Estimating Price Paths for Residential Real Estate

Authors John W. Birch and Mark A. Sunderman

Abstract

Several approaches have been used to estimate and adjust for price movements in residential real estate; however, weaknesses remain in current systems. This study incorporates a different way of measuring temporal price patterns. The method involves a time series model, an approach not previously employed when estimating real estate price movements. The finding indicate that the proposed technique is likely more accurate than current procedures. The method also represents a significant adaptation of standard time series models. For the task at hand, the new model is arguably preferable to the more standard versions.

Introduction

Having an accurate price series is useful and sometimes critical for researchers, as well as for real estate practitioners. A price series is often needed to adjust the sales price of a house to either an earlier or later time period. On other occasions, knowledge of the precise movement of prices of houses is important.[1]

The price of a property often changes over short time periods, sometimes in unpredictable ways. After adjusting or standardizing prices for any changes in real property characteristics, there often remains a significant movement in typical price levels. A sequence of monthly or other time intervals can be used to approximate the pure price path following these property adjustments. Several approaches have been taken when measuring movements in the dollar value of property. However, there are weaknesses in each, and these can lead to distortions in results. The purpose of this article is to develop a method to obtain a more accurate picture of market price patterns than may occur using current methodologies.[2]

As a description of the market for residential property, Bryan and Colwell (1980) state: "It is a mistake to imagine that housing prices move upward without interruption. On the surface, it may appear that housing prices have exhibited such behavior, however, there may be hidden cyclical and seasonal fluctuations. Over the past several decades, the tendency, as reflected in U.S. price data, had been for new home prices to rise. However, substantial deviations from this tendency have occurred."

Housing prices move through time based on secular trends, cyclical, seasonal and other measurable effects. In addition, the observed time series will always contain the impact of a large number of random forces. By definition, each of these random forces has too small an impact on any time series to be separately explained. But the combined effect of these forces can be measured. This combined effect is characterized by the random nature of its impact over time.[3] Given this way of viewing a time series of price movements as a combination of explicable and inexplicable forces, it is important to look for some method that attempts to separate the impact of these two kinds of forces in an optimal way. The final market price pattern should include the impact of these identifiable movements and omit the random, inexplicable effects. The field of time series is characterized by this view, where a main purpose of the analysis is to separate the explainable component in the data sequence from the random movement that is always present in some degree.[4]

The method introduced in this study ensures all explainable forces influencing real estate price movements are accounted for in the final market price pattern depicted. The process thus isolates and removes from the observed series the combined effects of all the inexplicable, random forces that are present. In contrast, currently used methods for determining price index movements, including standard regression, do not contain a procedure that guarantees an optimal breakdown between the effects of these two kinds of time related forces. As will be seen, the impact of unusual or unpredictable phenomena on price movements is not likely to be properly accounted for when applying standard regression or repeat sales methodologies. One example of such an event is a change in property tax rates within the time frame of the analysis. Other examples will be given later. Part of the problem is the subjective choice of the length of the time intervals to be employed in regression models. If intervals are too lengthy, the market pattern tends to be too smooth, excluding some of the explainable phenomenon. If too short, it encompasses some of the random elements in the series and there is spurious accuracy in the resulting market pattern depicted.

Time Adjustment Systems

There are numerous systems suggested in the real estate literature for measuring price movements. All such systems entail the use of regression models in one form or another (see Bryan and Colwell, 1982; Meese and Wallace, 1997; and Clapp and Giaccottto, 2002).

The regression approach to real estate valuation was first used in 1939 (see Bryan and Colwell, 1982). It also remains as a currently popular method. Several types of models have been employed over the period since its inception. In order of introduction, the first involves the use of market value characteristics of sold properties, with sales price as a function of these varying property traits. Both real property characteristics and neighborhood location variables are used in such models to account for variations in sales prices caused by real differences in sold properties. In this way, properties are standardized for differences across the sample of available property sales. Typically this approach involves pooling into one model all sales prices and associated property and location characteristics data for several successive time periods, with time incorporated as a separate variable or variables. A price index series is derived from the coefficient(s) on the time variable(s).[5] The model can be summarized as:

$$SP_{it} = F\{PL_{ijt}, G_t, U_t\}, \quad (1)$$

where SP_{it} represents the property selling price of the *i*th property at sale date *t*. PL_{ijt} stands for the set of *j* property and location variables affecting the market value of the *i*th property at sale date *t*. G_t is a general representation for the set of variables that model time movements of standardized prices and U_t is the random variable representation.

A major problem with this specific method is the large number of independent variables generally required for all the factors needed to explain the range of values of sold properties.[6] As with the repeat sales approach, the number of property sales has to be fairly large to attain reliable results.

Regression modeling for time effects at first involved simple linear and quadratic trends. Later, Bryan and Colwell (1982) introduced more flexibility into the modeling of time movements. In terms of annual movements of price, one of their regression models incorporates a novel methodology in the form of a set of annual dummy variables. There is one variable to represent the beginning of each of the years in the analysis period.[7] The two dummies closest to the sale date are assigned values that sum to unity, with the two values being proportionate in each case to the closeness of the sale to that year's beginning and end. For that particular sale, all other beginning year dummies take a value of zero. The resulting estimated path of price is a point on a log linear function that moves smoothly from the beginning of each year to the beginning of the next year. Shifts in log linear slope occur only at the beginning of each new year. The system provides more annual flexibility than linear or quadratic movements, being essentially an unconventional piecewise linear technique, with nodes at each year end within the period analyzed.

For intra-year patterns, Bryan and Colwell used monthly dummy variables, with each variable accounting for variations in the monthly price level of standardized housing. This seasonal specification, along with the annual dummy modeling already mentioned, gives the extra flexibility "sufficient to capture the full range of price variability" of standardized housing (Bryan and Colwell, 1982: 58).

This Bryan-Colwell model is certainly an improvement over earlier models, which lack an intra-year seasonal component and contain linear or quadratic forms to take care of longer term movements. At the same time, there is a problem when introducing a model that uses specific time interval specifications to show changes in short-term price levels, such as in the case of monthly seasonal dummies. There is no guarantee that these time intervals are of optimal length. The monthly time intervals for the Bryan-Colwell intra-year dummies result in approximations of the price pattern. Thus, if the intervals are too short, the resulting dummy variables will pick up a portion of the random component in the time series under analysis. If too long, the resulting movement may not vary sufficiently and some of the within year pattern will be too smooth, missing a part of the seasonal or any other significant intra-year fluctuation. The same idea can be applied to the annual dummies used in the model, although the amount of error in distinguishing between the market and the random elements in the series may be of lesser importance because of the longer time intervals inherently involved. In effect, the specification of the time intervals for the two sets of dummies is likely to generate error when arriving at the final market price pattern. As will be noted later, it is likely that this problem cannot be properly identified by the use of tests for serial correlation of residuals.

