t :

$$(1) \quad V_t = f(x, t) \quad .$$

Interpretation of this relationship depends crucially upon the inclusion of the correct set of property characteristics, x , and the correct functional form, $f(\cdot)$, for the hedonic regression. Conditional upon these two issues, however, the hedonic function can be used to disaggregate the variation in real estate prices into that attributable to changes in the characteristics of properties sold and that attributable to intertemporal variation. In particular, the statistical results can be used to forecast the market price for a standardized or "quality adjusted" property over time.

Because the set of property characteristics is not known with certainty, it has been suggested that the characteristics of each property be standardized with reference only to themselves, by confining the analysis to properties which have been sold more than once (Bailey, *et al* [1963]):

$$(2) \quad V_t/V_\tau = g(t, \tau) \quad .$$

In this formulation, changes in the selling price of a property between time τ and time t are related to the timing of the two transactions, or perhaps to the time interval $(t-\tau)$ 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 [1986] and Karl Case and Shiller [1987a, 1987b].

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 is extremely wasteful of transactions information. In any market run, the fraction of properties which are repeat sales 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. Moreover, this latter technique is inappropriate when any of the characteristics of the properties have been changed between sale dates. Although it may be possible to identify properties whose physical characteristics have changed between sales and to exclude them from the statistical analysis, it is more difficult to identify properties whose locational characteristics (for example, neighborhood or public service attributes) have changed. Identification of properties with changed characteristics requires specifying and measuring those characteristics, and it is a curious research strategy indeed that completely ignores those measurements.

Elimination of properties which sold only once and those whose characteristics had been changed reduced Palmquist's sample of single family houses in King County, Washington by two thirds, from 4,785 to 1,613. Similarly, the implementation of this research strategy reduced the sample size of house sales

available to Case and Shiller by 96 percent in Atlanta and Chicago (from 221,876 to 8,945 and from 397,183 to 15,530, respectively), by 97 percent in Dallas (from 211,638 to 6,669), and by 93 percent in San Francisco (from 121,909 to 8,066).¹

The small fraction of repeat sales in the samples analyzed by Palmquist, Mark and Goldberg, and Case and Shiller, despite the long time periods included in the analyses,² suggests that sample selectivity may have been an important phenomenon affecting the results.³

This paper presents and tests a simple model of real estate prices which includes the desirable features of both approaches to the estimation of price appreciation. On the one hand, it uses all available information on property sales, whether single or repeat transactions. On the other hand, it capitalizes on the added precision possible when there exist multiple transactions, by comparing transaction prices for the same properties whenever possible. The model makes the appropriate comparison regardless of whether or not property characteristics have been changed.

¹ The sample of single family houses analyzed by Mark and Goldberg [1984] was reduced by only 61 percent (from 4,376 to 1,695) in Fraser, Vancouver and by 57 percent (from 1,398 to 660) in Kerrisdale, Vancouver. It is not clear from their paper, however, whether dwellings whose physical or locational characteristics changed between sales were excluded from subsequent analysis.

² Palmquist's data included house sales over a 14-1/2 year period. Mark and Goldberg considered a 22 year period, and Case and Shiller considered house sales over a 16-1/2 year period.

³ Indeed, Mark and Goldberg speculate that their statistical results, comparing price indices estimated using equations (1) and (2), may have arisen "... due to the characteristics of houses being resold. This [problem] could be serious as this index [the estimation of equation (2),] uses substantially less information than [estimation of equation (1)] and is likely to be biased accordingly" [1984, p 37].

The model can thus be used to estimate price appreciation over time for a standardized unit by combining data from three kinds of samples: single transactions where one sale is observed; multiple transactions where the physical and locational characteristics of properties are unchanged; and multiple transactions where the physical or locational characteristics of properties have been modified.

