THE DYNAMICS OF REAL ESTATE PRICES

Bradford Case and John M. Quigley *

Abstract—Several studies of housing price trends recommend confining statistical analysis to repeat sales of residential properties. Recently, price indices derived from these techniques have formed the basis for inferences about the "efficiency" of housing markets.

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

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

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, compiled by the National Association of Realtors, are 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 (1982) 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:

$$V_t = f(x, t). \tag{1}$$

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 produce indices of the market price for a standardized or "quality adjusted" property over time.[1]

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

$$V_t/V_\tau = g(t, \tau). \tag{2}$$

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).[2]

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 mar-

Received for publication October 25, 1988. Revision accepted for publication February 27, 1990.

* College of William and Mary and University of California, Berkeley, respectively.

Financial support for this research was 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.

[1] See Sirmans (1982) for a review and criticism of U.S. Census Bureau procedures for estimating "quality adjusted" indices of new home prices using these hedonic procedures.

[2] Indices so computed have also been relied upon by Karl Case and Shiller (1989) in subsequent tests of the "efficiency" of metropolitan housing markets.

ket 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 or those whose characteristics had been changed reduced Palmquist's sample of single family houses in King County, Washington by about two-thirds. Similarly, the implementation of this research strategy reduced the sample size of house sales available to Case and Shiller by 96% in Atlanta and Chicago, by 97% in Dallas, and by 93% in San Francisco.[3]

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 (fourteen to twenty-two years) suggests that sample selectivity may have been an important phenomenon affecting the results.[4]

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, as long as the changes have been measured.

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 between sale dates by rehabilitation or other forms of investment.

II. A Simple Model

Suppose that at time zero property values, V_0, vary with continuously measured qualitative and quantitative aspects of properties, say x_1 and x_2, and some discrete binary attribute, say x_3, according to the simple exponential relation

$$V_0 = Ax_1^{a_1}x_2^{a_2}e^{a_3x_3}, \tag{3}$$

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:

$$V_t = V_0x_1^{b_1t}x_2^{b_2t}e^{b_3tx_3}. \tag{4}$$

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

$$\begin{aligned}\log V_t = \log A &+ a_1 \log x_1 + a_2 \log x_2 \\ &+ a_3x_3 + b_1t \log x_1 \\ &+ b_2t \log x_2 + b_3tx_3.\end{aligned} \tag{5}$$

In a more realistic context, the intertemporal relationship can be expressed over discrete time intervals; the selling price of a property is

$$\begin{aligned}\log V_t = \log A &+ a_1 \log x_1 + a_2 \log x_2 + a_3x_3 \\ &+ \sum_{n=1}^{t} b_{1n} \log x_1 + \sum_{n=1}^{t} b_{2n} \log x_2 \\ &+ \sum_{n=1}^{t} b_{3n}x_3.\end{aligned} \tag{5$'$}$$

[3] The sample of single family houses analyzed by Mark and Goldberg (1984) was reduced by only 61% in Fraser, Vancouver and by 57% 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.

[4] 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 [, based on equation (2),] uses substantially less information than [an index based on equation (1)] and is likely to be biased accordingly" [1984, p. 37].

If we observe two sales of a property at t and τ, $t > \tau$, whose characteristics are unchanged during the interval $[\tau, t]$, then with continuous time the selling price at t is

$$\log V_t = \log V_\tau + b_1(t-\tau)\log x_1 + b_2(t-\tau)\log x_2 + b_3(t-\tau)x_3, \tag{6}$$

and with discrete time intervals, the selling price is

$$\log V_t = \log V_\tau + \sum_{n=\tau+1}^{t} b_{1n}\log x_1 + \sum_{n=\tau+1}^{t} b_{2n}\log x_2 + \sum_{n=\tau+1}^{t} b_{3n}x_3. \tag{6$'$}$$

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 the case of continuous time, 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, $V_{\overline{t^*}}$, is