An alternative regression-based technique, referred to as the repeat sales method, was introduced in 1963.[8] This approach was so described since only repeat sales data were employed in the analysis. The technique involves comparing the price of a property resale with its initial earlier price, where both transactions occur within the period under analysis. The procedure can be shown to be equivalent to the use of two multiplicative regression models, one for initial sales and one for repeat sales of the same set of properties (Bryan and Colwell, 1982). The simplicity of this repeat sales method is appealing, but the method has at least one serious drawback (*e.g.*, Case and Shiller, 1987; Haurin and Hendershott, 1991; Clapp and Giaccotto, 1992; Dombrow, Knight and Sirmans, 1997; and Gatzlaff and Haurin, 1997). There is typically only a small fraction of properties resold within the time frame of many studies involving price movements. Case and Shiller found that less than 5% of houses resold within a 15-year time span. This percentage will tend to be even lower for the typically shorter time spans used in much real estate analysis. Clapp and Giaccotto also observed a small percentage of repeat sales in residential property transactions. In addition, smaller market sizes would worsen the problem, given the more limited number of sales, and thus repeat sales, generated from such areas. For these reasons alone, the repeat sales method is often an unreliable way to estimate market price movements compared to other approaches.

A third approach to obtaining price patterns within the multiple regression approach was initially suggested by Gloudemans (1990) and Jensen (1991). The procedure, sometimes called the assessed or appraised value (AV) approach, involves the use of a single proxy variable to replace all the property and location variables used in the method previously described.[9] Gloudemans pointed out that individual property AVs would be a sound proxy to replace all the independent variables previously used to explain differing property market values. Employing this proxy approach to standardize the cross section of property values has the advantage of not requiring nearly the detailed information needed when doing standard regression modeling involving a large number of independent variables. In many situations, such information is either not available or of questionable quality. A major disadvantage of this method can be the accuracy of AVs. As noted later, this disadvantage may not be too serious, although other issues remain. Basically the proxy regression model can be summarized as:

$$SP_{it} = F\{AV_{iT}, G_t, U_t\}, \quad (2)$$

where AV_{iT} represents one independent variable, the appraised market values of the i sold properties at some specified and fixed time, T. All other variables are as previously defined.

Various methods were employed by Gloudemans (1990) and Jensen (1991) to model time within their proxy regression models. Gloudemans used successively larger order time variables with annual time sub intervals. As pointed out by Clapp, Richo and Giaccotto (1994), this specification is subject to the problem that a choice of the number of order terms must be made, and that the specification is subjective and might be inappropriate. Also the annual time interval used implies intra-annual seasonal or other market movements are absent, an assumption that is often questionable.

Jensen (1991) also excludes forces less than one year in duration. In addition, his longer term movements are generated using continuous spline regression, where price trends at node points suddenly shift, by the nature of the model. Again the result becomes an approximation based on choice of the time location for node points and the linear or non-linear specification for movements between nodes, as well as on the assumption there are no intra-annual movements present in the data.

Clapp, Richo and Giaccotto (1994) use a log-based AV type model where dummies are employed to obtain various steps or levels over a sequence of three years of quarterly intervals. As with Bryan and Colwell (1982), the question again arises as to whether these time intervals are too long or too short. If too long, the model would generate market movements that are too smooth, deleting any important measurable patterns occurring within shorter time frames. Monthly or some other shorter intervals could be more appropriate. Alternately, too short a time interval specification will generate a spurious accuracy in the fit, including random movements within the depicted price pattern. The regression model is thus likely to contain a specification error based on the length of the basic time interval chosen for the analysis. It follows there is no guarantee of an optimal separation of systematic explainable market movements from random ones when arriving at the final depiction of the market pattern.[10] Regression models and repeat sales approaches both require a subjective specification of the sub interval unit for the analysis. In the event that a particular short term period, such as a year, contains unpredictable extra variability caused by unexpected and unusual events, these models are likely to pick up too little of the resulting price index movement. Yet changes in market price levels attributable to these unpredictable changes in conditions occur with varying incidence through time and across different communities. At the same time, these phenomena are important in their impacts on market price movement. Such unpredictable phenomenon could include large layoff effects, very adverse weather conditions, changes in property tax rates, announcements about the entry (or departure) of large business operations, and so on. Even variations in the timing and intensity of normally repeated phenomenon can cause difficulty. As an example, the February effect found by Bryan and Colwell might take place in some years more in the last week or two of January than in February. Further, there is no guarantee that autocorrelation on the residuals of a standard regression will result in the elimination of nonrandomness from the final price patterns depicted.[11] In general, there are likely misspecifications in time interval lengths that create errors in the estimated price path when employing current methods.

Given these interval specification problems associated with current methods, it is worth considering an alternative approach to modeling market price movements. First, though, it is important to see how to obtain a price indicator series that is standardized for variations in property characteristics without having to use any regression methods to adjust for resulting differences in selling prices.

Obtaining a Standardized Time Series Reflecting the Market

In regression analysis, standardizing is carried out using either the proxy independent variable (AV), or a whole set of variables to account for individual differences in property characteristics. There is another way to standardize data that does not require regression analysis. Dividing the sales price (SP) for each sale by the property's AV gives a set of ratios, one for each sold property. Keeping in mind that the AV values are all for an identical fixed date, whereas the sales prices vary over the entire period of analysis, these ratios represent a set of numbers that typically move in proportion to the market price of a standardized set of properties. If, for instance, the ratios tend to rise and are highest toward the end of the period, that directly indicates the market has risen to that extent. These SP/AV ratios can be used to obtain a single time series directly reflecting standardized residential property price movements. Once a basic time interval is chosen for the analysis, the mean of these ratios can be found for each interval within the whole period under analysis. These mean ratios represent a time series that directly approximates the relative movement of standardized property prices over the period.[12]

There is a possible issue concerning bias if AV statistics are used when estimating market price movements. The problem occurs whether the AV regression approach or the SP/AV mean ratios are employed to standardize prices. There are two conditions that must both hold if AV-based estimates of price movements are to be biased. One of these conditions is that properties sold within time intervals significantly change in character over successive intervals. For instance, if more property that is inherently cheaper is sold in earlier compared to later intervals, there is interval nonhomogeneity in property sales. The second condition required for the presence of bias involves vertical inequity. This is the situation where the ratio of AV to market values is typically not the same for less valuable properties relative to more valuable ones.

A test for at least one of these two conditions is necessary to ensure the subsequent AV-based analysis is valid. Tests for homogeneity of properties sold across the whole set of time intervals can be conducted to determine the presence of this bias.[13] Tests for vertical inequity are also available in the literature.[14] If both conditions are present, implying bias, some adjustment of the data is advisable prior to continuing with the AV-based analysis. Either interval sales samples must be made homogeneous across all intervals, or vertical inequity needs to be removed from the AV values before any AV-based procedure is employed.[15] There is no bias present if either condition is false. In particular, there is no bias if properties tend to be equally underestimated relative to their true market values.

Given these caveats, the time series of interval mean SP/AV ratios for the sequence of intervals can be used as a reflector of changes in values of standardized housing stock. However, since these mean ratios are subject to sampling error, the resulting time series contains a random component. The random portion can be eliminated by fitting the data using a time series model. These models are designed to separate the systematic, explicable movements in any data sequence from the random variation or noise that is present in the series.

