A Probabilistic Model for Price Levels in Discontinuous Markets

Cary Webb
Chicago State University
Department of Mathematics
Ninety-Fifth Street at King Drive
Chicago, Illinois 60628, USA

Abstract. A linear model is constructed that allows estimation of price levels in markets wherein goods are traded only irregularly. The model is tested with real estate data collected by the state of Arizona. The assumption of the model is that prices of individual properties follow independent random walks. The results presented here amplify work done by Court in 1938 and by Bailey, Muth and Nourse in 1963.

1. Introduction

Methods for computing price levels in the real estate market or any other market wherein a given good is traded only irregularly are less precise than those methods for computing price levels in other types of markets such as stocks or consumer goods. In both these markets, most goods are traded regularly, that is, goods which are perfect substitutes (e.g., shares of IBM) are traded at regular intervals (e.g., on trading days of the NYSE). Clearly similar conditions are not met in the real estate market, since any given property is usually quite distinct and is traded only irregularly.

To be sure, there are several price indexes currently used in the real estate market, for example, the mean price of new single-family houses [1], a hedonic (i.e., quality-adjusted) price index for new single-family houses [2], and the mean price of existing single-family houses [3], [4]. Of these, only the hedonic index in [2] takes account of changes through time in the set of properties traded. In particular, there is no definitive index that measures price levels of existing single-family houses.

Because of the relative lack of price information in real estate markets, the reliability of a real estate price index is necessarily less

Measurement in Economics
ed. by W. Eichhorn

than the reliability of financial indexes and of the Consumer Price Index (CPI). For this reason, one must realize that the accuracy of financial indexes and the CPI can never be duplicated in the real estate market or any market wherein a given good transacts only irregularly.

The purpose here is to establish a methodology that will produce the most accurate real estate index and to quantify the reliability of this index. The method to be developed will apply also to any other market wherein a sufficiently large subset of the goods are known to have transacted at least twice.

It may well turn out that the most accurate possible index for a particular market is not accurate enough for one of its intended purposes. In such a case, the work that we present here and elsewhere [15] will suggest the amount of additional information necessary to achieve the desired reliability.

One line of research on real estate and similar markets has produced the hedonic indexes. These were pioneered by Andrew Court in his 1938 study of automobiles [5]. More recent work on hedonic indexes is contained in [6], [7], and [8].

It is possible to derive the model developed in this paper from Court's 1938 model. Historically, however, models based on data from a minimum of two sales had a separate development. The earliest example is a chaining technique, dating from 1927 [9]. A further development of the chaining technique is contained in [10]. In 1963, Bailey, Muth, and Nourse introduced a regression model [11] which avoided the drawback of the chain method in not using information about the price level at a given time that may be contained in the returns on investments made in properties sold afterwards.

In this paper we present a regression method that is a refinement of [11] which avoids problems with heteroscedasticity. The underlying assumption that allows us to infer optimal estimates is that real estate prices follow a random walk (in discrete time). By this we mean that the return to every period is chosen from the same distribution. In other words, the random variables that are the returns of the various properties in the various periods are identical and identically distributed (i.i.d.).

2. Theory

In order to build the price index model, let P_i denote the (unobserved) market price level in period i, i = 0,1,...,n, and let w_i be the logarithm of the associated price relative, $w_i = \log(P_i/P_{i-1})$. Define w_{ij} to be the sum of the w_k's, k = i+1,...,j.

An implication of the random walk assumption is that the distributions for which the w_i's are drawn are independent and identically distributed. It is also true, for the same reason, that the distributions from which the quantities $w_{ij}(j-i)^{-1/2}$ are drawn are homoscedastic. Suppose property 1 was bought in period i for price p_i and next sold in period j for p_j. Let $Y_{ij\ell} = \log(p_j/p_i)(j-i)^{-1/2}$. Then $Y_{ij\ell}$ is an estimate for $w_{ij}(j-i)^{-1/2}$. In other words

$$Y_{ij\ell} = (j-i)^{-1/2}(w_{i+1} + \cdots w_j) + e_\ell, \qquad (1)$$

where e_ℓ is a disturbance term. When a regression is performed, it will be "weighted least squares" because of the factor $(j-i)^{-1/2}$ in equation (1).

