Taylor & Francis
Taylor & Francis Group

Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods

Author(s): John M. Clapp and Carmelo Giaccotto
Source: *Journal of the American Statistical Association*, Vol. 87, No. 418 (Jun., 1992), pp. 300-306
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2290260
Accessed: 17-10-2017 05:19 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

JSTOR

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods

JOHN M. CLAPP and CARMELO GIACCOTTO*

Accurate estimation of price indices for residential property is an essential feature of real estate research, especially in view of recent efforts to forecast price trends for the 1990s. In this article, price trends are estimated by using the sales price, assessed value and date of sale for every residential property transaction between independent parties. This assessed value (AV) methodology is compared to the repeat sales (RS) method. This article develops a simple method for correcting the effect of the measurement errors associated with assessed value. We demonstrate that the large samples available with the AV method allow the measurement error problem to be reduced to negligible proportions. Using data on the Hartford, Connecticut metropolitan area, we find that price trends estimated from the AV and RS methods are substantially similar over a seven-year period. But the RS method is inefficient because it uses a relatively small subset of the data. Our results indicate that it remains inefficient when the researcher has a dataset much richer in repeat sales than ours.

KEY WORDS: Assessed value; Measurement errors; Real estate price indices; Repeat sales; Residential property.

Accurate price series on a large number of assets, such as stocks and bonds, is an essential feature of financial market research. Analogous information for local real estate markets (e.g., relatively small cities and towns) would be useful not only to researchers but also to town officials and homeowners. However, real estate price indices reported in the literature are almost always for single-family properties at the national, regional, and metropolitan levels of aggregation; only relatively few (and usually large) metropolitan areas are covered (Bailey, Muth, and Nourse 1963; Bryan and Colwell 1982; Pollakowski 1987).

The need for accurate estimates of residential price changes has been emphasized by the recent debate over expectations for U.S. housing prices in the 1990s (Mankiw and Weil 1989; Hendershott and Peach 1990). For most persons, residential real estate is the largest component of their wealth. Changes in housing prices have a significant impact on individual investment decisions, which in turn may influence growth in the macroeconomy (Case 1990).

This article is motivated by the fact that none of the methodologies proposed in the literature (for example, Case and Shiller 1987) has been put into practice in the sense of regularly published price indices analogous to the published information on the stock market. This article develops a methodology designed to provide such information on constant quality price indices by property type. We apply this methodology to several towns in Hartford, Connecticut.

Two methods for price index estimation currently are established in the literature: hedonic regressions and repeat sales (RS) analysis. Hedonic regressions use data on a vector of characteristics for each property sold to control for quality. Because this approach requires a large amount of data over time, it is doubtful that it can be extended beyond single-family properties in metropolitan areas.

The RS method originally proposed by Bailey, Muth, and Nourse (1963) controls quality by using prices at different points in time for the same property; properties that have changed substantially due to additions or modifications are (or should be) screened out of the sample. There are two problems with the RS method, however. First, data on resales of the same property may not be readily available; second, it may be difficult to eliminate properties that have changed substantially between the two sales.

The average percentage of residential transactions where a single property is involved in repeat sales over a given time period is small; for example, less than 5% over a 15-year period for the data available to Case and Shiller (1987). Thus small sample sizes are to be expected if any degree of spatial detail is desired. Further, there may be selection bias in the repeat subsample; for example, "lemons" may trade more frequently than other properties. This article presents evidence on sample selection bias.

The purpose of this article is to evaluate and develop a new price index method and to compare the database and price indices derived using this method to those derived using the RS method. The proposed method uses the assessor's opinion of value to control for heterogeneous property and neighborhood characteristics. We use errors-in-variables analysis to show that errors in the assessor's opinion of value (Berry and Bednartz 1975; Kennedy 1984) do not bias the estimated price indices. The next section develops the assessed value (AV) method, shows its relationship to the repeat sales method, and develops a model for combining repeat information with information on properties that sold only once. The following sections outline the data and results for Hartford and for cities studied by Case and Shiller (1987): Atlanta, Dallas, Oakland, and Chicago.

* John M. Clapp is Professor of Finance and Real Estate and Carmelo Giaccotto is Associate Professor of Finance, University of Connecticut, Storrs, CT 06269-2041. Valuable comments on earlier drafts of this article were received from Robert Shiller, Yale University, Susan Woodward, U.S. Department of Housing and Urban Development, an associate editor, and two referees. The authors are grateful to the Center for Real Estate and Urban Economic Studies, University of Connecticut, for financial support, research assistance, and secretarial help.