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

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

$$V_{t^*} = V_{\overline{t^*}}(x_1^*/x_1)^{a_1+b_1t^*}(x_2^*/x_2)^{a_2+b_2t^*} \times e^{[a_3+b_3t^*][x_3^*-x_3]}. \tag{8}$$

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

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

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

$$\begin{aligned}\log V_t = \log V_\tau &+ a_1\log(x_1^*/x_1) \\ &+ a_2\log(x_2^*/x_2) + a_3(x_3^*-x_3) \\ &+ b_1[t\log x_1^* - \tau\log x_1] \\ &+ b_2[t\log x_2^* - \tau\log x_2] \\ &+ b_3[tx_3^* - \tau x_3].\end{aligned} \tag{10}$$

In the discrete case, the analogous expression is

$$\begin{aligned}\log V_t = \log V_\tau &+ a_1\log(x_1^*/x_1) + a_2\log(x_2^*/x_2) \\ &+ a_3(x_3^*-x_3) + \sum_{n=\tau+1}^{t^*-1} b_{1n}\log x_1 \\ &+ \sum_{n=t^*}^{t} b_{1n}\log x_1^* \\ &+ \sum_{n=\tau+1}^{t^*-1} b_{2n}\log x_2 \\ &+ \sum_{n=t^*}^{t} b_{2n}\log x_2^* + \sum_{n=\tau+1}^{t^*-1} b_{3n}x_3 \\ &+ \sum_{n=t^*}^{t} b_{3n}x_3^*.\end{aligned} \tag{10$'$}$$

Equations (5), (6), and (10), or their primed versions, provide alternative methods for estimating the parameters of the hedonic model for the 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. However, if information is available for two or more samples, the relevant equations can be estimated more efficiently by imposing the appropriate cross equation constraints. In the case of continuous time, this can be expressed quite compactly. The system of three equations formed by (5), (6), and (10) can be expressed as a single stacked equation:

$$\begin{pmatrix}\log V_t \\ \log V_t/V_\tau \\ \log V_t/V_\tau\end{pmatrix} = \begin{pmatrix}X_1 \\ X_2 \\ X_3\end{pmatrix}\beta + \begin{pmatrix}\epsilon_1 \\ \epsilon_2 \\ \epsilon_3\end{pmatrix}. \tag{11a}$$

This can be written more compactly as

$$\mathbf{Y} = \mathbf{Z}\beta + \epsilon, \tag{11b}$$

where each of the three elements of $\mathbf{Z}$ is a (1×7) matrix:

$$\begin{aligned}X_1 &= (1, & &\log x_1, & &\log x_2, & &x_3, & &t\log x_1, & &t\log x_2, & &tx_3) \\ X_2 &= (0, & &0, & &0, & &0, & &[t-\tau]\log x_1, & &[t-\tau]\log x_2, & &[t-\tau]x_3) \\ X_3 &= (0, & &\log(x_1^*/x_1), & &\log(x_2^*/x_2), & &[x_3^*-x_3], & &Q_1, & &Q_2, & &Q_3),\end{aligned}$$

with

$$Q_1 = t \log x_1^* - \tau \log x_1$$
$$Q_2 = t \log x_2^* - \tau \log x_2$$
$$Q_3 = tx_3^* - \tau x_3,$$

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:

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

As noted by Goldberger (1964), however, the restrictions can be imposed more efficiently. The system of equations can also be expressed as

$$\mathbf{Y} = \begin{pmatrix} X_1^* & 0 & 0 \\ 0 & X_2^* & 0 \\ 0 & 0 & X_3^* \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix} + \begin{pmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \epsilon_3^* \end{pmatrix} \tag{13a}$$

where X_i^* is X_i after deletion of the columns whose coefficients in equation i are zero, and β_i^* is β_i with the appropriate rows deleted. More compactly, this can be expressed as