Equation (1) can be written in the conventional form,

$$Y = b_1x_1 + \cdots + b_nx_n + e, \qquad (2)$$

where Y is of the form $Y_{ij\ell}(j-i)^{-1/2}$, $b_k = w_k$, and x_k is the 'dummy' variable, $x_k = (j-i)^{-1/2}$, k = i+1,...,j, and $x_k = 0$, otherwise. The object is to estimate the coefficients b_k, k = 1,...,n.

It is important to notice in equation (2) that there is no constant term. Hence we cannot expect the residuals from the regression to total (exactly) zero. There will be evidence in the data we analyze, however, that the actual mean residual is not significantly different from zero (section 3).

Assuming individual properties follow independent random walks, the true disturbances are uncorrelated and have mean zero. It is a consequence that the least-squares estimates for the b_k's are unbiased and of minimum variance among all unbiased estimates that are linear transformations of the vector of y values. These optimality properties are direct consequences of the Gauss-Markov theorem [12].

The end results of our calculations are the values P_i/P_{i-1}, the market price relatives. It is, however, the value $\log(P_i/P_{i-1})$ that is estimated by b_i. This means that if c is the expected value of b_i the expected value of P_i/P_{i-1} is $e^{v/2}e^c$ where v is the variance of the estimator b_i. In our data (section 3), however, the factor $e^{v/2}$ is very nearly 1. (For the three data sets that were analyzed, it was 1.0012, 1.0002, and 1.0001.) The difference between using the correction factor $e^{v/2}$ and not using it, is several orders of magnitude smaller than the standard deviations of the estimates (which average .0483 for the annual series and larger for the finer series). It was, therefore, neglected.

Notwithstanding the influence of the Gauss-Markov theorem in the formulation of our model, in [15] it will become apparent there is some natural multicollinearity in the model. This is especially true, in real estate markets, whenever the time periods are one month in duration. Therefore we should not be surprised if estimates of monthly price levels are unreliable. This will be true of the data we analyze in the next section.

Suppose there are N buy-sell pairs in the data. Then there is a system of N equations of the form of equation (2). Such a system can be represented with matrices, $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$, where $\mathbf{y}$ and $\mathbf{e}$ are N-dimensional column vectors, $\mathbf{X}$ is an N by n matrix, and $\mathbf{b}$ is an n-dimensional column vector.

There are two necessary and sufficient conditions on the data that guarantee the nonsingularity of the matrix $\mathbf{X}'\mathbf{X}$. First of all, there must be a transaction in every period for which we want to estimate a return. Second, in every period at least one of the properties must be held and not sold (i.e., in every period we must be strictly between the buying date and the selling date of one of the observed properties). We call data that satisfy these two conditions "connected". (This term has been applied to similar conditions; for a discussion see [13].)

It is not hard to see that these two conditions are necessary for non-singularity. That the conditions are also sufficient can be derived from a theorem in the theory of M-matrices [14]. This derivation is contained in section 4. We are grateful to Charles Johnson and the editors of Linear Algebra and Its Applications for this observation.

It is simple enough to modify the model to accommodate either kind of singularity. First of all, if there is no transaction in a certain period, one merely omits the corresponding variable from the model.

Of course, the regression procedure produces no estimate for the return and price level for the missing period. The estimate for the return of the preceding period includes the return of the "missing" period.

Secondly, suppose all properties purchased prior to a certain period either have been sold or are sold in that period. Then the regression procedure must be broken into two parts. The first part estimates the returns and prices up to the pivotal period. The second part estimates the other returns and prices.

The model can be illustrated with a small example (Table 1).