© 1992 American Statistical Association
Journal of the American Statistical Association
June 1992, Vol. 87, No. 418, Applications and Case Studies

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

1. METHODOLOGY

1.1 The Hedonic and Assessed Value Methods

An intuitively appealing method for estimating a cumulative price index, called the *hedonic regression method* after Bryan and Colwell (1982) and Pollakowski (1987), uses the following cross-sectional regression:

$$\ln P_{it} = \mathbf{b}'_1 \ln \mathbf{S}_{it} + \mathbf{b}'_2 \ln \mathbf{L}_{it} + c_1 Q1_i + c_2 Q2_i + \cdots + c_T QT_i + e_{it}, \quad (1)$$

where, P_{it} = transaction price i at time t, $i = 1, \ldots, n_t$, and $t = 1, \ldots, T$; $\mathbf{S}_{it}(\mathbf{L}_{it})$ = a vector of structure (locational) characteristics; $\mathbf{b}'_j$ = row vectors of coefficients on structural ($j = 1$) and locational ($j = 2$) characteristics; Qt_i = a time dummy with values of 1 if the ith house sold in period t and 0 otherwise; and e_{it} = random error with 0 mean and variance σ_e^2. The regression coefficients $c_1, c_2, \ldots, c_T$ represent the logarithm of the cumulative price index. To show that $\exp(c_t)$ is indeed a cumulative index, note that at time 0 (i.e., when all time dummies equal 0) the price of each property is represented by:

$$\ln P_{i0} = \mathbf{b}'_1 \ln \mathbf{S}_{i0} + \mathbf{b}'_2 \ln \mathbf{L}_{i0} + e_{i0} = \ln V_{i0} + e_{i0}, \quad (2)$$

where $\ln V_{i0}$ is the natural logarithm of the true market value (expected value of the sales prices) of the ith property. For the typical house sold at time t, $Qt = 1$ (all other dummies = 0); therefore $P_t = V_0 e^{c_t}$. We define $c_0 = 0$, setting the base of the index at 1.

The data base required by the hedonic regression method is so large and thus so expensive to obtain that the method is applied infrequently and then only for a few large cities. Furthermore, data compiled by the Society of Real Estate Appraisers (SREA) generally are used to estimate Equation (1). Although these data capture the large majority of transactions, properties sold by their owner may be underrepresented. In addition, some of these SREA transactions are between relatives (i.e., family members) or involve a substantial amount of personal property.

The AV method relies on the fact that the tax assessor is required to assess real property at some percentage of market value. Tax assessors use appraisal methods, typically including regressions similar to Equation (1), to estimate market value. When a major property characteristic changes (e.g., change in square footage, addition of a room, or rezoning), then assessed value is adjusted to value the changed property in a manner consistent with the assessment of similar properties. The value of the changed property does *not* reflect price trends since the last revaluation of all property in the area. Thus we model the assessed value of property i at time zero (A_{i0}) as

$$\ln A_{i0} = (1/c) \ln V_{i0} + z_{i0}. \quad (3)$$

Here c is a parameter usually referred to as *vertical assessment equity* (Kochin and Parks 1982) and z_{i0} is a random disturbance term with 0 mean and variance σ_z^2. The constant c in Equation (3) allows for departures from assessment uniformity. For example, if $c > 1$, then assessed value is inelastic with respect to true value: Assessment practices are biased against low-valued properties and therefore against minorities and low-income households.

Systematic assessment error is discouraged and corrected by: (a) appeals by those over-assessed; (b) the use of professional revaluation firms, independent of local politics, for computer-assisted mass assessment (CAMA); and (c) state government evaluation of local assessment practices. Note that appeals only by those overassessed tend to shift the assessment burden to those underassessed, restoring equity. Also CAMA methods, drawing heavily on pricing models similar to Equation (1), have become increasingly sophisticated.

Assessment errors (undervaluation and overvaluation) are captured by z. There are several sources for the error term z, including the lag between times of the sales used by the assessor and the date of assessed value (October 1 of the revaluation year in Connecticut). Also, the assessor does not have complete information about the market participants and their preferences, which would determine the actual value of a property (i.e., expected sales price). Other sources of z might include racial prejudice and political favoritism (Berry and Bednarz 1963).

The relationship between Equation (1) and the assessed value approach can be clarified by comparing Equations (2) and (3). The assessed value summarizes into a single number the locational and structural characteristics of a real property. Thus A_{i0} acts as a proxy for the true market value (as of time o) of the ith property.