Choosing a Time Series Model

Underlying observed price level movements, there is a continuous ongoing search for the correct present values inherent in the expected flow of future services of whatever asset is being considered. Expectations about the future have a direct impact on the market level attained at any particular point in time. Future price levels and current ones are related. However, it seems highly unlikely that expectations and thus price movements should follow any precise mathematical function. Functional forms can only generate approximations of the way the market is moving. Thus, it is more realistic to have a system that does not force a prior specification of the nature of the market movement. It is better to use methods that have enough flex to measure all systematic movements as dictated by the data themselves, and to exclude random effects where there is no explainable pattern.

Typical time series models can be considered as "one-way" processes, where the data are analyzed through time as it is usually looked at, from the past to the present and forward from there. Most time series models are of this type mainly because concern with such models has been heavily oriented toward forecasting, with its stress on using the past to project the future. However, interest here lies with time series where observations are already determined. The purpose is not to forecast but to smooth the data in an optimal fashion, deleting the unexplainable random effects in the series so as to isolate explainable movements in the price pattern. Since all data are available at the beginning of the analysis, it is quite possible to smooth the data in both directions, forward through time and backward as well. This dual approach is used to generate two-way time series models. The purpose is to smooth existing data as opposed to forecasting future values.

Using this two-way technique increases the ability to identify movements in the data that cannot be found as accurately when basing the smoothing on past values only. This can best be seen by realizing that random movements are not identical when data are viewed from the past to the future as opposed to looking at outcomes that have all occurred.[16] Some of the inexplicable movements that cannot be forecast might well be explicable after the fact, and can thus be considered a part of market movement from the ex post view taken here. A two-way process is required to pick up these ex post movements as a part of the market pattern. Thus, a two-way method picks up more information inherent in the time series, compared to one-way methodologies.

The two-way model developed here is an extension of standard time series (one-way) modeling. In particular, it builds on a simple (one-way) exponential smoothing model. Thus, it is helpful to begin with a brief review of the essentials of one-way (standard) time series modeling. Once that is complete, it is easier to understand a two-way time series smoothing model. More precisely, the method to be used here represents an extension of Brown's (1963) single exponential smoothing methodology.

Exponential Smoothing

Exponential smoothing methods were developed over the ten-year period following 1955. In addition to Brown (1956), major contributions included those of Holt (1957) and Winters (1960). Brown developed one parameter exponential smoothing for time series. Three different models of increasing complexity were introduced, for time series showing stationary, linear and nonlinear movements. These were called the single, double and triple exponential smoothing models, respectively. Later, Holt advanced a two-parameter procedure for series showing evidence of linear trend. This procedure turned out to be a small improvement on Brown's double exponential method for linear time series (see Makridakis, 1982).

Neither Brown's (1956) nor Holt's (1957) methods are designed for time series containing seasonal or other (cyclic) patterns. In contrast, Winter's (1960) method is a three-parameter system that solves the problem of adequate forecasting where fixed seasonality is present. At the same time, Winter's method may not perform all that well when there are patterns that are more irregular than seasonals, or when seasonals are changing [see Makridakis (1983: 118) and Gardner (1985: 15)]. It is also a problem that the task of finding jointly optimal values for the three required Winter's smoothing parameters is more complex and involves more computer work [*e.g.*, Gaynor and Kirkpatrick (1994: 395)].

Theoretical underpinnings for exponential smoothing were largely resolved when Brown and Meyer (1961) introduced the fundamental theorem of exponential smoothing. Following several years involving empirical comparisons and lesser theoretical contributions, most of the interest in time series shifted toward Box-Jenkins (ARIMA) methodologies.[17] These developments occupied much of the field of time series analysis over the ensuing twenty years.[18] ARIMA methodology involves more sophisticated time series models compared with those of early exponential smoothing techniques, but also requires a larger number of observations than is usually available for analysis of short-term market movements.[19] Thus, ARIMA modeling is of very limited value for short-term analysis of time series.

There is an important additional problem when using one-way orthodox time series models to fit a data sequence. These models are based on forecasting values from the present to the future, primarily since short-term forecasting was the original stimulus for their development. As already mentioned, any short-term patterns that do not mirror the prior movements in the series will be excluded in the final fit when using these forward looking or one-way types of models to fit the data sequence. These unpredictable values may turn out to be due to forces that are an explainable, important part of the market when examined after the fact. Such important effects need to be included within the market movement. At the same time, there always remain unexplainable random movements in the series that should be excluded from the final market movement picture. Thus, unpredictable movements prior to the experience are not the same as random movements in an ex post sense. It is only these ex post random movements that need to be eliminated from the final market picture.

In the analysis of observed data, two-way models can be employed to overcome this inherent problem in orthodox models. Such analysis, involving a combination of standard forward modeling with a reverse time identical backward process, will cover significant movements in the data, including those that were not predictable before the fact. The two-way process only leaves out the ex post, unexplainable random residuals that are not a part of the underlying short-term market process being evaluated.

Brown's (One-Way) Single Exponential Smoothing

Brown's (1956) single exponential smoothing model is a system that can be described as:

$$S_t = \alpha Y_t + (1 - \alpha)S_{t-1}, \qquad \text{for all } t = 0, 1, 2,... . \qquad (3)$$

Y_t represents the actual or observed time series value in period t.[20] S_t and S_{t-1} stand for the smoothed values for the periods t and $t - 1$.[21] The symbol α is called the smoothing constant.[22]

The procedure for smoothing can be described as follows. Assume that t is a number representing the time period for each of the observed data points, t = 1,2,3,... T. Thus, when $t = 1$, the first observed value in the actual series is represented as $Y_{1.}$ Also when $t = 1$, $S_{t-1} = S_0$, usually called the value for the initial benchmark interval, $t = 0$. It can be thought of as a typical value for the time series in period 0. The terminal benchmark value refers to a typical value in the subperiod or time interval immediately following the end of the whole time period under analysis.[23]

An α of 1 means there is no smoothing, since S_t would always equal Y_t and the ‘‘smoothed’’ series would be the same as the actual series. This would not be appropriate if the observed time series contains the effects of any random forces. At the other extreme, an α set to 0 would mean the smoothed values would never change, remaining at the starting value set for S_0 in the initial benchmark interval. Such an extreme would also be improper if the time series contain any systematic movement (trend, seasonal, etc.). It is apparent that a best time pattern fit will occur when α is set somewhere between 0 and 1 for any time series that contains elements coming from both random and systematic forces.[24]

Single exponential smoothing has a major flaw that makes it inappropriate to apply to data with either a rising or declining trend or any pattern dominated by such movements. The fitted or smoothed values found will lag behind the trend, being too high when the trend is declining and too low when it is rising.[25] Further, this bias worsens the more removed in time a fitted value from its initial benchmark interval.