Table 1

Property	Period of Trade	Price
1	1	1000
1	3	1500
2	1	1200
2	2	1600
3	2	2000
3	3	2500

For these data the **Y** vector and the **X** matrix are the following

$$\mathbf{Y} = \begin{bmatrix} 2^{-1/2}\log(1500/1000) \\ \log(1600/1200 \\ \log(2500/2000) \end{bmatrix} = \begin{bmatrix} .2867 \\ .2877 \\ .2231 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 2^{-1/2} & 2^{-1/2} \\ 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

From this it follows that the least squares estimate of $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is,

$$\mathbf{b} = \begin{bmatrix} .2614 \\ .1968 \end{bmatrix}.$$

Thus, setting $P_0 = 1$, we have $P_1/P_0 = e^{.2614} = 1.30$ and $P_2/P_1 = e^{.1968} = 1.22$

A limiting case for the model is when all holding periods are one period in duration. In this case, one of the variables x_i is 1 and the remainder are 0. In other words, the rows of the $\mathbf{X}$ matrix each contain a single 1. The remaining entries are 0.

In this situation it is easy to see the matrix $\mathbf{X}'\mathbf{X}$ is diagonal. The entry in the (k,k) position is the number of properties held during the kth period. Therefore the market return for period k is the geometric mean return of the individual properties held in this period. In this sense, the market return is known without the possibility of error. The non-zero variance of the return is merely a measure of the dispersion of the yields of the individual properties. Hence it is a measure of how much we can expect the market return to vary over time.

3. Empirical Results

Our data is the set of all transactions in Cochise County, Arizona, involving single-family housing which traded at least twice in the 78-month interval from January 1971 through June 1977. This data was originally collected by the state government of Arizona. We are grateful to Roger Ibbotson for making it available to us.

In the data there are 798 properties that changed hands at least twice in different months. Of these, 662 changed hands only twice. There were 116, 17, and 3 properties that changed hands 3, 4, and 5 times, respectively. None changed hands more than 5 times. These transactions constitute a total of 957 holding periods. The mean duration of a holding period was 22.991 months.

There were no transactions recorded in December 1971 and so in estimating monthly returns, month 12 has been omitted. Therefore, the estimated return for November 1971 is really an estimate for the two-month return over November and December.

The model was used to estimate monthly, quarterly, semiannual, and annual market returns. The least-squares estimates are presented in Table 2. Included are the derived price index numbers which are also displayed graphically in Figure 1. As is usual, the price in the base period was set to 100.

The most striking feature in the four series of index numbers is the differences in their volatility. There is a distinct rise in volatility from the annual series, through the intermediate series, to the monthly series.

For each series the mean standard error was computed (Table 3). Included in this table is the mean return for each series and the ratio of mean return to mean standard error. These ratios correspond closely to what we usually refer to as a t-statistic.

It should be noted that the mean return is not exactly the mean of the b_k's because the b_k's are the estimates for the logarithms of the price relatives. If these price relatives are close to 1, then w_k is close to the return from period k-1 to period k. For our numbers, the mean return is slightly larger than the mean of the b_k's. We have chosen to display the mean returns because they have more intuitive and practical appeal.

The ratio of the mean return to mean standard error declines monotonically, as we would expect, from a value of 3.76 for the annual series to one of .17 for the monthly series. Even the semiannual series is only 1.12. Therefore, our confidence in the annual series is relatively high but in the quarterly and monthly series is low. The semiannual series merits at best only moderate confidence.

One expects R^2 to decrease as the number of explanatory variables decreases, but this is only partially true for our results (Table 4). R^2 does indeed decrease from .28 in the case of the monthly series to .20 for the quarterly series and semiannual series. But then there is an increase to .23 for the annual series. (These same observations apply to the adjusted R^2 values except that the adjusted R^2 for the annual series is greater than the other adjusted R^2's.) The explanation of this increase is that the returns for the holding periods that are contained in the same calendar year are to some extent atypical of the other returns. When these shorter holding periods are omitted, the explanatory power of the model increases. Regressions that delete presumed outliers are caried out later in the paper. They suggest we can probably expect a significant increase in accuracy by being able to identify outliers properly.

The serial correlation of each of the four series is displayed in Table 5. All but the annual series exhibits negative correlation ranging from -.3132 for the monthly series to -.5290 for the semiannual series. The confidnece that the correlation cannot be explained by chance fluctuation in the estimates ranges from 99% for the monthly and quarterly series to 95% for the semiannual series.