When transaction prices occur at various points of time, the price index can be estimated by applying appropriate statistical techniques, discussed below, to:

$$\ln P_{it} = c \ln A_{i0} + c_1 Q1 + c_2 Q2 + \cdots + c_T QT - c z_{i0} + e_{it}, \quad (4)$$

where c is the assessment equity parameter from Equation (3). The AV method, Equation (4), provides a highly flexible technique for estimating a price index for the aggregate, for broad categories of real estate such as single-family vs. condominiums, or for a city, a metropolitan area, or even a state. Clearly, the main goal of the AV method is price index estimation; however, the $\hat{c}_t$'s can be useful for further analysis (such as market efficiency studies) or for estimating property appreciation and depreciation.

The AV method has two major advantages. First, the simplification of the data base—A_{i0} replaces $\mathbf{L}_{i0}$ and $\mathbf{S}_{i0}$ in Equation (2); and second, the data represent *all* transactions between independent parties, not just those handled by certain real estate professionals. In over 40 states, the data required for the AV method are collected by state agencies as part of "sales ratio" studies designed to evaluate local assessment practices. Thus, the data on real property sales are generally "clean" and available on magnetic tape from state offices.

1.1.1 Analysis of Errors in Assessed Value. Random errors in assessed value (Equation 3) pose a problem for the AV method, because measurement error will bias not only the ordinary least squares (OLS) estimator of c (the coefficient on the log of assessed value, $\ln A_{i0}$) but also all esti-

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

mators of price trends $c_1, \ldots, c_T$ (Greene 1990). In the present situation, two characteristics facilitate analysis: (1) assessed value, A_{i0}, is the only variable measured with error; and (2) the time dummy variables are orthogonal to each other. Thus we are able to obtain a closed-form solution for the asymptotic bias of $\hat{c}_t$ and then empirically estimate the effect of errors in Equation (4).

It may be shown (see Clapp and Giaccotto 1991) that the bias in $\hat{c}$, the assessment equity parameter, is captured by

$$\text{plim } \hat{c} - c = -(\sigma_z^2/\sigma_a^2)c, \quad (5)$$

where σ_z^2 is the variance of the measurement error in assessed value and σ_a^2 is the variance in observed assessed values. Also for each time period $t = 1, \ldots, T$, the bias in the price index is

$$\text{plim } \hat{c}_t - c_t = (\sigma_z^2/\sigma_a^2)\bar{v}_t, \quad (6)$$

where

$$\bar{v}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} \ln V_{i0} \quad (7)$$

is the average of (log) true market value, as of time zero, for the sample of properties that sold during period t. In most situations the actual level of the price index is not as meaningful as the rate of growth representing the rate of real property inflation. The probability limit for this rate is

$$\text{plim}(\hat{c}_t - \hat{c}_{t-1}) = c_t - c_{t-1} + \left(\frac{\sigma_z^2}{\sigma_a^2}\right)(\bar{v}_t - \bar{v}_{t-1}). \quad (8)$$

Equation (8) suggests that there will be no bias from measurement error if the change in $\bar{v}_t$ is 0. But as n_t becomes large for all t, there can be no change in average true value. That is, the only source of change is sampling variability, which can be measured by examining $\bar{a}_t - \bar{a}_{t-1}$ for each t, where $\bar{a}_t = 1/n_t \sum_{i=1}^{n_t} \ln A_{i0}$.

A second important conclusion from Equation (8) is that any change in $\bar{v}_t$ must be multiplied by σ_z^2/σ_a^2. Although this "error ratio" is not directly observable (only σ_a^2 can be measured), reasonable estimates can be made. The mean squared error from fitting Equation (4) provides an estimate of $c^2\sigma_z^2 + \sigma_e^2$. The vertical equity parameter, c, is typically near unity for our data (Clapp 1990). Thus we may obtain an upper bound estimate of σ_z^2.

We conclude that Equation (8) allows us to: (a) determine if measurement error biases the $\Delta\hat{c}_t$'s; and (b) make reasonable corrections where errors are important. We will implement this strategy in Section 3.2. Note that high correlation between repeat sales and AV price trend estimates will strongly support our contention that both systematic and random errors in assessment are not important, at least for the Connecticut data.

1.2 The Repeat Sales Method

Bailey, Muth, and Nourse (1963) proposed an alternative method for estimating a constant quality real estate price index. This approach uses sales prices for the same property at different points in time, provided that property characteristics do not change between the two sales.