$$\mathbf{Y} = \mathbf{Z}^*\beta^* + \epsilon^*. \tag{13b}$$

Since $X_1^* \neq X_2^* \neq X_3^*$, it follows that $\phi = E(\overset{*}{\epsilon}'\overset{*}{\epsilon}) \neq \sigma^2\mathbf{I}$, with constant σ^2. Thus the parameters of (13b) can be estimated more efficiently by generalized least squares:

$$\overset{*}{\beta} = \left(\overset{*}{\mathbf{Z}}'\phi^{-1}\overset{*}{\mathbf{Z}}\right)^{-1}\left(\overset{*}{\mathbf{Z}}'\phi^{-1}\mathbf{Y}\right), \tag{14}$$

where again

$$\phi = E(\overset{*}{\epsilon}'\overset{*}{\epsilon}). \tag{15}$$

Since $\overset{*}{\epsilon}$ is unknown, generalized least squares estimation proceeds in two steps.[5] First, equation (11) is estimated by ordinary least squares, and the residuals, $\mathbf{r}' = (r_1, r_2, r_3)$, are used to compute $\hat{\phi}$:

$$(1/N)E(\mathbf{r}'\mathbf{r}) = \hat{\phi}, \tag{16}$$

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

$$\tilde{\beta} = \left(\overset{*}{\mathbf{Z}}'\hat{\phi}^{-1}\overset{*}{\mathbf{Z}}\right)^{-1}\left(\overset{*}{\mathbf{Z}}'\hat{\phi}^{-1}\mathbf{Y}\right). \tag{17}$$

When the intertemporal effects are modelled as a series of discrete intervals, the estimation procedure is analogous. The observations in (5′), (6′), and (10′) can be "stacked" and the coefficients of (11) can be estimated by ordinary least squares. More efficient estimates may be obtained by generalized least squares estimation of (13).

[5] 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.

III. Empirical Analysis

The empirical analysis comparing the proposed estimation technique with the so-called repeat sales method relies upon a unique body of data on every housing transaction in a single neighborhood during a seven-year period. We compare the coefficient estimates obtained under the different methods, and the implications for housing price indices are derived therefrom. In the text we report estimates of the continuous time model (that is, equations (5), (6), and (10) above) largely because they are simpler to compare and interpret. The comparison of price indices, reported in section IV, is based upon the discrete time model (that is, equations (5′), (6′), and (10′) above). The coefficients of this more complicated model are reported in the appendix.

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 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 1980s 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 \$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 measured characteris-

TABLE 1.—SUMMARY DATA ON HOUSING TRANSACTIONS (STANDARD DEVIATIONS IN PARENTHESES)

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

[a] Property characteristics and selling prices at the time of the second sale.
[b] Thousands of days elapsed from September 30, 1980 to date of sale.
[c] Fraction of sales conveying title in fee simple.

tics 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.

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 12,000 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 unchanged 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. 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 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 (17)).

TABLE 2.—DYNAMIC PRICE MODEL ESTIMATED FOR DIFFERENT SAMPLES (*t*-RATIOS IN PARENTHESES)

		Multiple Sales	
	Single Sales, β_1	Unchanged Properties, β_2	Changed Properties, β_3
$\log A$	7.335		
	(11.59)[b]		
a_1	0.279		
	(3.03)[b]		
a_2	−0.105		
	(6.65)[b]		
a_3	0.467		0.860
	(5.52)[b]		(3.32)[b]
a_4	−0.038		−0.039
	(4.64)[b]		(1.79)
a_5	−0.026		0.001
	(1.51)		(0.02)
a_6	0.035		0.054
	(0.71)		(0.31)
$b_1 \times 10^3$	0.024	0.306	0.208
	(0.59)	(1.79)	(1.74)
$b_2 \times 10^3$	−0.032	−0.023	−0.012
	(3.85)[b]	(0.45)	(0.60)
$b_3 \times 10^3$	0.007	−0.307	−0.230
	(0.14)	(1.47)	(1.51)
$b_4 \times 10^3$	0.034	−0.035	0.028
	(5.20)[b]	(0.58)	(1.31)
$b_5 \times 10^3$	−0.013	−0.040	−0.049
	(1.26)	(0.78)	(1.97)
$b_6 \times 10^3$	0.169	0.222	0.210
	(5.02)[b]	(1.45)	(2.12)[a]
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.
[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