II. A SIMPLE MODEL

Suppose initially property values, V_0 , vary with continuously measured qualitative and quantitative aspects of properties, say x_1 and x_2 , and discrete binary attributes, say x_3 , according to the simple exponential relation

$$(3) \quad V_0 = Ax_1^{a_1}x_2^{a_2}e^{a_3x_3}$$

where a_1 , a_2 , a_3 , and A are parameters. Suppose property values vary over time t according to demands and the relative scarcity of x_1 , x_2 , and x_3 . In the simplest representation, let the price vary continuously with time,⁴

$$(4) \quad V_t = V_0 x_1^{b_1t} x_2^{b_2t} e^{b_3tx_3}$$

The simple model expressed in Equations (3) and (4) implies that if we observe a transaction at time t , the selling price of the property is

⁴ In the text the model is specified using continuous time because the time interval in the empirical analysis below is rather short, about seven years (and the discrete time notation is cumbersome). The generalization to discrete time is straightforward:

$$(4') \quad V_t = V_0 x_1^{\sum b_{1n}T_n} x_2^{\sum b_{2n}T_n} x_3^{\sum b_{3n}T_n},$$

where T_n is a vector of dummy variables for each time period. $T_i=1$ for $0 \leq i \leq t$, and $T_i=0$ for $t < i \leq n$.

$$(5) \quad \log V_t = \log A + a_1 \log x_1 + a_2 \log x_2 + a_3 x_3 + b_1 t \log x_1 \\ + b_2 t \log x_2 + b_3 t x_3 \quad .$$

Suppose, however, we observe two sales of a property at t and τ , $t > \tau$, whose characteristics are unchanged during the interval $[\tau, t]$. From Equation (4) the selling price at t will be

$$(6) \quad \log V_t = \log V_\tau + b_1(t-\tau) \log x_1 + b_2(t-\tau) \log x_2 + b_3(t-\tau)x_3 \quad .$$

Finally, suppose we observe two sales of a property at t and τ , $t > \tau$. In this case, however, suppose the characteristics of the property are changed from (x_1, x_2, x_3) to (x_1^*, x_2^*, x_3^*) at t^* , $\tau < t^* < t$. In this case, from Equation (4) at time t^* , after the first sale is made and just before the change in property characteristics, the value of the property, $V_{t^*}^-$, is

$$(7) \quad V_{t^*}^- = V_\tau x_1^{b_1(t^*-\tau)} x_2^{b_2(t^*-\tau)} e^{b_3(t^*-\tau)x_3} \quad .$$

When the transformation is made, from Equations (3) and (4), the new value of the property, V_{t^*} , is:

$$(8) \quad V_{t^*} = V_{t^*}^- (x_1^*/x_1)^{a_1+b_1 t^*} (x_2^*/x_2)^{a_2+b_2 t^*} e^{[a_3+b_3 t^*] [x_3^*-x_3]} \quad .$$

Finally, from Equation (4) the value of the property at the time of the second sale, V_t is

$$(9) \quad V_t = V_{t^*} x_1^{*b_1(t-t^*)} x_2^{*b_2(t-t^*)} e^{b_3(t-t^*)x_3^*}$$

So, upon substitution of (7) and (8) into (9) and rearranging,

$$(10) \log V_t = \log V_\tau + a_1 \log(x_1^*/x_1) + a_2 \log(x_2^*/x_2) + a_3(x_3^* - x_3) \\ + b_1[t \log x_1^* - \tau \log x_1] + b_2[t \log x_2^* - \tau \log x_2] + b_3[t x_3^* - \tau x_3]$$

Equations (5), (6), and (10) provide alternative methods for estimating the parameters of the hedonic model for the three kinds of samples.⁵ Consistent estimates of the parameters can be obtained from any of the three samples, at least as long as the samples are random.⁶ However, if information is available for two or more samples, the relevant equations can be estimated more effi-

⁵ Using discrete time notation, the analogous equations are

$$(5') \log V_m = \log A + a_1 \log x_1 + a_2 \log x_2 + a_3 x_3 + \sum_{i=1}^m b_{1i} T_i \log x_1 + \\ \sum_{i=1}^m b_{2i} T_i \log x_2 + \sum_{i=1}^m b_{3i} T_i x_3$$

and

$$(6') \log V_m = \log V_n + \sum_{i=n}^m b_{1i} T_i \log x_1 + \sum_{i=n}^m b_{2i} T_i \log x_2 + \sum_{i=n}^m b_{3i} T_i x_3$$

for (5) and (6) respectively, where m is the period of the later sale and n is the period of the earlier sale, $T_i=1$ for $0 \leq i \leq m$; $T_i=0$ otherwise.

