A Regression Method for Real Estate Price Index Construction

Author(s): Martin J. Bailey, Richard F. Muth and Hugh O. Nourse

Source: *Journal of the American Statistical Association*, Dec., 1963, Vol. 58, No. 304 (Dec., 1963), pp. 933-942

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2283324

A REGRESSION METHOD FOR REAL ESTATE PRICE INDEX CONSTRUCTION

MARTIN J. BAILEY
RICHARD F. MUTH
University of Chicago
AND
HUGH O. NOURSE
Washington University

Quality differences make estimation of price indexes for real properties difficult, but these can be largely avoided by basing an index on sales prices of the same property at different times. The problem of combining price relatives of repeat sales of properties to obtain a price index can be converted into a regression problem, and standard techniques of regression analysis can be used to estimate the index. This method of estimation is more efficient than others for combining price relatives in that it utilizes information about the price index for earlier periods contained in sales prices in later periods. Standard errors of the estimated index numbers can be readily computed using the regression method, and it permits certain effects on the value of real properties to be eliminated from the index.

1. INTRODUCTION

INDEX numbers of the prices of real properties are difficult to construct. The major problem appears to be the great variation in quality among properties. Thus, indexes based upon the average sales prices of all properties of some particular kind sold in a given period, as in Laurenti [2], are likely to be deficient in two respects. First, variation in the quality of properties sold from period to period will cause the index to vary more widely than the value of any given property. Second, if there is a progressive change in the quality of properties sold at different times, the index will biased over time.

One way to avoid these difficulties is to eliminate quality differences using regression analysis. The regression approach has been applied to automobiles by Griliches [1] and to single-family houses by Pendleton [4],[1] and Bailey is currently using this approach in a study of prices of single-family houses in the Hyde Park area of Chicago. In this method, one introduces variables measuring important qualitative characteristics along with period effects into a regression analysis of sales prices. Dummy variables [5] are used to describe attributes such as brick construction or corner lot. Dummy variables are also used for period effects so that the data determine the specific functional form of the time index. Coefficients of the period effects yield a multiplicative price index if sales prices are expressed in logarithms. Such a method yields information on the influence of variation in quality characteristics on sales prices as well as a real estate price index. It also avoids the problem of selecting items of the same quality for comparison at different times. But, where quality characteristics are numerous and difficult to measure, as with multi-family residential and nonresidential properties and with all properties in the older parts of cities, the method may not yield useful results.

[1] A referee of this paper has informed us that the Statistical Division of the Department of Commerce is using this method to evaluate a Laspeyre index for the price of housing.

Most of the difficulties of specifying and measuring the numerous quality characteristics of real properties can be avoided by basing a price index on sales prices of the same properties at different times. The principal problem in so doing is the fact any given property is sold only at infrequent intervals. Somehow, a means must be found to combine price relatives based upon repeat sales of given properties into a single index. One way of doing so to obtain a multiplicative chain index, which has been employed with various modifications by Wenzlick [7] and Wyngarden [8], is as follows. First, the price relatives of all properties initially sold in period zero or the base period and resold in the first period are computed, and their geometric mean is taken as the index for the first period. The price relatives for all pairs of transactions for which the first period is that of initial sale are then adjusted or multiplied by the index for the first period. The price index for period two is the geometric mean of all adjusted price relatives with final sale in period two, where the adjustment factor for those price relatives with initial sale in period zero is unity (see equation (7), section 3). The process is then repeated, adjusting all price relatives by multiplying by the index for the period of initial sale and taking the geometric mean of all adjusted price relatives with final sale in a particular period as the index for that period.

The chain method just discussed is computationally simple, though perhaps tedious. But it is inefficient in that it neglects information about the index for earlier periods contained in price relatives with final sales in later periods. Furthermore, computation of the standard errors of the estimated index numbers can be quite difficult, and it is hard to estimate the effects of other changes affecting values of a particular property simultaneously with the index number. In this paper we shall discuss a regression method for combining price relatives based upon repeat sales prices of given properties which avoids the above noted difficulties.