The paired sales approach can be derived from Equation (4)—and therefore from Equation (1)—by first eliminating from the sample all properties that sold only once during the sample period and second dividing the repeat subsample into two equal halves: the first NR observations for the first sale and the last NR observations for the second sale. That is, there are NR repeat pairs. Then Equation (4) can be partitioned as follows:

$$p_1 = ca_1 + c_1Q1_1 + \cdots + c_TQT_1 + e_1 - cz_1 \quad (9)$$

and

$$p_2 = ca_2 + c_1Q1_2 + \cdots + c_TQT_2 + e_2 - cz_2, \quad (10)$$

where all variables are NR dimensional column vectors defined in the previous section (lower-case p and a stand for $\ln P$ and $\ln A$). The subscripts 1 and 2 index the first and second sales. Note that this partition of Equation (4) holds only for the repeat subsample.

Because there are no changes in property characteristics between the two sales, there are no changes in assessed value other than technical adjustments for a general revaluation. Thus $a_1 = a_2$ and $z_1 = z_2$. The estimating equation for the RS method is derived by subtracting Equation (9) from Equation (10):

$$y = p_2 - p_1 = c_1(Q1_2 - Q1_1) + \cdots + c_T(QT_2 - QT_1) + e_2 - e_1. \quad (11)$$

The familiar form of the RS equation (e.g., used by Bailey, Muth, and Nourse [1963] and by Case and Shiller [1989]) can be obtained from Equation (11) by combining the time dummies for the first and second sales. A single dummy is formed for each period with a value of minus 1 for the first sale and plus 1 for the second sale. A typical approach is to assume that the error term in Equation (11) is iid after adjusting for heteroscedasticity (see Case and Shiller 1989).

There are two negative aspects of the RS method. First, information about properties that sold only once must be discarded from the sample, which reduces the sample size by as much as 97%! Second, the remaining observations may be biased by the fact that those properties selling repeatedly may not be a representative sample from the population of all properties sold. There are some plausible sources for sample selectivity in the RS method. First, properties are bought to be repaired or rehabilitated for resale; second, there may be a "lemons" phenomenon in which properties that do not meet buyer expectations are repeatedly resold; and third, "starter homes" sell repeatedly as owners move to better housing. We will test for sample selectivity by comparing sales prices and assessed values for all properties sold to those for properties sold twice only and those sold three or more times.

Aging and changes in quality of location between the first and second sale (Palmquist 1979) can be viewed as part of general economic and environmental change that should be reflected in the price index rather than held constant. In particular, any loss in value due to aging should be reflected in our estimate of the price appreciation experience by the typical homeowner. Similarly, stock market indices do not

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

control for aging of the capital stock. Therefore, we follow Case and Shiller (1990, pp. 8–9) in ignoring aging and changes in location quality when implementing Equation (11); similar reasoning leads us to ignore these variables when using the assessed value approach, Equation (4).

1.3 The Relationship Between Repeat and Assessed Value Methods

Inspection of Equations (9) and (10) reveals that one could apply the AV method to the repeat subsample, augmented by the assessed value of each property. Equations (9) and (10) would be stacked, whereas they are subtracted for the RS method, Equation (11). Thus we have

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} C + \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{12}$$

where $y_j = p_j - ca_j$, $x_j = Q1_j, Q2_j, \ldots, QT_j$, $C' = (c_1, c_2 \cdots c_T)$, the cumulative price index, $f_j = e_j - cz_{i0}$, and $j = 1, 2$ indices the first and second sales.

The AV method (Equation 12) and the RS method (Equation 11) use the information contained in the RS subsample somewhat differently. The RS method averages the geometric rate of growth in price for each property over the time period between the two sales (Shiller 1990). The AV method on the other hand averages the growth rate from 0 to the time of the first sale for the first NR observations and from 0 to the second sale for the last NR observations. The sample size advantage of the AV method over the RS method (2NR vs. NR) is mitigated to some degree by the fact that the error terms in Equation (12) likely will be positively correlated across the two groups. This problem may be remedied by treating Equation (12) as a system of seemingly unrelated regressions (SUR) of Zellner (1962). When this is done, we expect the results to be very similar to OLS estimation of Equation (11). If this turns out to be the case, then we have credible evidence that assessed values augmented with sale prices are good substitutes for repeat sales data.

1.4 Combining RS Data with the AV Method

In the previous section, the database was restricted to those properties that sold more than once, because the RS method cannot handle one-only transactions. But there is no reason to exclude these observations from the AV approach, provided the assessed value of each property is in the database. The AV method makes it possible to use data both on properties that sold only once and on repeat sales. To accomplish this goal, simply generalize the system of Equation (12):

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} C + \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \tag{13}$$

where all terms are defined for Equation (12) and $j = 3$ indices those properties that sold only once during the sample period. The total number of transactions in the sample is $N = 2\text{NR} + \text{NO}$, where NO is the number of one-only sales.