The equation analogous to (10) for properties sold initially at n , modified from (x_1, x_2, x_3) to (x_1^*, x_2^*, x_3^*) at r and sold again at m , is

$$(10') \log V_m = \log V_r + a_1 \log x_1^* + a_2 \log x_2^* + a_3 x_3^* \\ + \sum_{i=r}^m b_{1i} T_i \log x_1 + \sum_{i=r}^m b_{2i} T_i \log x_2 + \sum_{i=r}^m b_{3i} T_i x_3$$

⁶ Note, however, that Equation (6) provides estimates of the b 's only, not the a 's in Equation (3).

ciently for these samples by imposing the appropriate cross equation constraints. These constraints can be imposed by considering the system of three equations formed by (5), (6) and (10):

$$(11a) \begin{pmatrix} \log V_t \\ \log V_t/V_\tau \\ \log V_t/V_\tau \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

with $\beta = \beta_1 = \beta_2 = \beta_3$.

This can be written more compactly as

$$(11b) \quad Y = Z \beta + \varepsilon,$$

where each of the nine elements of Z is a (1×7) matrix:

$$\begin{aligned} 0 &= (0, 0, 0, 0, 0, 0, 0) \\ X_1 &= (1, \log x_1, \log x_2, x_3, t \log x_1, t \log x_2, t x_3) \\ X_2 &= (0, 0, 0, 0, [t-\tau] \log x_1, [t-\tau] \log x_2, [t-\tau] x_3) \\ X_3 &= (0, \log(x_1^*/x_1), \log(x_2^*/x_2), [x_3^* - x_3], Q_1, Q_2, Q_3), \end{aligned}$$

with

$$Q_1 = t \log x_1^* - \tau \log x_1$$

$$Q_2 = t \log x_2^* - \tau \log x_2$$

$$Q_3 = t x_3^* - \tau x_3$$

and where β is a (7×1) matrix with $\beta' = (\log A, a_1, a_2, a_3, b_1, b_2, b_3)$.

Estimation of Equation (11) by ordinary least squares utilizes the data from the three samples and imposes the restrictions inherent in the model:

$$(12) \hat{\beta} = (Z' Z)^{-1} Z' Y$$

However, since $X_1 \neq X_2 \neq X_3$, it follows that $\phi = E(\varepsilon'\varepsilon)$ is not diagonal. Thus the parameters can be estimated more efficiently by generalized least squares:

$$(13) \beta = (Z' \phi^{-1} Z)^{-1} (Z' \phi^{-1} Y)$$

where again

$$(14) \phi = E(\varepsilon' \varepsilon)$$

Since ε is unknown, generalized least squares estimation proceeds in two steps.⁷ First, Equation (11) is estimated by ordinary least squares, and the residuals, $r' = (r_1, r_2, r_3)$, are used to estimate ϕ :

$$(15) (1/N)E(r'r) = \hat{\phi}$$

where N is the sample size. Then, $\hat{\phi}$ is utilized to estimate the coefficients:

⁷ See Goldberger [1964, pp 262-265] and Zellner and Huang [1962] for discussion of a similar problem in the imposition of extraneous restrictions in the estimation of a set of relations. Note that Equation (11) can be expressed equivalently as three regression relationships, each estimated from a sample of $(q+r+s)$ observations. For the first equation, based upon q observations on single sales, the values of the dependent variable are $(V_q, 0_r, 0_s)$ where $V_q = (\log V_{t1}, \log V_{t2}, \dots, \log V_{tq})$, 0_r is a $(1 \times r)$ vector of zeros, 0_s is a $(1 \times s)$ vector of zeros, and the residuals are $(\varepsilon_1, 0_r, 0_s)$. Analogously, the residuals from the second and third equations, based respectively upon r and s observations on multiple sales and multiple sales of changed properties, are $(0_q, \varepsilon_2, 0_s)$ and $(0_q, 0_r, \varepsilon_3)$ where 0_q is a $(1 \times q)$ vector of zeros. These sets of residuals can be treated as if they arise from three regressions "seemingly unrelated" in the sense of Zellner [1962].

$$(16) \quad \tilde{\beta} = (Z' \hat{\phi}^{-1} Z)^{-1} (Z' \hat{\phi}^{-1} Y) .$$

III. EMPIRICAL ANALYSIS

The model is estimated from observations on the sales of single detached housing from the Kahala neighborhood of Honolulu, Hawaii during the period October 1980 through October 1987. The neighborhood consists of about 1100 residential parcels and is bounded by Kahala Beach, the Waialae Golf Course, and the Kalaniana'ole Highway. The sample for this analysis consists of every sale in this neighborhood during the seven year period: 424 residential transactions involving 315 separate properties. Six of these transactions involving five properties were deleted from the sample because of missing data.