2. REGRESSION ANALYSIS OF REPEAT SALES PRICES

The model upon which our regression method is based is as follows. Let:

$$R_{itt'} = \frac{B_{t'}}{B_t} \times U_{itt'}, \quad \text{or}$$
$$r_{itt'} = -b_t + b_{t'} + u_{itt'}, \tag{1}$$

where lower case letters stand for the logarithms of the corresponding capital letters. $R_{itt'}$ is the ratio of the final sales price in period t' to initial sales price in period t for the i-th pair of transactions with initial and final sales in these two periods. B_t and $B_{t'}$ are the true but unknown indexes for period t and t', respectively, where $t=0, 1, \cdots, T-1$, and $t'=1, \cdots, T$. While other assumptions are possible, we shall assume for now that the residuals in log form, $u_{itt'}$, have zero means, the same variances σ^2, and are uncorrelated with each other. In section 4 we shall explore the problem of correlated residuals.

Estimation of the unknown B's may be treated as a regression problem. Let x_t take the value -1 if period t is the period of initial sale, $+1$ if the period of final sale, and 0 otherwise for each pair of transactions. To normalize the index, let $B_0=1$ or $b_0=0$.[2] Using the above conventions, equation (1) becomes:

$$r_{itt'} = \sum_{j=1}^{T} b_j x_j + u_{itt'},$$

or, in matrix notation:

$$r = xb + u. \tag{2}$$

In (2) r and u are n-dimensional column vectors, where $n = \sum_{t,t'} n_{tt'}$, $n_{tt'}$ being the number of pairs of transactions with initial sale in period t and final sale in t'; b is a T-dimensional column vector of unknown logarithms of the index numbers to be estimated; and x is an $n \times T$ matrix. For a pair of transactions whose initial period of sale is other than the base period, i.e., $t=1, \cdots, T-1$, the corresponding row of x has a -1 in the t-th column; for any pair the corresponding row has a $+1$ in the t'-th column; all other elements of x are zeros.

Given the assumptions made above about u, the least-squares estimator:

$$\hat{b} = (x'x)^{-1}(x'r), \tag{3}$$

is the minimum variance linear unbiased estimator of b. The t-th diagonal element of $(x'x)$, which is a $T \times T$ matrix, is simply the number of pairs of transactions with initial sale in period t plus the number of pairs with final sale in period t. The t, t'-th off-diagonal element of $(x'x)$ is $-n_{tt'}$. Finally, $(x'r)$ is a T-dimensional vector whose t-th element is the sum of all price relatives for which period t is the period of final sale less the sum of all those for which t is the period of initial sale.

One advantage of the regression method is that it can be easily modified to eliminate the effects on value of certain changes in a property between the periods of initial and final sale. Examples of such changes are remodeling of or addition to a structure, a change in the number of dwelling units in an apartment building, a change in the race of the residents of a building or a neighborhood, and sale for demolition and redevelopment of the property to a new use. Using p appropriate variables to describe changes in a property which affect its value, the regression equation to be estimated becomes

$$R_{itt'} = \frac{B_{t'}}{B_t} X_{T+1}^{b_{T+1}} \cdots X_{T+p}^{b_{T+p}} U_{itt'},$$

or, in log form,

$$r_{itt'} = \sum_{j=1}^{T+p} b_j x_j + u_{itt'}. \tag{4}$$

The element in the t-th row and t'-th column of $(x'x)$, $t=1, \cdots, T$, $t'=T+1, \cdots, T+p$, is simply the sum of all values of $x_{t'}$ for which the period t is the period of final sale less the sum of all $x_{t'}$ for which t is the period of initial sale. For t, $t'=T+1, \cdots, T+p$, of course, the element of $(x'x)$ is the sum of the

[2] Alternatively, we could have expressed the model in terms of period-to-period changes in the index number. To do so, let c_j, $j=1, \cdots, T$, be the log of the relative change in the index from period $j-1$ period to j and x_j equal $+1$ if the period $j-1$ to j is included in the interval between initial and final sale and 0 otherwise. It can be shown that the two forms of the model yield identical estimates of the index number for any period. The ratio of the index estimated by the method discussed in the text for any pair of periods is thus independent of the base period selected. This follows because the ratio of the estimated indexes for period t' to period t is merely the product of the estimated period-to-period changes from t to t'. We prefer the form of the model discussed in the text, however, because it is computationally simpler.

cross products over all observations, $\sum_n x_t x_{t'}$, while the last p elements of $(x'x)$ are of the form $\sum_n x_t r$.