TABLE 3.—DYNAMIC PRICE MODEL ESTIMATED FOR POOLED SAMPLES (*t*-RATIOS IN PARENTHESES)

	Naive Model, β^*	Correctly Specified Model	
		OLS, $\hat{\beta}$	GLS, $\tilde{\beta}$
$\log A$	7.511	7.528	7.525
	(11.86)[b]	(9.57)[b]	(14.40)[b]
a_1	0.212	0.135	0.085
	(2.26)[a]	(1.32)	(1.33)
a_2	−0.111	−0.109	−0.115
	(6.74)[b]	(6.15)[b]	(10.31)[b]
a_3	0.528	0.620	0.684
	(5.82)[b]	(7.09)[b]	(13.40)[b]
a_4	−0.041	−0.033	−0.032
	(4.69)[b]	(4.29)[b]	(7.38)[b]
a_5	−0.020	−0.024	−0.021
	(1.07)	(1.41)	(2.15)[a]
a_6	0.042	0.025	0.029
	(0.75)	(0.46)	(0.92)
$b_1 \times 10^3$	0.078	0.100	0.129
	(1.85)	(2.45)[a]	(5.45)[b]
$b_2 \times 10^3$	−0.034	−0.032	−0.029
	(4.12)[b]	(4.02)[b]	(6.56)[b]
$b_3 \times 10^3$	−0.060	−0.093	−0.127
	(1.16)	(1.84)	(4.33)[b]
$b_4 \times 10^3$	0.032	0.029	0.027
	(4.92)[b]	(4.34)[b]	(6.85)[b]
$b_5 \times 10^3$	−0.023	−0.028	−0.034
	(2.35)[a]	(2.82)[b]	(6.21)[b]
$b_6 \times 10^3$	0.162	0.188	0.195
	(4.67)[b]	(5.55)[b]	(9.97)[b]
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.
[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

The GLS estimates coefficients are quite precisely estimated indeed. Ten out of the thirteen coefficients are statistically significant at the 0.01 level, and the simple correlation between the actual sale price and its predicted value is almost 0.9. Utilizing all the information and all the restrictions improves the precision of the estimates.

Similar results are reported for the discrete time model in the appendix.

IV. Implications for Estimating Price Indices

The precision of price indices derived from these models depends upon the entire variance-covariance matrices of the estimated parameters.

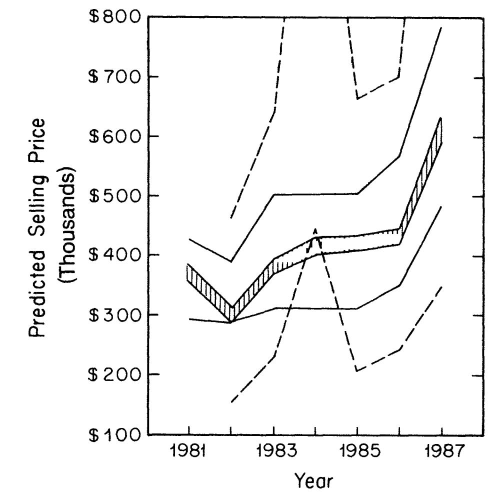
Figure 1 incorporates these factors using the discrete time model reported in the appendix. The solid lines indicate the 95% confidence interval for the price of a standardized property during the period 1981 through 1987. This estimate was obtained using the regression results in table A3, i.e., using the "naive" model ignoring the multiple sales aspects of the data.

The dashed lines present the same 95% confidence interval estimated from the repeat sales method recommended by Case and Shiller and others. These results rely upon the regression for unchanged properties reported in table A2. 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. Since repeat sales of unchanged properties are only a small fraction of total sales (only 47 out of 418 sales in this sample), the confidence with which price indices can be estimated is substantially reduced.