The 1980's have been 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 \$370,000. In 1980, the average transaction was for \$361,000. In 1987, unadjusted sales prices averaged \$845,000.

Table 1 provides summary information on these properties. It is worth noting that, of the 418 transactions recorded during the seven year period, only 108 were multiple sales of the same property. Thus an analysis based only upon Equation (2) using multiple sales would utilize only about one quarter of the information on house sales. Further, only 47 of those 108 multiple sales were of properties whose characteristics were unchanged between the first and the second sale. Clearly an estimate of Equation (2) based upon all 108 repeat sales would be misleading, while an estimate based upon only 47 sales out of 418 would be quite imprecise.

TABLE 1

Summary Data on Housing Transactions
(Standard Deviations in Parentheses)

	Single Sales	Multiple Sales** Identical Properties	Multiple Sales** Changed Properties	All Transactions Pooled
Land Area: x_1 (thousands of sq.ft.)	12.09 (5.05)	11.78 (4.28)	13.00 (6.82)	12.19 (5.27)
To Shore: x_2 (thousands of feet)	1.71 (1.07)	1.64 (1.17)	1.42 (1.05)	1.66 (1.08)
Living Area: x_3 (thousands of sq.ft.)	2.20 (0.70)	2.16 (0.63)	2.63 (0.98)	2.26 (0.76)
Other Covered Area: x_4 (thousands of sq.ft.)	0.50 (0.51)	0.61 (0.45)	0.79 (1.04)	0.56 (0.62)
Age: x_5 (years)	27.85 (14.54)	23.40 (16.99)	14.79 (17.81)	25.44 (15.99)
Time: t (thousands of days*)	1.37 (0.79)	1.91 (0.59)	1.93 (0.57)	1.51 (0.78)
Fee Simple***	0.68	0.74	0.74	0.69
Selling Price: V (thousands of dollars)	452.69 (427.62)	696.89 (864.53)	813.83 (773.26)	532.85 (568.63)
Median Sale Price (thousands of dollars)	350.00	399.00	610.00	370.00
Number of Observations	310	47	61	418

* Thousands of days elapsed from September 30, 1980 to date of sale.

** Property characteristics and selling prices at the time of the second sale.

*** Fraction of sales conveying title in fee simple.

As indicated in Table 1, the dwellings sold in the neighborhood averaged slightly more than 2000 square feet in living area, on lots of more than 12000 square feet. The dwellings are rather new, with an average age of about 25 years, and are well situated -- less than 2000 feet from the coast on average. On average, they also have about 560 square feet of covered area, carports, roofed patios and the like.

These averages mask a great deal of variation in the characteristics of properties sold, as noted by the standard deviations reported in the table. Of equal importance is the variation in the types of sales across the three categories of transaction. The physical and locational characteristics of properties which were sold only once are similar to those which were sold more than once and whose characteristics were unchanged. Sale prices in the latter category averaged almost \$700,000, or about \$250,000 more than those which were sold only once. The average sale date for identical houses sold twice was in December 1985, about a year later than the average for houses which were sold once.

The largest differences are between dwellings sold more than once whose characteristics were changed and transactions in the other two categories. Dwellings sold more than once, but whose characteristics were changed, are larger than others, by about 500 square feet, and have larger lots, by about 1000 square feet. They are ten years newer, are closer to the shore, and generally seem to be of higher quality. They sold for more than \$800,000 on average. Properties in this category are a combination of those which have been upgraded, sometimes slightly, and those which have been substantially improved (and in some cases even rebuilt).