An additional weakness is that exponential smoothers tend to lag systematically behind the observed peaks and troughs of an actual time series. The weakness is caused by the use of the one-way procedure for obtaining smoothers (from the past to the future). The problem characterizes one-way smoothing models when seasonal or other short-term patterns are observed in the time series.[26]

Two-Way Time Series Modeling

To overcome these weaknesses, first reverse the order of the data, to find a second single exponential smoothed time series, working from the last time interval back toward the first.[27] The biases in this backward exponential smoothing tend to be the mirror image of the biases in forward exponential smoothing. Thus, if there is a trend movement in the observed series, there will be a time lag in the backward smoothed series that is the opposite of that in the original smoothed series. As already indicated, the trend bias in the forward exponential smoothing gets steadily worse as intervals that are more and more removed from the beginning of the period under analysis are examined. It is at its smallest in the first interval of the period being examined. In contrast, working in reverse time order, the bias for the backward smoothing process is largest for this same first time interval of the whole period, since that is the interval furthest removed from the end of the whole period. Since the biases of the two systems are the opposite of each other, they can be eliminated by obtaining a weighted average of the forward and backward smoothed values. These two smoothed sets of values are called the forward and backward smoothers, respectively.

To illustrate the elements of the two-way process, imagine a set of monthly SP/AV mean values that are generally rising in a linear fashion over a short time interval such as a year. These are shown in Exhibit 1. Assume also, for simplicity, that the smoothing constant, α, is equal to zero. Note that because of the upward trend in the monthly values, the initial benchmark value (assumed to be 0.9) will be less than that for the terminal benchmark (assumed to be 1.02). Since $\alpha = 0$ in this example, both the forward and backward smoothers will be horizontal lines at levels equal to the initial and terminal benchmarks, respectively. Exhibit 1 shows that neither the forward nor backward one-way smoothed lines adequately fit the example data, and that their fits of actual data worsen for months further removed from their respective benchmark months.

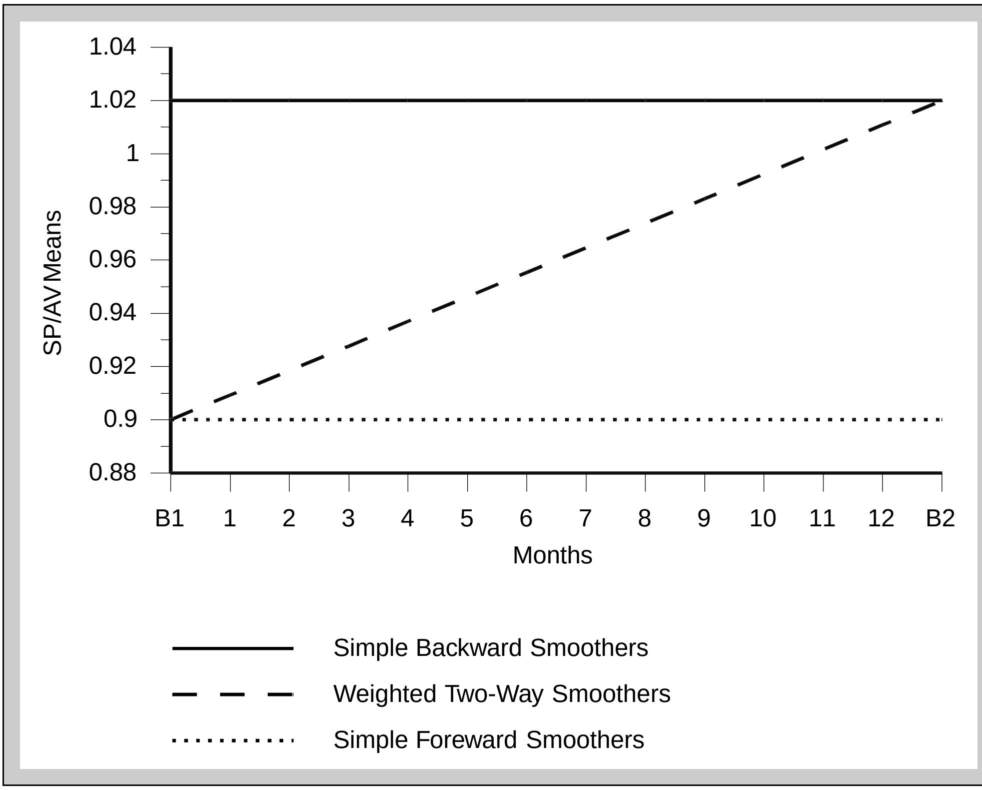
Now obtain weighted averages of the two smoothed lines for each month, using weights based on the amount of time from the month being measured to the two benchmark months. For example, the first month will be one month removed from the initial benchmark month, whereas this same month will be 12 months removed from the terminal benchmark month. Thus, the forward smoothed line value has a weight of 12/13 and the backward smoother weight is 1/13. Therefore, the weighted average for the first month in the 12-month series will be:

$$S_1^* = (12/13)Sf_1 + (1/13)Sb_1, \quad (4)$$

where Sf_1 is the simple forward smoothed value and Sb_1 is the simple backward smoothed value, both for the actual first month. S_1^* designates the weighted two-way exponential smoothed value, and the subscript 1 indicates it is the smoothed value for month one. The weights for month 2 would be 11/13 and 2/13, and so on for months 3 through 12.

This result, as shown in Exhibit 1 with $\alpha = 0$, will give a series of weighted smoothers that lie on a straight line joining the initial and terminal benchmark values.

Exhibit 1 | Simple Forward and Backward Smoothed Values and Weighted Two-Way Smoothed Valves (= 0)



Finding the Best α For Two-Way Weighted Exponential Smoothers

In exponential smoothing, the search for an optimal smoothing constant is based on the following idea. α smoothing constant values set at 0 or close to 0 give weighted smoothers that lie on or close to a straight line, respectively. Such fitted lines will often miss some of the systematic forces in observed time series (*e.g.*, seasonal movements). At the other extreme, values of α approaching 1 give weighted smoothers that fit observed data so closely that some random effects are apt to be included in the patterns they show. The question is how to find optimal α values.

The optimal smoothing constant is typically found by first setting α equal to 0, and deriving the fitted line that results. After this, a set of errors or residuals can be found. These residuals are defined as the actual or observed time series values minus the weighted smoothers or fitted values for each of the monthly time intervals. This set of residuals is then tested as a random sequence. If the test shows randomness, the optimal smoothing constant is an α of 0, since there is no further systematic element apparent in the residual sequence. If the test indicates the residual sequence is not random, the analysis is repeated using a slightly increased value for α. Once a new set of slightly more flexible fitted values has been found, a new set of residuals is determined and tested for randomness as before. Successive iterations of increased α values will make ensuing fits increasingly flexible, with an improved ability to identify any remaining pattern initially found in the residuals. The α iterations continue until the residual sequence test does not reject the hypothesis of randomness. The optimal α is the one used when the sequence of residuals is first found to be random. In this way, a smoothing sequence is found that contains all the systematic forces as determined by the behavior of the observed data.

A number of tests for randomness could be used for this procedure. In particular there are a variety of nonparametric tests.[28] Alternatively, the Durbin-Watson test is a more powerful option, although it does require fifteen or more time interval observations to be applied. Also the test can generate an indeterminate result, not an entirely satisfactory attribute [see Neter, Wasserman and Kutner (1996: 450–4)].

The Two-Way Model: Illustration and AV Regression Comparison

The purpose of this section is to use contrived data to show how the proposed model applies to a set of market prices over a sequence of time intervals. An AV regression model is also applied to the same data, and results of the two methods are compared.