For some purposes one might want to adjust the price index for depreciation. Unfortunately, a depreciation adjustment cannot be readily estimated along with the price index using our regression method. Assuming that properties depreciate at a constant rate per unit time, our model would become

$$R_{itt'} = \frac{B_{t'}}{B_t} e^{-c(t'-t)} U_{itt'}, \quad \text{or}$$

$$r_{itt'} = \sum_{j=1}^{T+1} b_j x_j + u_{itt'}, \tag{5}$$

where $x_{T+1} = (t'-t)$, the number of periods from initial to final sale, and $b_{T+1} = -c$, the negative of the rate of depreciation. But the vector of values of x_{T+1} can be obtained by multiplying the matrix x in (2) by the column vector whose transpose is $(1, 2, \cdots, T)$. Hence, the x matrix in (5) is singular. In applying our method, therefore, additional information would be needed in order to adjust the price index for depreciation.

The method described above has another characteristic which some may feel is a disadvantage. If the regression method were to be used to construct an index on a continuing basis, each time another period would be added to the index the whole regression would have to be re-estimated. Doing so would make most efficient use of the available data, since sales in any given period add to the information available about the index in all earlier periods. The additional computations would not be much of a burden if a computer could be used. But it might prove irksome continually to revise the index for earlier periods, particularly if the index is published and widely used.

Continual revision of the index could be avoided, however, by recomputing the regression in order to obtain the best estimate of the index for the latest period only. This period is the one of greatest practical importance in most instances. At the same time, the recomputed regression would indicate when the additional information made available by sales in the latest period changes the estimated values of the index for earlier periods by amounts of sufficient practical importance to warrant revision of previously published estimates. The whole published series might then be revised infrequently if desired. And, if the computational burden of recomputing the regression each period is excessive, it too can be avoided. After obtaining the index for an initial interval of time using the regression method, the index could be estimated for succeeding periods using a variant of the chain method described earlier. Price relatives whose final sales occur in the periods following the initial computation via the regression method could be adjusted to the base period by multiplying them by previously estimated values of the index for their period of initial sale. The geometric mean of these adjusted price relatives could then be taken as the estimate of the index for that period. As our discussion in sections 3 and 5 below suggests, the gain in efficiency of the regression method over the chain method tends to be greatest in the earlier periods of the index. The modification of the chain method described here would seem largely to overcome its greatest weakness, the scarcity of information upon which to base the estimated index for earlier periods.

3. ILLUSTRATION FOR THE THREE-PERIOD CASE

To see how the regression method works and to compare it with the method for chaining together price relatives discussed in the introduction, consider estimation of a price index for three periods. Under the assumptions made in (1), $\bar{r}_{01}$, the mean of the price relatives for all pairs with initial transaction in period 0 and final transaction in period 1, is an unbiased estimator of b_1. Another unbiased estimator, which is uncorrelated with $\bar{r}_{01}$, is $(\bar{r}_{02}-\bar{r}_{12})$. Further, given our assumptions about u:

$$\operatorname{var}(\bar{r}_{01}) = \frac{\sigma^2}{n_{01}} \quad \text{and} \quad \operatorname{var}(\bar{r}_{02} - \bar{r}_{12}) = \left(\frac{1}{n_{02}} + \frac{1}{n_{12}}\right)\sigma^2.$$

As is well known, the weighted average of two or more uncorrelated estimators with minimum variance is that in which the separate estimators are weighted in inverse proportion to their variances. Thus, the minimum variance weighted average of $\bar{r}_{01}$ and $(\bar{r}_{02}-\bar{r}_{12})$ is

$$\frac{n_{01}\bar{r}_{01} + \left(\dfrac{n_{02}n_{12}}{n_{02}+n_{12}}\right)(\bar{r}_{02}-\bar{r}_{12})}{n_{01} + \left(\dfrac{n_{02}n_{12}}{n_{02}+n_{12}}\right)} = \frac{n_{01}(n_{02}+n_{12})\bar{r}_{01} + n_{02}n_{12}(\bar{r}_{02}-\bar{r}_{12})}{n_{01}(n_{02}+n_{12}) + n_{02}n_{12}}\cdot$$