Table 2 reports the coefficients of the dynamic price model estimated separately for the three samples. The first column reports estimates of Equation (5), β_1 , using the 310 properties sold one time. The second column reports the coefficients of Equation (6), β_2 , estimated using the 47 repeat sales of properties whose characteristics were unchanged. The third column reports the coefficients of Equation (10), β_3 , estimated using the 61 properties sold more than once whose characteristics were changed between the first and the second sale. In Equation (5) eight of the thirteen coefficients are highly significant, and the equation explains a large proportion of the variation in selling prices. The regressions based upon repeat sales perform less well, in part because the samples are so small. For the 47 repeat sales, in spite of the very high correlation ($r^2=0.92$) between the actual selling price and the price predicted by the model, the coefficients are quite imprecisely estimated: none of the six coefficients is significantly different from zero at even the .05 level. Similarly, for the sample of repeat sales with changed physical characteristics, only two coefficients are significant by conventional criteria even though the model explains a very large proportion of the variation in selling prices ($r^2=0.90$).

Table 3 compares the conventional model, estimated using the entire sample of 418 sales, with the model and estimation technique proposed in this analysis. The results reported in column 1 ignore the multiple sales in the data and treat the entire sample as a group of unrelated transactions. The so called "naive model" reports the coefficients of Equation (5), β^* , estimated using the full sample of 418 sales. A comparison of this model with that reported in column 1 of Table 2 indicates that the larger sample improves the statistical properties of estimates somewhat. Nine of the thirteen coefficients are statistically

TABLE 2
Dynamic Price Model Estimated for Different Samples
(t ratios in parentheses)

	Single Sales, β_1	Multiple Sales	
		Identical Properties, β_2	Changed Properties, β_3
logA	7.335 (11.59)**		
a ₁	0.279 (3.03)**		
a ₂	-0.105 (6.65)**		
a ₃	0.467 (5.52)**		0.860 (3.32)**
a ₄	-0.038 (4.64)**		-0.039 (1.79)
a ₅	-0.026 (1.51)		0.001 (0.02)
a ₆	0.035 (0.71)		0.054 (0.31)
b ₁ x10 ³	0.024 (0.59)	0.306 (1.79)	0.208 (1.74)
b ₂ x10 ³	-0.032 (3.85)**	-0.023 (0.45)	-0.012 (0.60)
b ₃ x10 ³	0.007 (0.14)	-0.307 (1.47)	-0.230 (1.51)
b ₄ x10 ³	0.034 (5.20)**	-0.035 (0.58)	0.028 (1.31)
b ₅ x10 ³	-0.013 (1.26)	-0.040 (0.78)	-0.049 (1.97)
b ₆ x10 ³	0.169 (5.02)**	0.222 (1.45)	0.210 (2.12)*
R ²	0.832	0.277	0.839
r ²	0.832	0.920	0.900
Observations	310	47	61

Note: r^2 = Correlation coefficient between actual selling price and price predicted by model.

* Coefficient significantly different from zero at .05 level.

** Coefficient significantly different from zero at .01 level.

TABLE 3
Dynamic Price Model Estimated for Pooled Samples
(t ratios in parentheses)

	Naive Model, β^*	Correctly Specified Model OLS, $\hat{\beta}$	GLS, $\tilde{\beta}$
logA	7.511 (11.86)**	7.528 (9.57)**	7.525 (14.40)**
a_1	0.212 (2.26)*	0.135 (1.32)	0.085 (1.33)
a_2	-0.111 (6.74)**	-0.109 (6.15)**	-0.115 (10.31)**
a_3	0.528 (5.82)**	0.620 (7.09)**	0.684 (13.40)**
a_4	-0.041 (4.69)**	-0.033 (4.29)**	-0.032 (7.38)**
a_5	-0.020 (1.07)	-0.024 (1.41)	-0.021 (2.15)*
a_6	0.042 (0.75)	0.025 (0.46)	0.029 (0.92)
$b_1 \times 10^3$	0.078 (1.85)	0.100 (2.45)*	0.129 (5.45)**
$b_2 \times 10^3$	-0.034 (4.12)**	-0.032 (4.02)**	-0.029 (6.56)**
$b_3 \times 10^3$	-0.060 (1.16)	-0.093 (1.84)	-0.127 (4.33)**
$b_4 \times 10^3$	0.032 (4.92)**	0.029 (4.34)**	0.027 (6.85)**
$b_5 \times 10^3$	-0.023 (2.35)*	-0.028 (2.82)**	-0.034 (6.21)**
$b_6 \times 10^3$	0.162 (4.67)**	0.188 (5.55)**	0.195 (9.97)**
R^2	0.830	0.999	n.a.
r^2	0.830	0.882	0.881
Observations	418	418	418

Note: r^2 = Correlation coefficient between actual selling price and price predicted by model.