The data were generated as follows. First a linearly related set of fifty monthly mean SP/AV ratios was generated. These range from 0.855 to 1.10. Also a separate series of fifty ratios was determined to represent monthly seasonal movements for the same period. These center on 1.00 and fluctuate around a range from 0.94 to 1.06. These two monthly series were multiplied together to obtain the ‘‘true’’ monthly market value movement as measured by the resulting combined mean ratios.

The true market movement was then shocked by a randomly generated set of fifty data, with a mean ratio equal to 1.00 and standard deviation of 0.04. The resulting combined series became the ‘‘observed’’ sequence of monthly mean ratios.

Individual sales were found for each interval by first generating a set of sales prices for each of the fifty time intervals. The appraisals on each property were found such that the resulting SP/AV values for all sales in each monthly time interval would have a mean equal to the observed mean monthly ratio. This completed a set of data on which analysis could be conducted using both time series analysis of the monthly mean SP/AV ratios and regression AV analysis using SP and AV individual values within the data set. The advantage of constructing this kind of data is that it allows comparisons using both the regression AV and the SP/AV time series analysis. Actual data does not allow this type of comparison since true market values are never discernible. This approach is acceptable as an illustration provided the contrived data look reasonable. Certainly the data used here fit such a description. Thus, the two-way exponential smoothing method and the regression AV method (using monthly dummies to catch seasonal patterns) were applied to the data set.[29] The estimated market movements derived from each of the two different approaches were then compared with the true market pattern.

Exhibit 2 shows the earlier determined true market pattern as a heavy solid line. The solid white line shows the series that contains the random noise (the ‘‘observed’’ data used by the two-way and implicitly by the multiple regression approach). The estimated market movement based on the two-way exponential model is shown as a dotted line. The estimated market movement for the AV multiple regression model using monthly time intervals is shown on the figure as a fine dark line. From Exhibit 2 it is clear that the AV regression line follows very closely the observed data, whereas the two-way exponential results fit the market pattern much better.

The fitting error for each of the two models was found as the mean absolute error (MAE) between each model's estimated market levels and the corresponding *true* levels for the forty-eight months of model fit. The MAE using the two-way exponential model is .01998, *i.e.*, the predictions are, on average, 1.998% removed from (above or below) the true market level. For the AV regression model, using monthly time intervals, the MAE was found to be 0.0330, or the model predictions were typically in error by 3.30%. This is about 60% less accurate than the exponential model results.

The multiple regression model was fitted a second time, where in this case the data were analyzed using six-month time intervals. This biannual time interval arrangement gives results using the MAE approach that are typically 3.6% removed from the true market pattern, clearly a performance inferior to both of the prior models. These six month interval multiple regression fitted values are shown in Exhibit 3 along with both the true market movement and the actual or observed time series of mean ratios.

The two-way exponential procedure is better than either of the regression models. In the case of the regression using monthly intervals, the graph indicates that the dummy levels basically follow the observed series random fluctuations. The interval used is apparently too short and there is thus too close a fit of the observed time path of prices. The variations in the true market are overstated since the model does not separate the random movement from the true movement. In general, when using multiple regression, too short a time interval will give this spurious fit, since dummies pick up the average level within each interval, and

Exhibit 2 | Fitting Capacity: Two-Way Exponential Smoothing and AV (Monthly) Regression Models

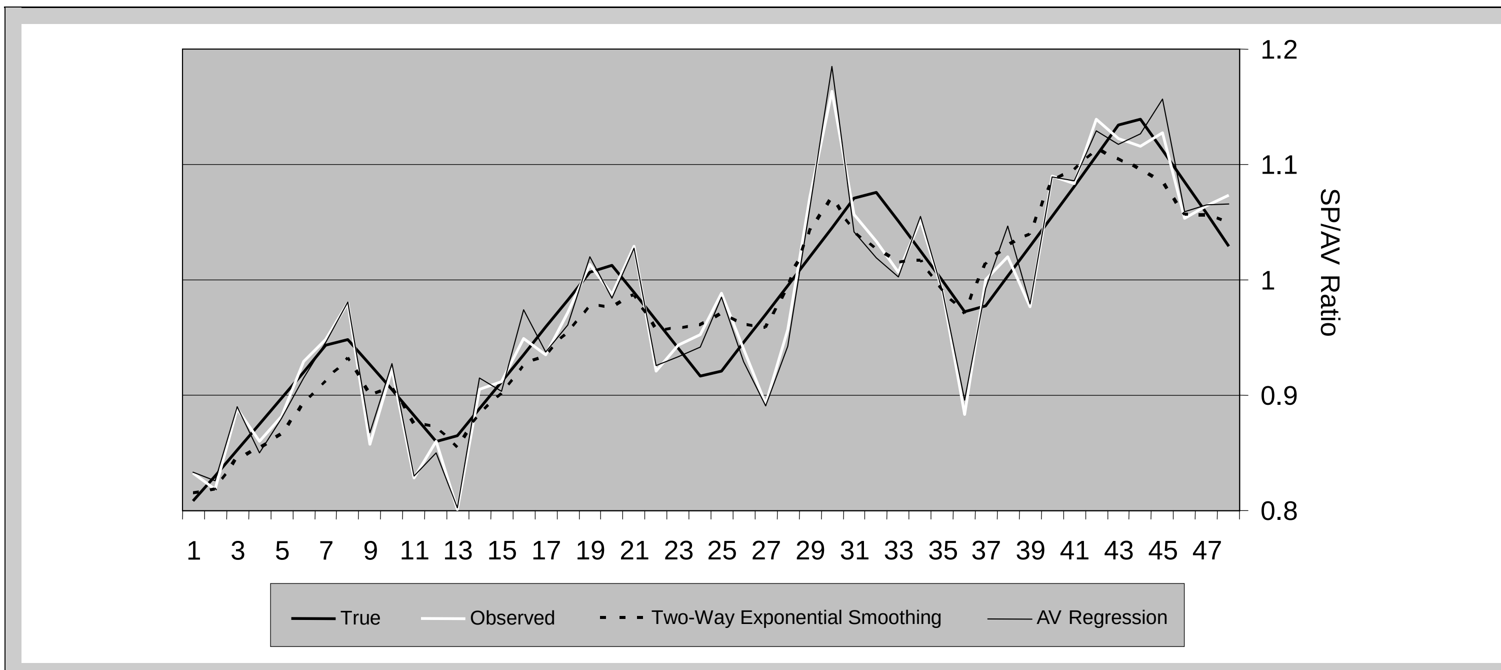
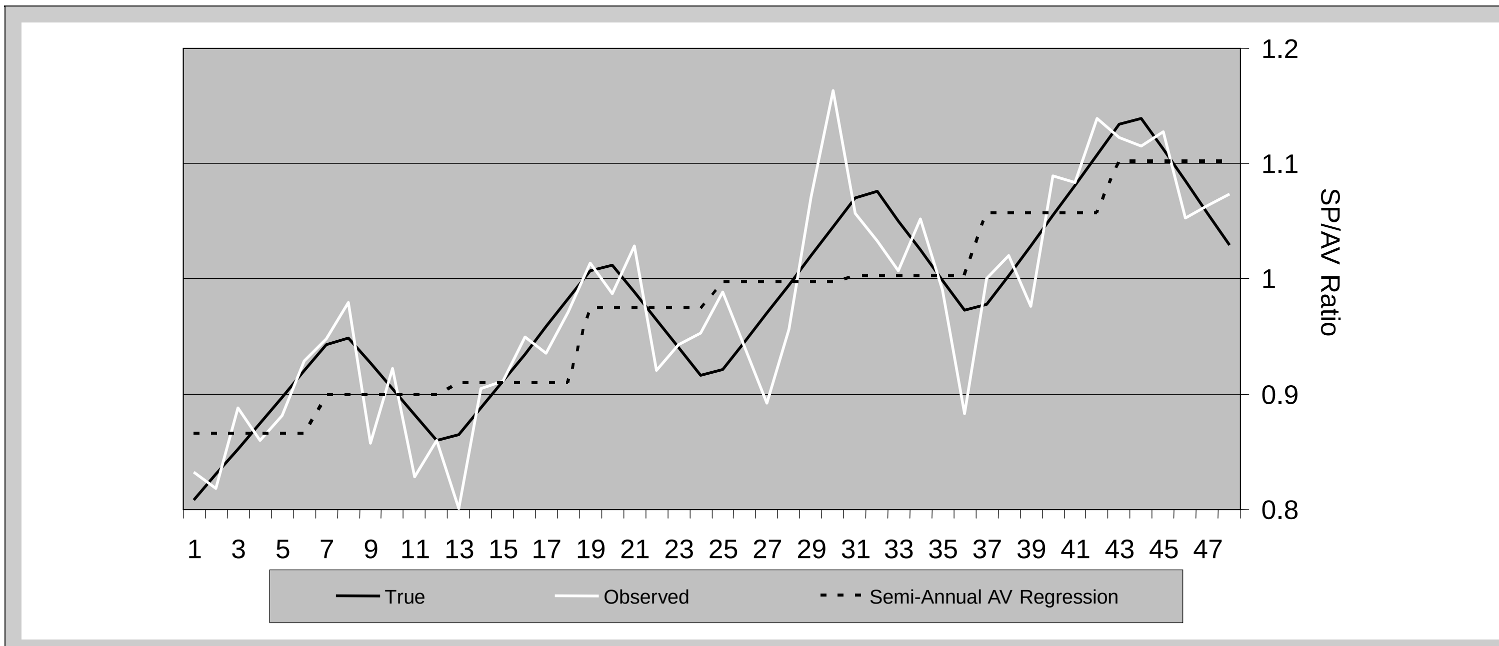