The variance-covariance matrix of error terms is

$$V = \sigma^2 \begin{pmatrix} I_{\text{NR}} & \rho I_{\text{NR}} & 0 \\ \rho I_{\text{NR}} & I_{\text{NR}} & 0 \\ 0 & 0 & I_{\text{NO}} \end{pmatrix}, \tag{14}$$

where I_J is a $J \times J$ identity matrix with J = NR or NO and ρ is the covariance between f_{1i} and f_{2i}.

Once ρ is estimated from Equation (12) and σ^2 is estimated from Equation (13), it is straightforward to estimate Equation (14) and its inverse, V^{-1}. Thus we estimate the correct variance-covariance of the OLS cumulative price index, $\hat{C}$:

$$C - \text{OLS} = (X'X)^{-1}X'VX(X'X)^{-1}, \tag{15}$$

where $X' = (x_1', x_2', x_3')'$. Likewise, the generalized least squares (GLS) variance-covariance matrix can be calculated as GLS $= [X'V^{-1}X]^{-1}$. We will use these computational formulas to investigate the efficiency of the estimators.

2. DATA SOURCES

Two sources of data are used. First, we have repeat sales (1970–1986) for single-family residential (SFR) properties provided by Case and Shiller (1987, 1989). Although four major metropolitan areas are covered, there are no data for assessed value or for properties that sold only once. Thus the Case-Shiller data are used to examine sample selectivity and to validate the accuracy with which price trends can be measured from the Hartford repeat data.

The Hartford data include all one- to three-family (hereinafter referred to as SFR) residential transactions over the period from October 1981 to September 1988. The data are for four large (more than 15,000 population) towns in the core of the metropolitan area (i.e., surrounding the central business district). None of these towns had a general change in assessed value over the 1981–1988 period. The data were collected, verified, and compiled by Connecticut's Office of Policy and Management using standards determined by the International Association of Assessing Officers (IAAO).

3. PRICE INDICES FOR FIVE CITIES

This section has three parts. In the first section, we evaluate sample selectivity by comparing average sales price and average assessed value (where available) for properties that sold only once, only twice, and three or more times. In the second section, we compare price trends for different samples and methods by evaluating the cumulative index (c_t) and the continuously compounded rate of change ($c_t - c_{t-1}$). In the third section, we evaluate the precision with which price trends can be measured by the AV and RS methods.

3.1 Sample Selectivity

In Hartford, the prices and assessed values of SFR properties that sell twice are about 15% less than that of properties that sell only once during the sample period (see Table 1). Similar results are obtained for those properties that sell three or more times. These relationships are statistically significant and they hold when the quarterly data are examined.

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

Table 1. Average Sale Prices in Five Metropolitan Areas (All Quarters Combined)

	Number of sales in sample period						
	1	2	% Difference	t Value	3+ Sales	% Difference 2–3+	t Value
Hartford	$131,962	$115,042	14.7%	8.66	$107,277	7.2%	2.65
Atlanta	—	66,300	—	—	52,405	11.0%	7.67
Chicago	—	54,727	—	—	52,094	4.8%	4.00
Dallas	—	64,080	—	—	60,487	5.6%	2.30
Oakland	—	140,132	—	—	67,422	48.1%	13.00

NOTE: Prices are not adjusted for inflation, but examination of each quarter individually confirms the above findings for the majority of quarters; see Clapp, Giaccotto, and Tirtiroglu (1990).

The data provided by Case and Shiller (1987, 1989) support the conclusion that properties selling more frequently typically command lower sales prices. However, there is enormous variability across metropolitan areas in the amount of the difference; for example, it is less than 5% in Chicago and almost 50% in Oakland. The quarterly data (not shown) support the same conclusions.

3.2 Comparison of Price Indices

Given our finding of sample selectivity, what is the relationship between a price index using all the data and the AV method versus one using the repeat subsample and the RS method? A complete answer to this question requires examination of differences between the samples as well as between the methods.

Figure 1 shows the cumulative price index ($\hat{c}_t$) for SFR properties in Hartford. Quarter 1 is the fourth quarter of 1981, when both indices have values of 1 (by construction), and Quarter 27 is the third quarter of 1988. The cumulative growth in the value of residential real estate is 90% or approximately $10\frac{1}{2}$% per year, a return comparable to that of the stock market for the same period.

Figure 1 shows that the two approaches (the RS method using the repeat subsample and the AV method using all the data) generally track each other well over long periods. There are some periods when gaps occur between the two price indices, but differences generally are transitory. One exception, however, is the period from Quarter 18 (the first quarter of 1986) to Quarter 24 (the third quarter of 1987).