* Coefficient significantly different from zero at .05 level.

** Coefficient significantly different from zero at .01 level.

significant, and the model explains approximately the same proportion of the variance in sales prices.

Columns 2 and 3 report the results when the panel nature of the sample is recognized and is incorporated explicitly into the estimation. Column 2 reports the ordinary least squares results ($\hat{\beta}$, Equation 12), and column 3 reports the generalized least squares results ($\tilde{\beta}$, Equation 16) using the residuals to estimate ϕ by Equation 15.

The GLS estimated coefficients are quite precisely estimated indeed. Ten out of the thirteen coefficients are statistically significant at the .01 level, and the simple correlation between the actual sale price and its predicted value is almost 0.9. A comparison of the OLS and GLS results clearly indicates the increased precision arising from the latter technique. Utilizing all the information and all the restrictions improves the precision of the estimates.

IV. IMPLICATIONS FOR FORECASTING MARKET PRICES

The precision of forecasts derived from these models depends upon factors not presented in Tables 2 and 3, namely the entire variance-covariance matrices of the estimated parameters, the forecast values of the exogeneous variables, and the standard error of the estimate.⁸ Figure 1 incorporates these factors. The dotted line indicates the price pattern for the average property in the sample, forecast from the point of means of the full data set (roughly July

⁸ That is, the width of any confidence interval is proportional to $s^2 [1 + \tilde{X}(\tilde{M})\tilde{X}']$ where s is the standard error of the regression, \tilde{X} is the vector of forecast values for the exogeneous variables, and \tilde{M} is the variance-covariance matrix of the estimated coefficients.

1985) through January 1995 using the regression results reported in column 1 of Table 3, *i.e.*, using the "naive model" ignoring the multiple sales aspects of the data. As the diagram indicates, the price of the average house is forecast to rise from almost \$450,000 to almost \$1,250,000 during the 9-1/2 year period. The figure also presents the 80 percent confidence interval for that forecast; the dotted line indicates the band which contains the true unknown price with 80 percent confidence. Note that the interval is not symmetric about the mean forecast and that the width of the confidence interval increases over time, reflecting the greater uncertainty about future prices.

The dashed line presents the same information utilizing the repeat sales method recommended by Case and Shiller and others. Using the results from column 2, Table 2, it reports the forecast price for the average house during the period 1985-1995 (increasing from almost \$420,000 to almost \$820,000) and the 80 percent confidence interval for that forecast.