The annual series, on the other hand, exhibits a positive serial correlation of .4741. There are only 5 returns in this series, however, so although that figure does suggest a certain amount of positive correlation, for this number of returns it is significant only at the less than 75% level.

A deeper understanding of the model would lead us to expect negatively correlated return estimates, even though the true returns are not correlated. This subject is explored in [15] where negative correlation is explained as part of a unified theory that contains theoretical confidence intervals for the coefficient estimates as functions only of the parameters n and N and the standard deviation of the y variables.

Since there is no constant term in the model we cannot expect the residuals to total exactly zero. The mean residual, however, is .027 with a standard deviation of .466. Therefore the mean is not significantly different from zero. This is consistent with the assumption that the true disturbances have a mean of exactly zero.

The distribution of residuals is more peaked than the normal distribution with 91% within one standard deviation of the mean. In their tails, the residuals are skewed appreciably to the right with 8% to the right of one standard deviation from the mean and only 1% to the left of one standard deviation from the mean.

Observations whose residuals are greater than 1 correspond to properties whose prices have grown 172% more than is explained by the model. 6% of the observations are in this category. If we assume for the moment that these observations are outliers because, perhaps major additions or other improvements have been made in the properties, then the right-skewness of the residuals largely disappears and their (small) positive bias also is diminished.

The coefficient estimates for the annual series together with their standard errors and t-statistics for the null-hypothesis are given in Table 6. One notes the t-statistics range from 1.81 to 8.27. Most of this variation is explained by the variation in coefficient estimates and not by the variation in standard errors all of which are in the range .052 plus or minus .012.

It is likely that much of what variation there is in the standard errors, a steady decrease from .0627 to .0397 from 1972 to 1976, can be explained by the increase in portfolio size throughout most of the interval of study. This increase is monotonic from 136 in 1972 to 352 in 1975. (There is a decrease to 278 in 1976.) The relationship between portfolio size and accuracy is investigated more thoroughly in [15].

In order to investigate the increase in accuracy from higher quality data, the Cochise data which produced the annual series were analyzed further by deleting the 42 out of 672 observations that produced residuals either greater than 1 or less than -1. (Recall, these correspond to properties that appreciated more than 172% or less than 37% of that of the market.) The rationale behind this approach is that these observations are quite likely outliers for reasons that have already been discussed.

Table 7 displays the results of applying the model to the edited data. The mean standard error is .0196 which implies a 95% confidnece of being, on average, within 4.3% of the true return in any given year. This compares with 11.2% for the unedited data. This is a significant increase in reliability.

The ending price level is 173.6. We are 95% confident that the true ending price is within $(5^{1/2})(4.3) = 9.6$ of the true ending price level. This compares with 95% confidence of being within $(5^{1/2})(11.2) = 25.0$ of the true ending price level with unedited data. Hence, we expect significantly more reliable results if the quality of the data can be moderately increased.

For the edited data, the mean residual is .004 with a standard deviation of .183. In other words, the mean differs only insignificantly from zero. This is strong evidence that the true disturbances have been precisely zero.

In section 2 the values of the variables x_k were specified to assure the homoschedasticity of the model. In order to confirm empirically the correctness of this specification, the market returns were estimated with an unweighted least-squares regression (which is the method employed in [11]). Compared with results from the weighted least-squares regression, (a) the mean residual increased 130%, from .027 to .062, and (b) the standard deviation of the residuals increased 18% from .466 to .550. Both of these results are indications that our estimates become poorer when the specification is changed. For the same reasons, these results are also confirmation that the i.i.d. hypothesis on returns (i.e., the random walk assumption) is appropriate.

4. Connectivity Theorem

The purpose of this section is to demonstrate the equivalence of connected data and the nonsingularity of the matrix $\mathbf{X}'\mathbf{X}$. Our method will be to reduce the problem to a theorem in the theory of M-matrices.