The shaded area indicates the confidence interval obtained using the techniques proposed in this paper. The GLS model generates a price index that increases by 66% during the period 1981–1987. The reported confidence interval is much narrower using this method, indicating much more precision in estimating market price trends.

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

FIGURE 1.—PRICE INDEX FOR AVERAGE PROPERTY



APPENDIX

This appendix presents the results of the discrete time variant of the model reported in the text. In this variant, separate slope and intercept coefficients are estimated for each year 1981 through 1987. The variable definitions and sample sizes are the same as those reported in table 1 in the text. Table A1 presents the coefficients of the dynamic price model estimated using the 310 observations on single sales (analogous to the results reported in column 1, table 2 in the text). Table A2 presents the coefficients of the models estimated on the multiple sales of identical and changed properties (analogous to the results reported in columns 2 and 3, respectively, of table 2). Table A3 presents the results of the "naive" model ignoring multiple sales information (analogous to the results reported in column 1 of table 3). Table A4 presents the results estimated for the pooled sample using the

TABLE A1.—DYNAMIC PRICE MODEL ESTIMATED FOR SINGLE SALES β_1; 310 OBSERVATIONS (t-RATIOS IN PARENTHESES)

	x_1	x_2	x_3	x_4	x_5	x_6
a	0.245	−0.121	0.473	−0.012	−0.001	0.144
	(2.76)[a]	(8.22)[b]	(5.43)[b]	(0.68)	(0.09)	(2.81)[a]
b_{1982}	0.073	0.025	−0.109	−0.032	—	−0.003
	(0.56)	(0.60)	(0.68)	(1.08)	—	(0.04)
b_{1983}	−0.188	−0.001	0.192	−0.149	0.127	0.091
	(1.17)	(0.02)	(0.97)	(3.10)[b]	(4.28)[b]	(0.99)
b_{1984}	0.000	−0.027	0.059	0.142	−0.095	−0.081
	(0.00)	(0.53)	(0.31)	(1.88)	(3.03)[a]	(0.91)
b_{1985}	0.276	−0.037	−0.315	−0.015	0.006	0.102
	(2.14)[a]	(0.76)	(2.12)[a]	(0.22)	(0.20)	(1.02)
b_{1986}	0.054	−0.011	−0.042	0.028	−0.039	0.184
	(0.54)	(0.42)	(0.33)	(1.00)	(1.25)	(1.81)
b_{1987}	−0.199	−0.023	0.271	−0.034	0.047	0.067
	(2.32)[a]	(1.10)	(2.23)[a]	(1.29)	(2.03)[a]	(0.68)
R^2	0.999			log A	7.643	
					(12.42)[b]	

[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

TABLE A2.—DYNAMIC PRICE MODEL ESTIMATED FOR MULTIPLE SALES, β_2 and β_3 (t-RATIOS IN PARENTHESES)