We used White's covariance matrix to evaluate the statistical significance of these gaps. Because of the relatively small sample size for repeat sales (3,600 observations versus 13,600 for all the data), the gaps are not statistically significant. However, they may be *economically* significant. The gaps range from 0% to +2.8% or more, with a .5% difference not uncommon. Figure 1 suggests that the market is sometimes slow to eliminate differences between the repeat subsample and all transactions.

In addition to graphs of cumulative price indices, we used correlation coefficients between percent changes in price (i.e., first differences in $\hat{c}_t$) to evaluate differences among samples and methods. If one were only interested in the rough accuracy of percentage changes, then high correlation coefficients suggest that either method would be adequate. The difference between the samples (.8 coefficient for the AV method using all the data, AV-All and with the AV method using the repeat subsample, AV-Rpt) appears to be about as important as the difference between methods (the coefficient for AV-Rpt on Rpt-Rpt is .82). When compared to the .78 correlation between AV-All and Rpt-Rpt, we conclude that the difference between methods compounds slightly with the difference between samples. That is, the correlations indicate that the strongest similarity is between the RS method using the repeat data and the AV method using all the data.

We also experimented with the adjustment for errors in variables, Equation (8). The adjusted series is virtually identical to the unadjusted series. Further, the adjustment does not significantly improve the accuracy of the AV method: The correlation coefficient is .82 between Rpt-Rpt and AV-Rpt and .81 between Rpt-Rpt and AV (Adjusted)-Rpt. We conclude that errors in variables are not a serious problem for this database. This finding is consistent with our theory, Equation (8), which shows that measurement bias should go to 0 in large samples.

3.3 Precision of Price Index Estimates

The precision of the estimators of price changes are affected by two factors. First, sample size is generally much larger for the AV method. But this advantage may be offset by the second factor: the assumption, implicit in OLS estimation of Equation (4), of zero covariance between the errors associated with the first and second sales. The RS method, on the other hand, allows a nonzero covariance.

In Table 2 we compare OLS, ACOV (White's [1980] asymptotically consistent variance), and GLS estimates of the standard errors. In general, the researcher using the AV method will *not* have information on which properties (and sales) are involved in repeat transactions. Thus the OLS estimators will not be as efficient, and the standard errors will be estimated by an incorrect formula. We assess the magnitude of this problem by computing the correct OLS standard errors. We also compute some GLS standard errors to determine whether it is worthwhile to collect the additional information about repeat transactions.

Because of the large number of zeroes in X, correct OLS standard errors are not substantially different from those calculated from the OLS formula (see Table 2). Moreover, the standard errors calculated using the ACOV matrix were not substantially different from those calculated by OLS.

An interesting question is whether efficiency can be substantially improved by using the full sample *and* retaining

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

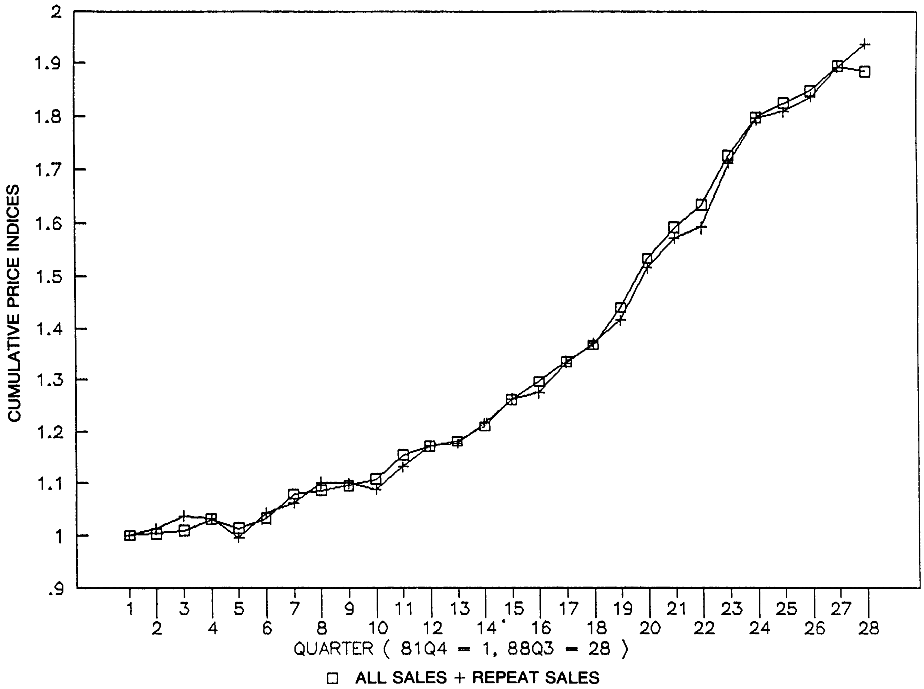

Figure 1. Cumulative Price Indices for Hartford, CT.

information on the properties that sold repeatedly. The GLS estimators for the full sample have standard errors that are approximately 50% of those from the RS method (more for early quarters, less for later quarters). But if one had simply used the AV method on the full sample, estimated by OLS, then the standard errors would have been between 55% and 75% of those obtained by using the RS method on the repeat subsample. An additional 10%–20% reduction can be achieved by using the full sample with information from the repeat subsample (GLS). However, obtaining this additional precision requires considerable data collection and manipulation.