Exhibit 3 | Fitting Capacity: AV (Semi-Annual) Regression Model



these individual mean ratio values can be significantly impacted by ex post unexplainable random effects on sales prices and appraised values. This is particularly the case if the time interval generates few sales, giving rise to typically larger mean standard errors. As for the regression model using six-month data, there is clearly too much smoothing and this is of course because the six-month time interval is too lengthy.[30]

An additional comparative analysis was run using monthly intervals applied to a changed data set. A structural shift involving a 20% reduction in observed sales prices (and thus in SP/AV mean ratios) was imposed beginning with month 27 in the time series. The relative accuracy of the same two approaches remains essentially unchanged.

The two-way exponential model, with its built-in procedure for finding the optimal smoothing constant, does not require an iterative changing of time intervals. The iterative movement of the smoothing constant effectively replaces a repeated changing of the time interval required for optimal regression modeling. The result is a solution for fitting time series requiring only one application of the exponential model, with the iteration process encompassed within the overall procedure.

There is a further advantage to the use of the two-way exponential method compared with standard regression modeling. An integral part of the method is that a specific probability can be attached to tests for rejecting randomness in the sequence of residuals. Thus, using the time series approach permits a statistically-based belief that the residuals to be omitted from the final market movement are indeed a random sequence. For regression, one-way standard autocorrelation analysis of residuals is an added requirement before such a belief about randomness can be attained.

Finally, even randomness in the sequence of residuals for a given regression fit is not a sufficient condition for determining interval length correctness for that regression model. Models with intervals that are too short generally result in final market estimates with residuals that are a random sequence. If that is the case, some randomness is apt to be included in the fitted market estimates and it is not obvious that the interval length used is long enough. This problem does not exist when using an iterative exponential smoothing model.[31]

Conclusion

There have been different approaches to measuring short run market price movements. One of these methods limits itself to the use of repeat sales data. Another involves the direct application of regression models to all sales. At first, many independent variables were introduced into regression models to account for the effect of individual property traits on sales prices. In this way, the time pattern of market prices could be estimated using sales prices standardized for variation in property character. Bryan and Colwell (1982) represent a good example of this specific regression method.

Another advance occurred when appraised values (AVs) were used as a proxy to replace the large number of independent variables needed to represent these property and location characteristics. The resulting gain in degrees of freedom made it possible to better estimate market patterns when limited data were available. Gloudemans (1990) and Jensen (1991) began using this method to standardize property sales prices, while also obtaining time patterns for residential markets. Clapp and Giaccotto (1992) also used the AV approach while applying dummy variables to measure variations in price levels over a series of quarterly periods. Schwann (1998) employed the same approach but replaced dummy variables with an autoregressive scheme.

All of these regression models require specification of some fixed time interval length with results depending on the choice made. There is no systematic, built-in test that the interval chosen is the correct one. The optimal length depends on the presence and character of any events that may have impacted the data within the period under analysis. The two-way exponential smoothing methodology escapes this difficulty. This same approach also requires less data than most previously used methods.[32] Third, the system largely overcomes the problem of standard or one-way autoregression and one-way exponential smoothing modeling, where fitted values lag systematic movements in time series. The two-way model incorporates all the flex necessary to describe any systematic market patterns, including seasonal and cyclical movements. There is also sufficient sensitivity to cover unpredictable movements that are, at the same time, ex post explainable parts of the market, should they occur. At the same time the method does not allow so much flexibility to include inexplicable, after the fact, random effects.

In general the following conclusions seem in order. First, the regression modeling process can generate serious fitting errors as a result of a misspecified basic time interval used in the analysis. An iterative regression procedure, involving systematic changes in the interval length used, could be employed to overcome the problem to a degree, but such an approach is awkward at best. In contrast, the two-way exponential time series procedure tends to avoid these problems. The time series model is preferable to the regression approach essentially because it incorporates an iterative process for estimating the smoothing constant α that results in an optimal fit for the model. Iterations for α in time series models are an alternative for iterations involving the length of the time interval used in regression models.

The main purpose of this article has been to report on the development and application of a two-way exponential smoothing system for effectively estimating true market movements in residential property prices. The advantage of the method is that it is neither too rigid nor too flexible, containing a procedure that is designed to accurately find the systematic movement evidenced by the observed price time series. Thus, the approach seems to overcome some difficulties attached to earlier methodology.

Endnotes

[1] On a more specific level, in appraisal work there is often a need to determine what has happened to the market over short periods of time, perhaps less than six months. Without this knowledge, it is impossible to make accurate adjustments to property value for changes in the real estate market. In the same vein, assessors need to adjust sales prices on property sold over recent months (or years) to a given point in time to compare with assessed valuation in order to determine the accuracy of the assessment process. Determining assessment accuracy is critical to homeowners, as it will impact their property taxes, as well as to local and state oversight agencies.