	x_1	x_2	x_3	x_4	x_5	x_6
	I. Unchanged Properties (47 observations): $R^2 = 0.758$					
b_{1982}	−0.423	−0.018	0.419	0.277	—	−0.329
	(0.51)	(0.18)	(0.45)	(0.46)	—	(0.41)
b_{1983}	1.266	−0.072	−1.434	0.126	−0.062	0.044
	(1.26)	(0.25)	(1.32)	(0.42)	(0.13)	(0.04)
b_{1984}	−0.039	0.214	0.236	0.022	−0.489	0.649
	(0.03)	(0.25)	(0.13)	(0.05)	(0.52)	(0.45)
b_{1985}	−1.258	−0.167	1.274	−0.018	0.462	−0.321
	(0.91)	(0.19)	(0.70)	(0.05)	(0.58)	(0.26)
b_{1986}	0.432	−0.144	−0.469	−0.086	0.120	0.243
	(0.38)	(0.62)	(0.33)	(0.38)	(0.49)	(0.45)
b_{1987}	−0.018	−0.075	0.148	0.104	−0.015	−0.123
	(0.02)	(0.54)	(0.10)	(0.64)	(0.14)	(0.23)
	II. Changed Properties (61 observations): $R^2 = 0.972$					
a	—	—	0.576	−0.037	0.005	0.229
	—	—	(4.34)[b]	(1.16)	(0.31)	(1.96)
b_{1982}	−0.129	0.185	0.001	0.021	—	−0.405
	(1.00)	(1.02)	(0.22)	(0.81)	—	(0.88)
b_{1983}	0.114	−0.240	0.008	−0.003	0.081	0.536
	(0.86)	(1.30)	(1.89)	(0.19)	(1.39)	(1.17)
b_{1984}	0.159	−0.215	0.013	−0.010	−0.004	0.173
	(1.38)	(1.37)	(1.10)	(0.14)	(0.39)	(0.76)
b_{1985}	−0.074	0.068	0.011	0.013	−0.013	0.257
	(0.87)	(0.55)	(2.37)[a]	(0.44)	(1.78)	(0.87)
b_{1986}	−0.097	0.122	0.007	−0.013	−0.000	−0.224
	(1.14)	(1.09)	(1.85)	(1.05)	(0.09)	(0.92)
b_{1987}	0.104	−0.138	0.008	0.014	0.007	−0.184
	(0.83)	(1.26)	(0.56)	(1.23)	(1.50)	(1.14)

[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

TABLE A3.—DYNAMIC PRICE MODEL ESTIMATED FOR POOLED SAMPLE, β^*; 418 OBSERVATIONS (t-RATIOS IN PARENTHESES)

	x_1	x_2	x_3	x_4	x_5	x_6
a	0.250	−0.119	0.502	−0.018	−0.006	0.145
	(2.55)[b]	(6.93)[b]	(4.94)[b]	(0.88)	(0.44)	(2.39)[b]
b_{1982}	0.113	−0.005	−0.132	−0.018	—	−0.004
	(0.83)	(0.16)	(0.78)	(0.58)	—	(0.04)
b_{1983}	−0.150	−0.019	0.176	−0.088	0.077	0.161
	(1.01)	(0.58)	(0.97)	(2.11)[a]	(2.86)[b]	(1.60)
b_{1984}	−0.029	0.002	0.045	0.097	−0.039	−0.159
	(0.19)	(0.04)	(0.26)	(1.22)	(1.32)	(1.68)
b_{1985}	0.201	−0.009	−0.219	−0.033	0.022	0.111
	(1.42)	(0.17)	(1.36)	(0.44)	(0.78)	(1.18)
b_{1986}	0.101	−0.023	−0.102	0.026	−0.009	0.100
	(0.98)	(0.98)	(0.83)	(0.98)	(0.35)	(1.07)
b_{1987}	−0.082	−0.044	0.166	−0.038	0.033	0.055
	(0.88)[a]	(2.48)[b]	(1.48)	(1.86)	(1.51)	(0.62)
R^2	0.998			$\log A$	7.376	
					(12.32)[b]	

[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

TABLE A4.—DYNAMIC PRICE MODEL ESTIMATED FOR POOLED SAMPLE, $\hat{\beta}$ AND $\tilde{\beta}$; 418 OBSERVATIONS (t-RATIOS IN PARENTHESES)