The standard errors of the $\Delta\hat{c}_t$'s are essentially the same for the two methods (see the last line of Table 2). From this result and from Figure 1, we conclude that the AV method accomplishes the same purpose as pairing repeat sales. Further, even when repeats are 100% of the data, Table 2 shows that OLS estimation of Equation (4) gives standard errors similar to those from the RS method. Together with results reported in Clapp and Giaccotto (1991, table 4), this indicates that the RS method is inefficient even when the database is much richer in paired sales than ours (25% repeats) or the four datasets of Case and Shiller (averaging less than 5% repeats).

Table 2. Standard Errors for Assessed Value Method, Selected Quarters (SE's for $\Delta\hat{c}_t$, Equations 12 and 13)

	Quarter					
	3	4	15	16	27	28
All Transactions						
ACOV	.014	.013	.010	.009	.015	.012
OLS: $(X'X)^{-1}$	.016	.014	.011	.009	.013	.011
C-OLS: $(X'X)^{-1}X'VX(X'X)^{-1}$	.016	.014	.011	.009	.013	.011
GLS: $(X'V^{-1}X)^{-1}$	.013	.012	.009	.008	.011	.010
Repeat Data						
OLS: $(X'X)^{-1}$	.019	.018	.018	.015	.035	.027
C-OLS: $(X'X)^{-1}X'VX(X'X)^{-1}$	.022	.020	.020	.017	.030	.023
SUR(GLS): $(X'V^{-1}X)^{-1}$	.021	.020	.016	.016	.024	.019

NOTE: ACOV is White's (1980) asymptotically consistent covariance matrix. C-OLS is the corrected OLS variance-covariance matrix, Equations (14) and (15) with $\hat{\rho}$ = .755 and $\hat{\sigma}^2$ from the OLS regression; $\hat{\rho}$ was estimated from Zellner's (1962) seemingly unrelated regression (SUR). The SUR regression (last line of the table) gave SE's substantially equal to the repeat method, Equation (11). For GLS, full sample, V^{-1} is estimated with Equation (14).

4. SUMMARY AND CONCLUSIONS

We have developed a new AV method for estimating real estate price indices and compared it to the RS method. Also we have developed an analytical solution showing that the effect of errors in the measurement of assessed values on price indices is negligible when the quarterly change in average assessed value (i.e., change in the composition of the sample) is small. For the Hartford single-family residential (SFR) data, measurement error is not important. The estimated price indices shown in Figure 1 reveal the growth in real estate values during the 1980s. Future research based on the AV method developed here might reveal whether this growth reflected fundamental values or was merely a speculative bubble.

The Hartford data indicate that the AV and RS methods give substantially the same estimates of price trends over a five- to seven-year period. However, the AV method data base is relatively easy to obtain, and the data have been cleaned by state agencies.

The most important differences between the two methods are: (a) the subsample of repeat sales is not representative of all sales (sample selectivity), and (b) the efficiency of the two methods differs because of differences in sample size and

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms

because of positive covariance between the first and second sales.

Evidence for five cities indicates that the repeat subsamples are influenced by a "lemons" or starter home phenomenon. The repeat subsamples have lower average sales prices and lower assessed values than does the sample of properties that sold only once. Furthermore, properties that sold three or more times have lower values than those that sold only twice. This does not appear to affect the price index estimates over periods of several years or more, but it may affect those estimates over periods as long as six quarters. Thus the RS method coefficient estimates may apply only to the repeat subsample over short time periods.

Turning to the efficiency issue, there is a substantial positive covariance between the regression residuals for the first and second sales. OLS applied to the RS method takes this covariance into account, whereas the AV method ignores this covariance. Although this makes the RS method more efficient, the effect is not very large in the Hartford data. Even when the sample is composed entirely of repeat sales, the AV method is almost as efficient as the RS method. On the other hand, the larger sample sizes available to the AV method increase efficiency by 25%–45%.