The first state in the reduction is to multiply $\mathbf{X}$ on the right by the n by n+1 matrix $\mathbf{A} = (a_{ij})$ where $a_{ii} = -1$, $a_{i,i+1} = 1$, and all other entries are 0. Clearly, $\mathbf{W} = \mathbf{XA}$ has the same rank as $\mathbf{X}$. It is not hard to see the entries of $\mathbf{W}$ are, (a) $w_{k\ell} = -(j-i)^{-1/2}$ if the holding period began in period $\ell = i$ and ended in period j, (b) $w_{k\ell} = (j-i)^{-1/2}$ if the holding period began in period i and ended in period $\ell = j$, and (c) $w_{k\ell} = 0$ otherwise.

The second stage in the reduction is to multiply $\mathbf{W}$ on the left by the diagonal matrix $\mathbf{D}$ that transforms each entry of $\mathbf{W}$ of the form $-(j-i)^{-1/2}$ into -1. Clearly, the rank of $\mathbf{Z} = \mathbf{DW} = \mathbf{DXA}$ is the same as the rank of $\mathbf{W}$, hence, the same rank as $\mathbf{X}$.

Let n_{ij} be the (i,j)th entry inthe (n+1)-dimensioal square matrix, $\mathbf{Z}'\mathbf{Z}$. It is not hard to see that if $i < j$ then n_{ij} is minus the number of observations with purchase date in period i and selling date in period j. n_{ij} is the number (without sign change) of the transactions, both "buys" and "sells", that occurred in period i. Thus, a diagonal element of $\mathbf{Z}'\mathbf{Z}$ is the sum of the other elements in the same row (or column) with the sign changed. This means, of course, that the trace of $\mathbf{Z}'\mathbf{Z}$ is the total number of transactions represented by the data.

In terms of **Z'Z**, the first condition of connectedness, that in every period there was at least one transaction, means every diagonal element of **Z'Z** is nonzero. The second condition of connectedness, that at all times at least one property is contained in the portfolio being observed, means that **Z'Z** is not a direct sum of two submatrices.

That **Z'Z** is an M-matrix is a consequence of (ch. 6, Theorem 4.67 in [14]), and that **Z'Z** has rank n if the data is connected is a consequence of (ch. 6, Theorem 4.16 in [14]). The converse is obviously true. Therefore we have the following theorem.

Connectivity Theorem. The **X** matrix in the price index model is of full rank if and only if the data are connected.