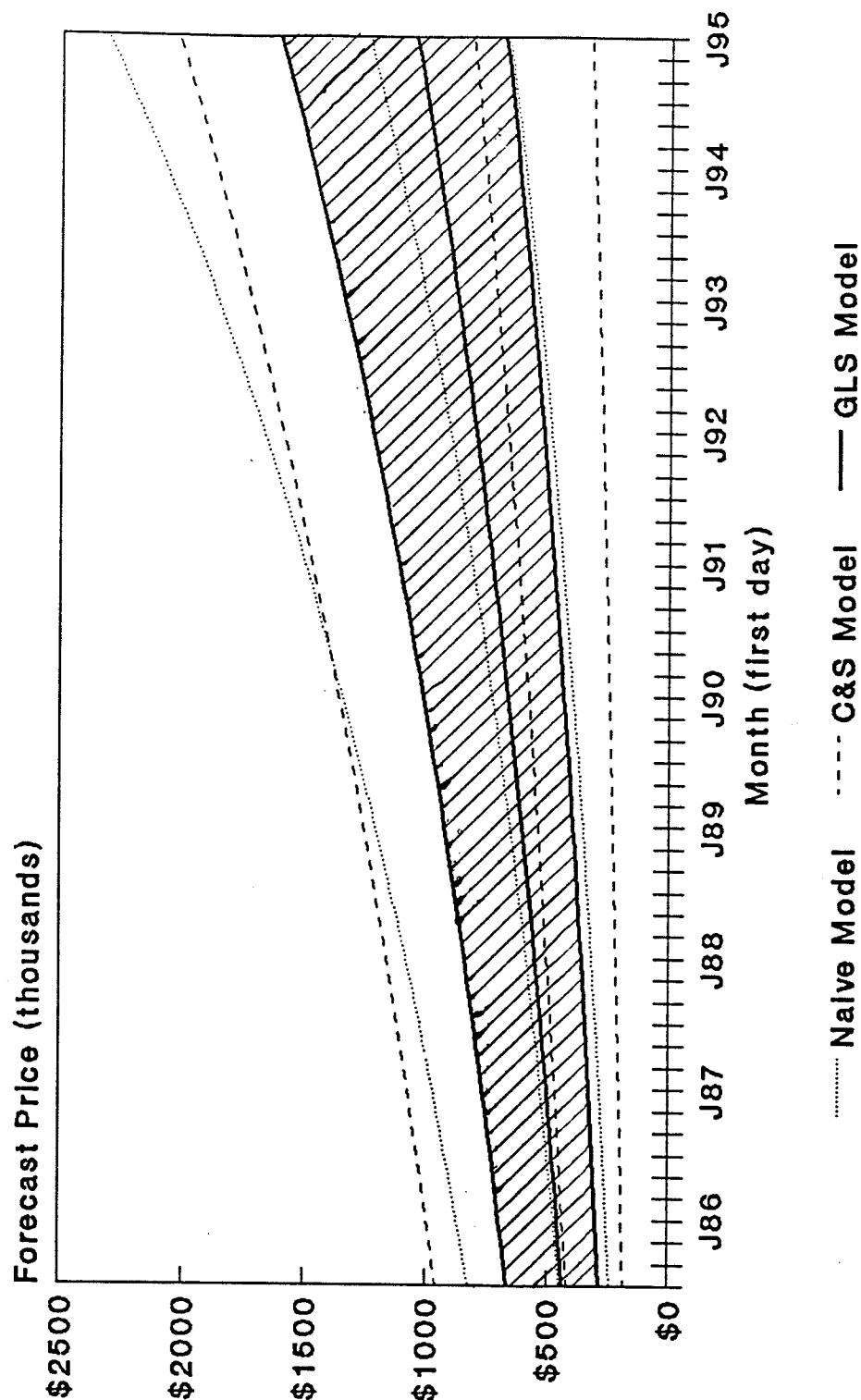
From the figure, it seems that the naive model has more desirable properties than the repeat sales model, at least for this sample of data. The principal reason is that repeat sales of unchanged properties are only a small fraction of total sales (only 47 out of 418 sales in this sample). This reduces the confidence with which forecasts of price changes can be made.

The solid lines indicate the price forecast and the confidence interval obtained using the techniques proposed in this paper. The GLS model generates a forecast sale price that increases from about \$430,000 to about \$1,050,000 during the period 1985-1995. The reported confidence interval is much narrower, indicating much more precision in forecasting market prices.

The evidence based on this sample suggests that for the problem of inferring the market prices of unsold properties or of forecasting the future prices of those properties, this hybrid technique offers practical as well as theoretical advantages over the other more conventional approaches.

FIGURE 1

Forecast Price of Average Property July 1985 through January 1995



V. 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, Dec 1963, pp. 933-942.

Case, Karl E., "The Market for Single Family Homes in the Boston Area," *New England Economic Review*, May/June 1986, pp. 38-48.

Case, Karl E. and Robert J. Shiller, "Prices of Single Family Homes Since 1970: New Indexes for Four Cities," NBER Working Paper #2393, National Bureau of Economic Research, Cambridge, MA., Sept 1987 [1987a].

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

Goldberger, Arthur S., *Econometric Theory*, New York, John Wiley and Sons, 1964.

Greenless, John S., "An Empirical Evaluation of the CPI Home Purchase Index, 1973-1978," *AREUEA Journal*, Vol 10, 1982, pp. 1-24.

Griliches, Zvi, ed, *Price Indices and Quality Change*, Cambridge: Harvard University Press, 1971.

Kain, John F. and John M. Quigley, "Measuring the Value of Housing Quality," *Journal of the American Statistical Association*, Vol 65, #440, June 1970, pp. 532-548.

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.

Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, Vol 57, 1962, pp. 348-368.

Zellner, Arnold and D.S. Huang, "Further Properties of Efficient Estimators for Seemingly Unrelated Regressions," *International Economic Review*, Vol 3, Sept 1962, pp. 300-313.