Now, consider (3), In the three-period case

$$(x'x) = \begin{Bmatrix} (n_{01}+n_{12}) & -n_{12} \\ -n_{12} & (n_{02}+n_{12}) \end{Bmatrix}$$

$$(x'x)^{-1} = \left[n_{01}(n_{02}+n_{12}) + n_{02}n_{12}\right]^{-1}\begin{Bmatrix} (n_{02}+n_{12}) & n_{12} \\ n_{12} & (n_{01}+n_{12}) \end{Bmatrix},$$

$$(x'r) = \begin{Bmatrix} n_{01}\bar{r}_{01} - n_{12}\bar{r}_{12} \\ n_{02}\bar{r}_{02} + n_{12}\bar{r}_{12} \end{Bmatrix}.$$

Substituting in (3) yields

$$\begin{aligned} \hat{b}_1 &= \frac{(n_{02}+n_{12})(n_{01}\bar{r}_{01} - n_{12}\bar{r}_{12}) + n_{12}(n_{02}\bar{r}_{02} + n_{12}\bar{r}_{12})}{n_{01}(n_{02}+n_{12}) + n_{02}n_{12}} \\ &= \frac{n_{01}(n_{02}+n_{12})\bar{r}_{01} + n_{02}n_{12}(\bar{r}_{02}-\bar{r}_{12})}{n_{01}(n_{02}+n_{12}) + n_{02}n_{12}}, \end{aligned} \tag{6a}$$

which is identical with minimum variance weighted average of $\bar{r}_{01}$ and $(\bar{r}_{02}-\bar{r}_{12})$. Similarly for the second period

$$\hat{b}_2 = \frac{n_{02}(n_{01}+n_{12})\bar{r}_{02} + n_{01}n_{12}(\bar{r}_{01}+\bar{r}_{12})}{n_{01}(n_{02}+n_{12}) + n_{02}n_{12}}\cdot \tag{6b}$$

In (6b) $\bar{r}_{02}$ and $(\bar{r}_{01}+\bar{r}_{12})$ are, on our assumptions, uncorrelated unbiased estimators of b_2.

We now turn to the chain index discussed in the introduction; the estimators of the logs of the unknown indexes described there are

$$\bar{b}_1 = \bar{r}_{01} \tag{7a}$$

$$\bar{b}_2 = \frac{n_{02}\bar{r}_{02} + n_{12}(\bar{r}_{01} + \bar{r}_{12})}{n_{02} + n_{12}}. \tag{7b}$$

Here the estimator $\bar{b}_1$ is seen to ignore completely the information about b_1 provided by the price relatives for which the final transaction took place in period two, namely $(\bar{r}_{02} - \bar{r}_{12})$. The estimator $\bar{b}_2$ takes into account both $\bar{r}_{02}$ and $(\bar{r}_{01} + \bar{r}_{12})$ but weights them inefficiently—in proportion to the number of final sales in period two rather than in inverse proportion to their variances.

For the three-period case it is relatively easy to compute the variance of $\bar{b}$ and hence its relative efficiency. On the assumptions we have made about u

$$\text{var}\,(\bar{b}_1) = \frac{\sigma^2}{n_{01}} \quad \text{and} \quad \text{var}(\bar{b}_2) = \left[\frac{n_{01}(n_{02} + n_{12}) + n_{12}^2}{n_{01}(n_{02} + n_{12})^2}\right]\sigma^2.$$

Since $\text{var}(\hat{b}) = \sigma^2(x'x)^{-1}$:

$$\text{var}(\hat{b}_1) = \frac{(n_{02} + n_{12})\sigma^2}{n_{01}(n_{02} + n_{12}) + n_{02}n_{12}},$$

and similarly for $\hat{b}_2$. Thus the relative efficiencies of $\bar{b}_1$ and $\bar{b}_2$ are:

$$\begin{aligned} \frac{V(\hat{b}_1)}{V(\bar{b}_1)} &= \frac{n_{01}(n_{02} + n_{12})}{n_{01}(n_{02} + n_{12}) + n_{02}n_{12}}, \quad \text{and} \\ \frac{V(\hat{b}_2)}{V(\bar{b}_2)} &= \frac{n_{01}(n_{01} + n_{12})(n_{02} + n_{12})^2}{n_{01}(n_{01} + n_{12})(n_{02} + n_{12})^2 + n_{02}n_{12}^3}. \end{aligned} \tag{8}$$

For $n_{01} = n_{12} = m$, $n_{02} = 2m$, not implausible relative magnitudes, the relative efficiency of $\bar{b}_1$ is 0.6 and that of $\bar{b}_2$ is 0.9; however, as n_{01} goes to zero both relative efficiencies go to zero. This example suggests that the advantage of the regression method is greater for the earlier periods of the index and where there are relatively few sales in the earlier periods. Also, as either n_{02} or n_{12} goes to zero the estimators $\bar{b}$ and $\hat{b}$ become the same and the relative efficiency of $\bar{b}$ goes to one. These results suggest a conjecture which we do not attempt to prove for the general case, namely that the advantage of the regression method for a particular period is greatest where the excess of initial over final transactions in all preceding periods is large and where the number of initial transactions in this particular period is large.