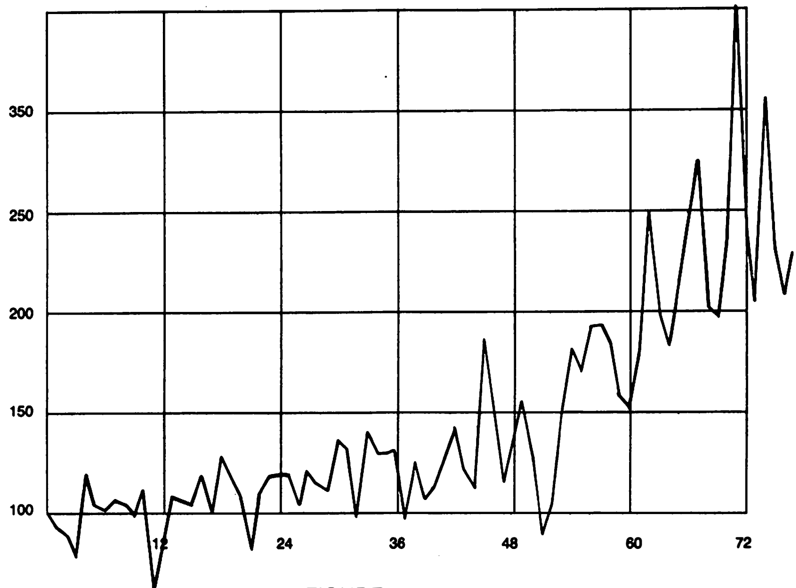

FIGURE 1a
ESTIMATES OF MONTHLY PRICE LEVELS

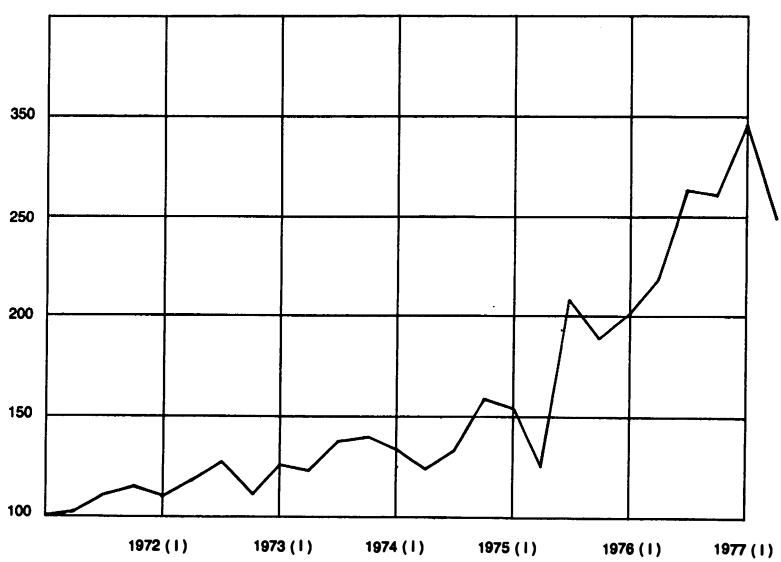

FIGURE 1b
ESTIMATES OF QUARTERLY PRICE LEVELS

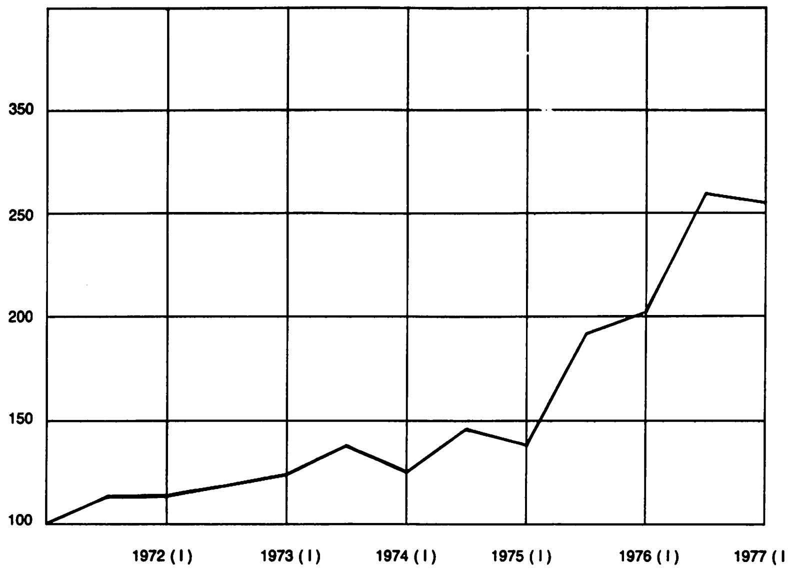

FIGURE 1c
ESTIMATES OF SEMI-ANNUAL PRICE LEVELS

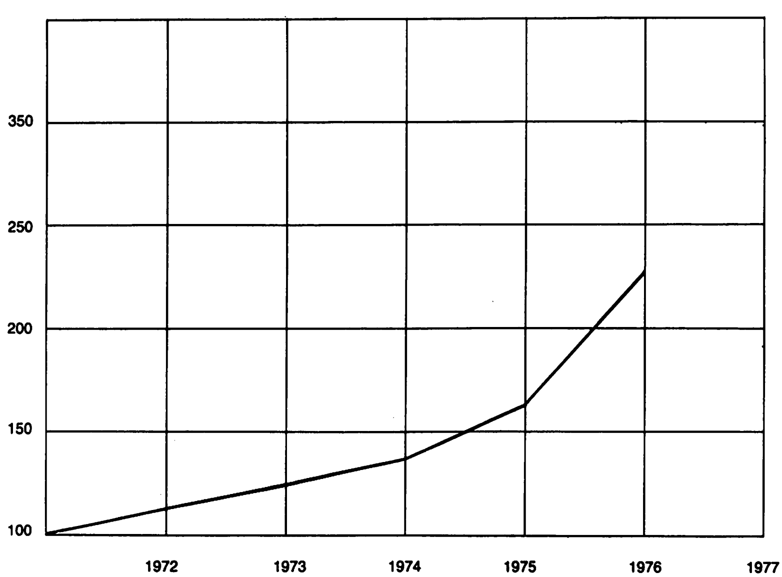

FIGURE 1d
ESTIMATES OF ANNUAL PRICE LEVELS

Table 2a

MONTHLY MARKET RETURNS AND PRICE LEVELS

Month	Return	Price
1/71	-	100.0
2/	-.0924	90.76
3/	-.0276	88.26
4/	-.1392	75.98
5/	.5898	120.79
6/	-.1393	103.97
7/	-.0261	101.26
8/	.0528	106.60
9/	-.0180	104.69
10/	-.0453	99.95
11& 12/	.1244	112.38
1/72	-.4565	61.08
2/	.7695	108.08
3/	-.0200	105.92
4/	-.0151	104.32
5/	.1431	119.24
6/	-.1522	101.09
7/	.2639	127.77
8/	-.0608	120.00
9/	-.0958	108.51
10/	-.2458	81.84
11/	.3305	108.89
12/	.0844	118.07
1/73	.0144	119.77
2/	-.0002	119.75
3/	-.1417	102.78
4/	.1701	120.27
5/	-.0472	114.58
6/	-.0252	111.69
7/	.2255	136.88
8/	-.0269	133.21
9/	-.2723	96.94
10/	.4638	141.91
11/	-.0865	129.63
12/	-.0050	128.98
1/74	.0109	130.39
2/	-.2483	98.01
3/	.2920	126.62
4/	-.1677	105.38
5/	.0767	113.47
6/	.1191	126.98
7/	.1344	144.04
8/	-.1589	121.16
9/	-.0755	112.02
10/	.6900	189.30
11/	-.1820	154.86
12/	-.2540	115.52
1/75	.1730	135.50
2/	.1566	156.72

(cont'd)

Table 2a (cont'd)

3/	-.1785	128.75
4/	-.3117	88.61
5/	.1774	104.33
6/	.4210	148.26