	x_1	x_2	x_3	x_4	x_5	x_6
			I. Ordinary Least Squares Estimates $\hat{\beta}$: $R^2 = 0.999$			
a	0.297	−0.116	0.388	−0.044	−0.007	0.189
	(4.43)[b]	(7.86)[b]	(9.59)[b]	(3.78)[b]	(1.00)	(4.47)[b]
b_{1982}	−0.010	0.011	0.001	0.000	—	−0.141
	(0.60)	(0.52)	(0.35)	(0.02)	—	(1.75)
b_{1983}	0.008	−0.054	0.009	−0.015	0.045	0.216
	(0.35)	(2.11)[a]	(3.58)[b]	(1.20)	(3.34)[b]	(2.44)[a]
b_{1984}	−0.003	0.005	0.008	0.008	−0.001	−0.014
	(0.11)	(0.12)	(1.44)	(0.32)	(0.09)	(0.17)
b_{1985}	−0.017	0.003	0.011	0.011	−0.010	0.051
	(0.54)	(0.07)	(5.23)[b]	(0.58)	(1.96)	(0.56)
b_{1986}	0.026	−0.037	0.005	−0.004	−0.004	0.126
	(2.32)[a]	(2.05)[a]	(3.06)[b]	(0.51)	(1.35)	(1.48)
b_{1987}	0.061	−0.024	0.001	−0.004	0.004	−0.097
	(4.93)[b]	(1.56)	(0.43)	(0.46)	(1.38)	(1.29)
log A	7.838					
	(11.64)[b]					
			II. Generalized Least Squares Estimates, $\tilde{\beta}$: $R^2 = 0.999$			
a	0.281	−0.115	0.419	−0.049	−0.001	0.213
	(6.29)[b]	(13.14)[b]	(22.57)[b]	(12.28)[b]	(0.26)	(13.07)[b]
b_{1982}	−0.021	0.025	0.002	0.018	—	−0.273
	(3.53)[b]	(2.90)[b]	(2.76)[b]	(3.68)[b]	—	(7.09)[b]
b_{1983}	0.009	−0.078	0.009	−0.009	0.055	0.348
	(1.13)	(7.96)[b]	(11.31)[b]	(2.81)[b]	(10.71)[b]	(8.21)[b]
b_{1984}	0.037	−0.056	0.010	0.004	−0.001	0.072
	(2.59)[a]	(2.92)[b]	(5.94)[b]	(0.52)	(0.65)	(2.10)[a]
b_{1985}	−0.060	0.062	0.012	0.008	−0.010	0.067
	(4.47)[b]	(3.29)[b]	(16.98)[b]	(1.53)	(7.38)[b]	(1.71)
b_{1986}	0.023	−0.032	0.006	−0.005	−0.004	0.048
	(4.88)[b]	(4.35)[b]	(10.78)[b]	(2.38)[a]	(5.00)[b]	(1.41)
b_{1987}	0.076	−0.026	−0.001	0.001	0.005	−0.197
	(14.92)[b]	(3.94)[b]	(1.27)	(0.25)	(6.30)[b]	(7.52)[b]
log A	7.738					
	(16.97)					

[a] Coefficient significantly different from zero at the 0.05 level.
[b] Coefficient significantly different from zero at the 0.01 level.

fully specified dynamic model. The table presents the ordinary least squares regression results and the generalized least squares regression results (analogous to those reported in columns 2 and 3, respectively, of table 3).

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* 58 (Dec. 1963), 933–942.

Case, Karl E., "The Market for Single Family Homes in the Boston Area," *New England Economic Review* (May/June 1986), 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. 1987a.

______, "Prices of Single Family Homes Since 1970," *New England Economic Review* (Sept./Oct. 1987b), 44–56.

______, "The Efficiency of the Market for Single Family Homes," *American Economic Review* 79 (Mar. 1979), 125–137.

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

Greenlees, John S., "An Empirical Evaluation of the CPI Home Purchase Index, 1973–1978," *AREUEA Journal* 10 (1982), 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* 65 (June 1970), 532–548.

Mark, Jonathan H., and Michael A. Goldberg, "Alternative Housing Price Indices: An Evaluation," *AREUEA Journal* 12 (1) (1984), 31–49.

Palmquist, Raymond B., "Alternative Techniques for Developing Real Estate Price Indices," this REVIEW 62 (Aug. 1980), 442–448.

Sirmans, C. F. (ed.), *Research in Real Estate*, Volume 2 (Greenwich, CT: JAI Press, 1982).

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