[2] A residential property may have experienced a significant change in lot or improvement characteristics over the time period under study. Thus, for example, a room might be added to a house, a sidewalk installed, and so on. It is assumed that sales prices or appraised values have been adjusted for these kinds of changes, so that these prices and values are comparable for the same property. Data initially cleaned in this way is needed to properly determine changes in price of a constant or unchanging stock of housing. To the extent the data are not cleaned, all methods for finding price movements suffer accordingly.

[3] For a more detailed description of the nature of random forces, see Johnston (1972: 10,11). Also of note is the classic statement on the same subject in Klein (1962). Another description of these random forces can be found in Nazen (1988: 9).

[4] See Gardner (1985) for a summary of the development of time series methodology.

[5] One variation is a strictly cross-sectional technique. In this approach, cross-sectional data describe conditions that exist at a given point in time (or within a very short time interval). Thus, implicit prices of the housing characteristics are estimated in separate hedonic equations for each time interval. Conventional price indexes can then be calculated by linking the results of each cross-sectional analysis. Examples of conventional price indexes include Paasche and Laspeyres indexes. See Valliant and Miller (1989) and Wallace (1996) for such examples. See also Ferri (1977) and Palmquist (1980) for a variation of this technique.

[6] For a discussion of the relative merits of regression models, see Bryan and Colwell (1982).

[7] Bryan and Colwell (1982) state: ‘‘Each date of sale is defined as a linear combination of the end points of the year in which the sale occurs. Thus, the B(y) variables are the proportionate weights.’’ In their approach there is a B(y) variable for each year in which sales occur with B signifying the beginning of the year and y symbolizing a year.

[8] Bailey, Muth and Nourse (1963) first developed this technique.

[9] Clapp and Giaccotto (1992), Clapp, Richo and Giaccotto (1994) and Schwann (1998) use assessed valuations in the context of regression modeling. Clapp et al. uses dummy variables and Schwann employs an autoregressive structure to estimate time patterns. For a detailed description of the AV method, see Clapp and Giaccotto (1992).

[10] An iterative regression procedure could be used to overcome this problem, although the methodology is a bit awkward and can itself also generate problems. Thus, the first regression model might specify an interval that is quite lengthy, perhaps even one year in length. If tests show the unexplained residuals are not a random sequence, the interval

is shortened to provide increased flexibility to pick up this nonrandom sequence. The procedure repeats just until a short enough time interval is used so that the hypothesis of randomness is no longer rejected. This would solve one of the system's weaknesses, but at a certain cost. Aside from substantially increased computation, it should be noted that shorter intervals mean the number of intervals used in the model increases. Thus, the degrees of freedom for statistical testing declines. If the number of sales is at all limited, these developments can cause a serious deficiency in the suggested procedure.

11 For an example illustrating this weakness under autoregressive modeling, see Schwann (1998: 279).

12 Manson (1993) advocated the direct use of monthly median appraisal-sales ratios, with regression models fitted to ratio time series directly. Thus, Manson proposed that median sales ratios be calculated for each month, and that this series be regressed on time variables directly. Weighted regression was suggested, where the weights are the number of sales observed within each month. As determined later, problems also remain with this alternative regression approach.

13 The Kruskel-Wallis nonparametric test is an example of a simple test for homogeneity across intervals. See Conover (1980) for a discussion of this test. As an example, the test could involve the null hypothesis that houses sold in each interval have the same square footage, across all intervals.

14 For a review of several of the major approaches to test for vertical inequity, see Sirmans, Diskin and Friday (1996).

15 Birch, Sunderman and Hamilton (1992) describe a way to test and adjust for vertical inequity.

16 This is analogous to the well known situation that Monday morning quarterbacks can understand more about optimal game strategy than the quarterback who is actually making decisions as the game progresses. This is why Monday morning quarterbacks seem to be wiser than the actual participants during the game!

17 ARIMA is used to represent the term autoregressive integrated moving average, which is an alternate way used to refer to Box-Jenkins models.

18 It should be noted, however, that exponential smoothing became a popular applied business methodology in these years [see Ledolter and Abraham (1984: 79)].

19 An elegant review of both exponential smoothing and Box-Jenkins modeling is available in Gardner (1985). For a clear introduction to ARIMA modeling theory and practice, see Pankratz (1983).

20 The word ''exponential'' is used to refer to the fact that the S_t is an additive function of exponentially declining weights of past Y_t values. Thus, the model as given in the text is equivalent to:

$$S_t = \alpha Y_t + (1 - \alpha)\alpha Y_{t-1} + (1 - \alpha)^2\alpha Y_{t-2} + (1 - \alpha)^3\alpha Y_{t-3} + \cdots + (1 - \alpha)^n\alpha Y_{t-n}.$$

Thus, if $\alpha = .5$, then:

$$S_t = .5Y_t + .25Y_{t-1} + .125Y_{t-2} + .0625Y_{t-3} + \cdots + (1 - .5)^n(.5Y_{t-n}).$$

21 See Gaynor and Kirkpatrick (1994) for applications of simple exponential smoothing.

22 This α is not the α commonly used in referring to the type I error for tests of hypotheses.

23 There are several ways for determining reasonable estimates for these benchmark periods [see Makridakis, Wheelwright and McGee (1983: 121–23) and Ledolter and Abraham (1984)].

[24] See, Gardner (1985) for an excellent review of the whole family of exponential smoothing methods. Also, for a recent text, see Gaynor and Kirkpatrick (1994).

[25] See Mendenhall, Reinmuth and Beaver (1993: 693–94) for excellent graphs showing this lag.

[26] There has been some work using damped smoothers to try to improve Winters (1960) modeling of time series. But the problem of time lags is the result of the one-way fitting process inherent in these forecasting models, which makes for smoothers that lag behind unforeseen changing cyclic and seasonal movements, situations that may characterize real estate price index patterns.

[27] It might be thought that double, triple or higher order exponential smoothing would handle this problem of higher order changes in time series sequential patterns. Unfortunately, using these more complex procedures again requires a choice about the nature of the functional form. Such an approach is not easily justified when dealing with short-term changes in time series. In contrast, the technique developed here avoids both the complexity and subjectiveness implied when choosing between using higher order exponential smoothing models.

[28] See Levine, Ramsey and Berenson (1995: 645–52) for a description of the formula and general procedure for both the small and large sample runs tests. See also Swed and Eisenhart (1943) for an earlier example of a similar test.

[29] The AV model reported on here is a specific version of the generalized functional form already referred to as Equation (2). This version can be written as:

$$SP_{it} = b_0 + b_1 AV_{iT} + \Sigma b_{ij} D_{ij} + U_t,$$

where D_{ij} takes on a value of 1 for sales in the *j*th interval and 0 otherwise. The log version of this model was also tried and showed very similar results.

[30] It is apparent that interval specification error can significantly reduce effectiveness of the regression model as a way of accurately estimating the true market path of prices. It might be thought that an iterative process involving a systematic changing of regression model time intervals will solve this problem involving the choice of an optimal time interval. There are problems with this regression modeling adaptation. The procedure is awkward, requiring as it does a continuous rearranging of sales data into new time interval sets for each regression run. For a statistically based conclusion that price indexes are inferior when founded on regressions using quarterly or longer time intervals, see Englund, Quigley and Redfearn (1999).