We have developed a three-equation system of seemingly unrelated regression for combining information on repeat and one-only sales and have developed computational formulas for estimating this model. For the Hartford data, the additional efficiency gain is between 10% and 20%. That is, the three-equation system produces standard errors that average about 50% of those derived from the RS method.

[*Received January 1991. Revised October 1991.*]

REFERENCES

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963), "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association,* 58, 933–942.

Berry, B., and Bednarz, R. S. (1975), "A Hedonic Model of Prices and Assessments for Single Family Homes: Does the Assessor Follow the Market or the Market Follow the Assessor?" *Land Economics,* LI, 1, 21–40.

Blakeley, D. M., and Follain, J. R., with Lee, H. (1986), "An Evaluation of Hedonic Price Indices for Thirty-Four Large SMSAs," *AREUEA Journal,* 14, 179–205.

Bryan, T. B., and Colwell, P. F. (1982), "Housing Price Indices," *Research in Real Estate,* 2, 57–84.

Case, B., and Quigley, J. M. (1991), "Dynamics of Real Estate Prices," *Review of Economics and Statistics,* 73, 50–58.

Case, K. E., and Shiller, R. J. (1989), "The Efficiency of the Market for Single-Family Homes," *The American Economic Review,* 79(1), 125–137.

——— (1987), "Prices of Single-Family Homes Since 1970: New Indices for Four Cities," *New England Economic Review,* Sept./Oct., 45–56.

Case, K. E. (1990), "Real Estate Cycles and the Regional Economy: New England in the 1980s," draft paper presented at the North American Regional Service Meetings, Boston, November 9–11, 1990.

Clapp, J. M. (1990a), "A Methodology for Constructing Vacant Land Price Indices," *AREUEA Journal,* 18(3), 274–293.

——— (1990b), "A New Test for Equitable Real Estate Tax Assessment," *Journal of Real Estate Finance and Economics,* 3, 233–249.

Clapp, J. M., and Giaccotto, C. (1991), "Estimating Price Trends for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods," Working Paper W 91-18, University of Connecticut, Finance Dept.

Clapp, J. M., Giaccotto, C., and Tirtiroglu, D. (1990), "Repeat Sales Methodology for Price Trend Estimation: An Evaluation of Sample Selectivity," Working Paper W 90-16, University of Connecticut, Finance Dept.

DiPasquale, D., and Wheaton, W. C. (1990), "Housing Market Dynamics and the Future of Housing Prices," draft paper presented at the AREUEA mid-year meetings in Washington, DC, May 27–28, 1990.

Griliches, Z., (ed.) (1971), *Price Indices and Quality Change,* Cambridge, MA: Harvard University Press.

Greene, W. H. (1990), *Econometric Analysis,* New York: Macmillan.

Hendershott, P. H., and Peach, R. (1990), "Is the Mankiw and Weil Forecast Believable?" draft paper presented at AREUEA mid-year meetings in Washington, DC, May 27–28, 1990.

Kennedy, P. (1984), "On an Unfair Appraisal of Vertical Equity in Real Estate Assessment," *Economic Inquiry,* 22, 287–289.

Knickerbocker, N. N. (1990), "Aircraft Noise and Property Values," draft dissertation, University of Maryland, Dept. of Economics.

Kochin, L. A., and Parks, R. W. (1982), "Vertical Equity in Real Estate Assessment: A Fair Appraisal," *Economic Inquiry,* 20, 511–531.

Mankiw, N. G., and Weil, D. N. (1989), "The Baby Boom, The Baby Bust, and the Housing Market," *Regional Science and Urban Economics,* 19, 235–258.

Palmquist, R. B. (1979), "Hedonic Price and Depreciation Indices for Residential Housing: A Comment," *Journal of Urban Economics,* 6, 267–271.

Pollakowski, H. (1987), "Owner-Occupied Housing Price Change in the U.S., 1974–1983: A Disaggregated Approach," report from the Joint Center for Housing Studies of MIT and Harvard University, Cambridge, MA.

Pollakowski, H. O., and Wachter, S. M. (1990), "The Effects of Land-Use Constraints on Housing Prices," draft paper presented at AREUEA mid-year meetings in Washington, DC, May 27–28, 1990.

Shiller, R. J. (1990), "Arithmetic Repeat Sales Price Estimators," *Journal of Housing Economics,* 1, 110–126.

Thibodeau, T. G. (1989), "Housing Price Indexes from the 1974–1983 Annual Housing Surveys," *AREUEA Journal,* 1, 100–117.

White, H. (1980), "A Hetroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica,* 48, 817–838.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association,* 57, 348–368.

This content downloaded from 131.123.92.13 on Tue, 17 Oct 2017 05:19:17 UTC
All use subject to http://about.jstor.org/terms