4. MULTIPLE REPEAT SALES AND CORRELATED ERRORS

In many cases there may be data on more than two sales of a given property for the time period covered by the index. If so, there is no unique way to reduce these sales to price relatives, and any way of computing price relatives is likely to run into the problem of correlated residuals.

To see why this is so, let the sales price of the i-th property in the t-th period,

P_{it}, be the product of a property effect, A_i, a period effect, B_t, and a residual effect, V_{it}:

$$\begin{aligned} P_{it} &= A_i B_t V_{it}, \quad \text{or} \\ p_{it} &= a_i + b_t + v_{it} \end{aligned} \tag{9}$$

in logs. We assume that the v's have zero mean, constant variance, and are uncorrelated with each other. Here, the v's represent, for example, the effect on the sales price due to the peculiarities of a particular combination of buyer and seller. If, now, data on sales prices exist for periods t, t', and t'', there are three price relatives which could be computed, namely:

$$\begin{aligned} r_{itt'} &= p_{it'} - p_{it} = -b_t + b_{t'} + (v_{it'} - v_{it}) \\ r_{itt''} &= -b_t + b_{t''} + (v_{it''} - v_{it}) \\ r_{it't''} &= -b_{t'} + b_{t''} + (v_{it''} - v_{it'}). \end{aligned} \tag{10}$$

Under our assumptions about the v's,

$$\begin{aligned} \mathrm{Var}(r_{itt'}) &= \mathrm{Var}(r_{itt''}) = \mathrm{Var}(r_{it't''}) = 2\,\mathrm{Var}(v_{it}) \\ \mathrm{Cov}(r_{itt'}, r_{itt''}) &= \mathrm{Cov}(r_{itt''}, r_{it't''}) = \mathrm{Var}(v_{it}) \\ \mathrm{Cov}(r_{itt'}, r_{it't''}) &= -\mathrm{Var}(v_{it}). \end{aligned} \tag{11}$$

One way out of this problem would be to work directly with sales prices and estimate the property effect for each property along with the period effects or index numbers. To do so would not be computationally feasible, however, with existing computer regression routines.

A more feasible alternative would be to use weighted regression methods, since under the assumptions made above the variance-covariance matrix, M, of the residuals is known up to a scalar. As is well known, the minimum variance linear unbiased estimator is:

$$\hat{b} = (x'M^{-1}x)^{-1}(x'M^{-1}r). \tag{12}$$

Here, M has $+1$'s down the diagonal, and, if $r_{itt'}$ and $r_{it't''}$ are included, $-\frac{1}{2}$ in the row and column for any pair of relatives involving the same sales price in period t', and 0's elsewhere. In practice the computations might be simplified by dividing the price relatives into two groups, the second containing all n_2 price relatives containing a common sales price. The matrices x, M, and r then become:

$$x = \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}, \qquad M = \begin{Bmatrix} I & 0 \\ 0 & M^* \end{Bmatrix}, \qquad r = \begin{Bmatrix} r_1 \\ r_2 \end{Bmatrix},$$

so (12) can be simplified to:

$$\hat{b} = (x_1'x_1 + x_2'M^{*-1}x_2)^{-1}(x_1'r_1 + x_2'M^{*-1}r_2). \tag{13}$$

In (13), $(x_1'x_1)$ and $(x_1'r_1)$ can be evaluated directly as discussed above. Provided that there are few properties with more than two sales prices for the period covered, it should not be difficult to obtain $(x_2'M^{*-1}x_2)$ and $(x_2'M^{*-1}r_2)$ by direct matrix multiplication.