7/	.2376	183.47
8/	-.0732	170.04
9/	.1451	194.71
10/	.0013	194.96
11/	-.0545	184.34
12/	-.1329	159.83
1/76	-.0352	154.20
2/	.1758	181.31
3/	.4061	254.94
4/	-.2101	201.38
5/	-.0829	184.68
6/	.1373	210.02
7/	.1661	244.91
8/	.1501	281.66
9/	-.2764	203.81
10/	-.0206	199.62
11/	.1907	237.67
12/	.5446	367.11
1/77	-.3502	238.55
2/	-.1239	208.98
3/	.4822	309.76
4/	-.2431	234.46
5/	-.1071	209.36
6/	.1383	238.30

Table 2b

QUARTERLY MARKET RETURNS AND PRICE LEVELS

Quarter	Return	Price
1971 (I)		100.00
(II)	.0258	102.58
(III)	.0899	111.80
(IV)	.0312	115.29
1972 (I)	-.0437	110.25
(II)	.0750	118.51
(III)	.0822	128.85
(IV)	-.1357	110.85
1973 (I)	.1392	126.29
(II)	-.0217	123.55
(III)	.1173	138.05
(IV)	.0187	140.62
1974 (I)	-.0404	134.94
(II)	-.0775	124.49
(III)	.0720	133.46
(IV)	.1935	159.28
1975 (I)	-.0304	154.43
(II)	-.1842	125.98
(III)	.6517	208.09
(IV)	-.0835	190.72
1976 (I)	.0546	201.14
(II)	.0902	219.29
(III)	.2027	263.73
(IV)	-.0057	262.24
1977 (I)	.1297	296.24
(II)	-.1590	249.14

Table 2c

SEMI ANNUAL MARKET RETURNS AND PRICE LEVELS

Half-year		Return	Price
1971	(I)	--	100.00
	(II)	.1259	112.59
1972	(I)	.0039	113.03
	(II)	.0520	118.91
1973	(I)	.0440	124.14
	(II)	.1124	138.09
1974	(I)	-.0905	125.59
	(II)	.1624	146.01
1975	(I)	-.0498	138.75
	(II)	.3862	192.33
1976	(I)	.0524	202.42
	(II)	.2846	260.03
1977	(I)	-.0176	255.44