[31] In this case it turns out that a 3-month interval comes close to minimizing error between estimates and true market values, but of course that is not always going to be the case. For other data, 5, 6, 11 weeks or some other length might be the best interval to use. Subjective judgments about optimal lengths for time intervals can be seriously wrong.

[32] Its robustness in the face of small samples could be further improved by replacing mean ratios with median ratios when constructing the time series fit.

References

Bailey, M. J., R. F. Muth and H. O. Nourse, A Regression Method for Real Estate Price Index Construction, *Journal of the American Statistical Association*, 1963, 58, 933–42.

Birch, J. W., M. A. Sunderman and T. W. Hamilton, Adjusting for Vertical and Horizontal Inequity: Supplementing Mass Appraisal Systems, *Property Tax Journal*, 1992, 11:3, 257–76.

Brown, R. G., *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall, 1963.

——., Exponential Smoothing for Predicting Demand, Paper presented to the 10th National Meeting of ORSA, San Francisco, 1956.

Brown, R. G. and R. F. Meyer, The Fundamental Theorem of Exponential Smoothing, *Operations Research*, 1963, 9:5, 673–85.

Bryan, T. B. and P. F. Colwell, Real Estate Market Performance: The Champaign County Case, *Illinois Business Review*, 1980, 37:6, 6–12.

Bryan, T. B. and P. F. Colwell, Housing Price Indexes, In C. F. Sirmans (Ed.), *Research in Real Estate, A Research Annual*, Greenwich, CT: JAI Press, 1982.

Case, K. E., and R. J. Shiller, Prices of Single Family Homes Since 1970: New Indices for Four Cities, *New England Economic Review*, 1987, 45–56.

Clapp, J. M. and C. Giaccotto, Estimating Price Indices for Residential Property: A Comparison of Repeat-Sales and Assessed Value Methods, *Journal of the American Statistical Association*, 1992, 87, 300–06.

——., Evaluating House Price Forecasts, *Journal of Real Estate Research*, 2002, 24:1, 1–26.

Clapp, J. M., G. F. Richo and C. Giaccotto, Adjusting for Time Using Sales-Ratio Data, *Assessment Journal*, 1994, 1:1, 56–62.

Conover, W. J., *Practical Nonparametric Statistics*, Second edition, New York, NY: Wiley, 1980.

Dombrow, J., J. R. Knight and C. F. Sirmans, Aggregation Bias in Repeat-Sales Indices, *Journal of Real Estate Finance and Economics*, 1997, 14, 75–88.

Englund, P., J. M. Quigley and C. L. Redfearn, The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition, *Journal of Real Estate Finance and Economics*, 1999, 19:2, 91–112.

Ferri, M. G., An Application of Hedonic Indexing Methods to Monthly Changes in Housing Prices: 1965–1975, *Journal of the American Real Estate and Urban Economics Association*, 1977, 5, 455–67.

Gardner, E. S., Exponential Smoothing: The State of the Art, *Journal of Forecasting*, 1985, 4:1, 1–28.

Gatzlaff, D. H. and D. R. Haurin, Sample Selection Bias and Repeat-Sales Index Estimates, *Journal of Real Estate Finance and Economics*, 1997, 14, 33–50.

Gaynor, P. E., and R. C. Kirkpatrick, *Introduction To Time-Series Modeling and Forecasting in Business and Economics*, McGraw-Hill, 1994.

Gloudemans, R. J., Adjusting for Time in Computer-Assisted Mass Appraisal, *Property Tax Journal*, 1990, 9, 83–99.

Haurin, D. and P. Hendershott, House Price Indexes: Issues and Results, *American Real Estate and Urban Economics Association Journal*, 1991, 19:3, 259–69.

Holt, C. C., *Forecasting Seasonal & Trends by Exponential Weighted Moving Averages*, Office of Naval Research, Memorandum No. 52, 1957.

Jensen, D. L., Modeling Time Effects in Computer-Assisted Mass Appraisal, *Property Tax Journal*, 1991, 10, 95–123.

Johnston, J., *Econometric Methods*, Second edition, McGraw-Hill, 1972.

Klein, L. R., *An Introduction to Econometrics*, Englewood Cliffs, NJ: Prentice Hall, 1962.

Ledolter, J. and B. Abraham, Some Comments on the Initialization of Exponential Smoothing, *Journal of Forecasting*, 1984, 3, 79–84.

Levine, D. M., P. P. Ramsey and M. L. Berenson, *Business Statistics for Quality and Productivity*, Englewood Cliffs, NJ: Prentice Hall, 1995.

Makridakis, S., The Accuracy of Extrapolation (Time Series) Method: Results in a Forecasting Competition, *Journal of Forecasting*, 1982, 1, 111–53.

Makridakis, S., S. C. Wheelwright and V. E. McGee, *Forecasting Methods and Applications*, Second edition, New York, NY: Wiley, 1983.

Manson, A. R., Time Trending of Ratios in Sales Ratio Studies, Unpublished manuscript, December 1993.

Mendenhall, W., J. E. Reinmuth and R. J. Beaver, *Statistics for Management and Economics*, Seventh edition, Wadsworth, 1993.

Meese R. E. and N. E. Wallace, The Construction of Residential Housing Price Indices: A Comparison of Repeat Sales, Hedonic-Regression and Hybrid Approaches, *Journal of Real Estate Finance and Economics*, 1997, 14:1/2, 51–73.

Nazen, S. M., *Applied Time Series for Business and Economic Forecasting*, Marcel Dekker, 1988.

Neter, J., W. Wasserman and Kutner, *Applied Linear Statistical Models*, Forth edition, Irwin, 1996.

Palmquist, R. B., Alternative Techniques for Developing Real Estate Price Indexes, *Review of Economics and Statistics*, 1980, 442–48.

Pankratz, A., *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, New York, NY: Wiley, 1983.

Schwann, G. M., A Real Estate Price Index for Thin Markets, *Journal of Real Estate Finance and Economics*, 1998, 16:3, 269–87.

Sirmans, G, S., B. A. Diskin and H. S. Friday, Vertical Inequity in the Taxation of Real Property, *National Tax Journal*, 1996, 48, 71–84.

Swed, F. and C. Eisenhart, Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *Annals of Mathematical Statistics*, 1943, XIV, 66–87.

Valliant, R. and S. M. Miller, A Class of Multiplicative Estimators of Laspeyres Price Indexes, *Journal of Business and Economic Statistics*, 1989, 7:3, 382–94.

Wallace, N. E., Hedonic-Based Price Indexes for Housing: Theory, Estimation, and Index Construction, *Federal Reserve Bank of San Francisco Economic Review*, 1996, 3, 34–48.

Winters, P. R., Forecasting Sales by Exponentially Weighted Moving Averages, *Management Science*, 1960, 6, 324–42.

John W. Birch, University of Wyoming, Laramie, WY 82071 or birch@uwyo.edu.

Mark A. Sunderman, University of Wyoming, Laramie, WY 82071 or sundermn@uwyo.edu.