It might be argued, however, that in addition to the v's in (10), the u's in (1) contain another component, w, which represents the deviation of the relative change in value of a particular property between the periods of initial and final sale from the average of all properties covered by the index. Furthermore, if successive w's for a particular property are uncorrelated with each other, then one clearly should include $r_{itt'}$ and $r_{it't''}$ rather than either of the other two pairs in (10) in calculating the index to avoid correlation of residuals. Without knowing the relative variances of the v's and w's, the method of weighted regression discussed above cannot be applied. Using (3), however, would probably be reasonably efficient if there are few properties with two or more sales prices and if the variance of the v's is small relative to the variances of the w's. In any case, the estimator (3) is unbiased if the u's have zero mean, regardless of their intercorrelation. And (3) is no doubt superior to the chain method because it uses more of the information provided by the data.

5. A NUMERICAL COMPARISON OF THE CHAIN AND REGRESSION METHODS

The following example, taken from a study by Nourse [3], illustrates the advantages of the regression method over the chain method of real estate price

COMPARISON OF REAL ESTATE PRICE INDEXES FOR A SMALL AREA IN ST. LOUIS ESTIMATED BY THE CHAIN AND REGRESSION METHODS

Year	Chain Link Method	Regression Method			Number of Initial and Final Sales in Each Year
	Index 1937 = 1.00	Index 1937 = 1.00	Logarithm of Index	Standard Error in Logarithms	
1937	1.00	1.00	0	—	52
1938	.94	.99	−.0048	.0546	35
1939	2.50	.96	−.0188	.0490	60
1940	1.35	1.05	.0206	.0468	65
1941	2.34	1.07	.0312	.0464	71
1942	3.37	1.41	.1498	.0463	77
1943	2.46	1.52	.1803	.0471	59
1944	2.59	1.51	.1793	.0428	103
1945	2.98	1.75	.2423	.0424	109
1946	3.55	2.10	.3232	.0410	128
1947	4.38	2.34	.3687	.0432	105
1948	4.79	2.59	.4137	.0438	90
1949	4.53	2.21	.3438	.0466	67
1950	5.57	2.79	.4458	.0470	64
1951	5.62	2.90	.4624	.0464	70
1952	5.51	2.99	.4757	.0473	64
1953	5.72	2.96	.4718	.0507	50
1954	4.82	2.90	.4620	.0540	39
1955	5.51	3.09	.4901	.0499	47
1956	6.34	3.38	.5288	.0580	31
1957	5.74	3.06	.4863	.0528	44
1958	5.77	3.13	.4955	.0551	38
1959	4.85	2.62	.4184	.0523	44
					1512 Total

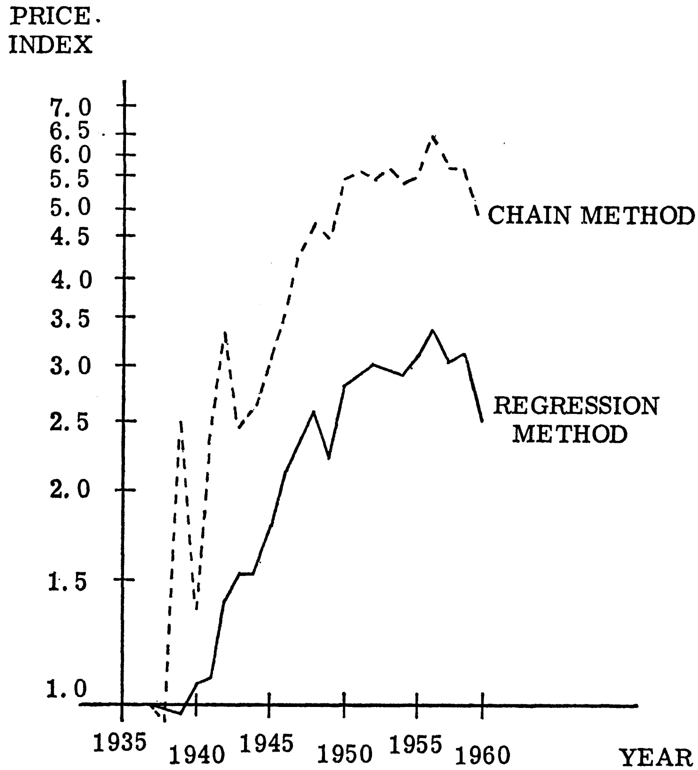

CHART 1. Comparison of real estate price indexes for a small area in St. Louis estimated by the chain and regression methods.

index construction. Indeed, it was the erratic behavior of the index estimated by the chain method that first prompted us to devise a better method.