Table 2d

YEARLY MARKET RETURN AND PRICE LEVELS

Year	Return	Price
1971	--	100.00
1972	.1204	112.04
1973	.1182	125.29
1974	.0981	137.58
1975	.1815	162.55
1976	.3885	225.71

Table 3

ACCURACY OF INDEX NUMBERS

Price Series	Mean Return r	Mean Std. Error e	$\frac{r}{e}$
Annual	.1814	.0483	3.76
Semi-Annual	.0888	.0791	1.12
Quarterly	.0477	.1231	.39
Monthly	.0373	.2232	.17

Table 4

R^2 OF THE PRICE SERIES

Price Series	R^2	Adjusted R^2 *
Annual	.227	.222
Semi-Annual	.202	.192
Quarterly	.201	.180
Monthly	.277	.216

*For the method of adjusting, see [12].

Table 5

SERIAL CORRELATION IN THE RETURN SERIES

Series	Serial Correlation	Level of Confidence
Monthly	-.3131	99%
Quarterly	-.4882	99%
Semi-annual	-.5290	95%
Annual	.4741	Less than 75%

Table 6

RELIABILITY OF ANNUAL RETURNS

Year	Coefficient Estimate	Standard Error	t - statistic
1972	.1137	.0627	1.81
1973	.1117	.0506	2.22
1974	.0936	.0491	1.91
1975	.1668	.0398	4.19
1976	.3282	.0397	8.27
mean	.1628	.0483	3.68

Table 7

COCHISE COUNTY, 1971-76

Edited Data Set

Year	Coefficient Estimate	Standard Error	t - Statistic	Return Estimate	Price Level
1972	.1001	.0250	4.00	.1053	110.5
1973	.1101	.0201	5.48	.1164	123.4
1974	.0761	.0198	3.84	.0790	133.2
1975	.1184	.0165	7.19	.1183	148.9
1976	.1532	.0168	9.11	.1656	173.6
Mean	.1116	.0196	5.92	.1180*	---

*Geometric mean

REFERENCES

[1] US Bureau of the Census, *New One-Family Houses Sold and for Sale*, Washington, D.C.

[2] US Bureau of the Census, *Price Index of New One Family Houses*, Washington, D.C.

[3] US Bureau of the Census, *Annual Housing Survey*, Part C, Financial Characteristics of the Housing Inventory, Washington, D.C.

[4] National Association of Realtors, *Existing Home Sales*, Washington, D.C.

[5] Court, Andrew T. 1939. "Hedonic Price Indexes with Automobile Examples", *The Dynamnics of Automobile Demand* (based on a joint meeting of the American Statistical Association and the Econometric Society in Detroit, Dec. 27, 1938) New York: General Motors Corp.

[6] Griliches, Zvi 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change", *The Price Statistics of the Federal Government*, General Series, no. 73, New York: National Bureau of Econometric Research.

[7] Griliches, Zvi 1965. "Hedonic Parice Indexes Revisted: some Notes on the State of the Art", University of Chicago, Center for Mathematical Studies in Business and Economics, Report no. 6802.

[8] Griliches, Zvi 1971. *Price Indexes and Quality Change*. Cambridge.

[9] Wyngarden, Herman 1927. *An Index of Local Real Estate Prices*. Ann Arbor: University of Michigan Press.

[10] Wenzlick, Roy 1952. "As I See the Fluctuations in the Selling Prices of Single-Family Residences", *Real Estate Analyst* 21:541-48.

[11] Bailey, M. J., Muth. R. F., Nourse, H. O. 1963. "A Regression Method for Real Estate Price Index Construction", *J. of Amer. Stat. Assoc.* 58:933-42.

[12] Theil, Henri 1971. *Principles of Econometrics*. New York: John Wiley.

[13] Searle, S. R. 1971. *Linear Models*. New York: John Wiley.

[14] Berman, A. and Plemmons, R. J. 1979. *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic Press.

[15] Webb, C. 1981. "The Expected Accuracy of a Price Index for Discontinuous Markets", preprint.