In this study it was desired to construct an index of the prices of real properties for a small area of the near northwest part of St. Louis. The price data were obtained from the amount of tax stamps affixed to warranty deeds.[3] Values estimated from tax stamp values were then adjusted on the basis of other information contained in the warranty deed. For any transaction in which only a partial interest was conveyed, the value indicated by tax stamps was divided by the fraction of the interest conveyed. Where an outstanding mortgage was assumed by the buyer and its amount was stated in the deed, this amount was added to the equity conveyed by the deed; otherwise, this transaction was not included in the study. A transaction was not included if more than one property was conveyed by the deed, unless all of the properties conveyed were in the area studied and had been transferred together at least twice during the period covered by the study.

[3] Federal tax stamps must be affixed to the warranty deed in an amount not less than $.55 per $500.00 of value conveyed. While the law does not prohibit affixing more stamps, there would seem to be no incentive to pay a higher tax than necessary. But since tax stamps are available only in multiples of $.55, only a $500.00 range of values can be determined from the tax stamp amount. In this study the sales price was taken to be the upper limit of this $500.00 range, since studies cited by Tonty, *et al.*, [6] indicate that the actual amount is generally greater than the mid-point of the range and frequently at the upper limit. These latter studies also suggest that the difference between the actual consideration and that estimated from the tax stamp amount is less than 5 per cent.

Next, price relatives and their logs were computed from these adjusted tax stamp values. To obtain the regression estimate, equation (3), of the price index, the number and sum of the logs of price relatives with initial and final sales in a particular pair of years were entered above the diagonals of two 23 by 23 matrices. The elements of $(x'x)$ and $(x'r)$ were then evaluated using the rules described in section 2, above: The t-th diagonal element of $(x'x)$ was found by adding elements of the t-th column of the number of price relatives matrix down to the diagonal and then across the t-th row; the t, t'-th off diagonal element of $(x'x)$ is, of course, the negative of the corresponding element of the number of price relatives matrix. The t-th element of the 23-dimensional vector $(x'r)$ was found by adding all elements of the sums of logs of price relatives matrix down the t-th column to the diagonal and then subtracting those elements across the t-th row. The information so obtained was then inserted into a computer regression routine prior to the matrix inversion phase. The remainder of the computations were performed by the computer, and its output included the logs of the estimated index and their standard errors.

Price indexes estimated from these data by the chain and regression methods are shown in Table 1 and plotted in Chart 1. The two indexes diverge greatly in the years prior to 1943. In fact, most of the difference in the levels of the two is accounted for by the change in the chain index from 1938 to 1939. In addition, the year-to-year changes in the index estimated by the chain method are very erratic from 1938 to 1943, much more so than the regression estimate of the index. For 1944 and following years, however, the indexes estimated by the two methods are very similar apart from the higher level of the chain estimate. We suspect that the major reason for the difference between the two estimates is the fact that there were relatively few final sales upon which to base the chain estimate for the earlier years.

REFERENCES

[1] Griliches, Z., "Hedonic price indexes for automobiles: An econometric analysis of quality change," U. S. Congress, Joint Economic Committee, *Government Price Statistics*. Hearings, January 24, 1961. Washington: Government Printing Office, 1961. Pp. 173–96,

[2] Laurenti, Luigi, *Property Values and Race*. Berkeley: University of California Press, 1960.

[3] Nourse, Hugh O., "The Effect of Public Housing on Property Values in St. Louis." Unpublished Ph.D. dissertation, Department of Economics, University of Chicago, 1962.

[4] Pendleton, William C., "The Value of Accessibility." Unpublished Ph.D. dissertation, Department of Economics, University of Chicago, 1963.

[5] Suits, Daniel B., "Use of dummy variables in regression equations," *Journal of the American Statistical Association*, 52 (1957), 548–51.

[6] Tonty, Robert L., *et al.*, "Reliability of deed samples as indicators of land market activity," *Land Economics*, XXX (1954), 47–8.

[7] Wenzlick, Roy, "As I see the fluctuations in the selling prices of single-family residences," *The Real Estate Analyst*, XXI (December 24, 1952), 541–8.

[8] Wyngarden, Herman, *An Index of Local Real Estate Prices*. Michigan Business Studies, Vol. 1, No. 2. Ann Arbor: University of Michigan, 1927.