

Estimating Property Values by Replication: An Alternative to the Traditional Grid and Regression Methods

Authors Tsong-Yue Lai, Kerry Vandell, Ko Wang, and Gerd Welke

Abstract This paper proposes and develops a replication method for estimating property values, in which optimal weights of comparable property attributes that best duplicate the subject property are determined. In a setting where the number of comparables is large compared to the number of attributes, replication weakly outperforms traditional general least squares regression by making use of potential correlations in the error structure. A similar result obtains in comparison to the grid method, which may suffer from subjective price adjustment factors. The replication method suggests using a large sample regression analysis to obtain the functional form of the error variance-covariance, and then replicating the subject with a smaller, attribute-close set of comparable properties.

The valuation of real estate assets differs dramatically from the valuation of financial assets. When applying a valuation technique such as the Capital Asset Pricing Model (CAPM) or Arbitrage Pricing Theory (APT) to financial assets, data availability is rarely a major consideration—especially for publicly-traded securities—and solutions are often closed form. Under the assumption of efficient markets, the analyst is interested simply in obtaining the price dynamic parameters such as betas for the individual asset relative to the market. These parameters are objectively available from reported market performance data and, in principle, identical for everyone. Thus, value estimates derived from application of the models do not vary across analysts, and there is little subjectivity in the valuation process.¹

The situation is different for real estate assets. Because of location boundedness, heterogeneous production, and non-continuous trading, the quantity and quality of data available might not allow individuals to use a model that offers a closed-form solution. For example, the regression method (or hedonic pricing model), which has been advocated by academics as an objective valuation method, has not received a warm welcome from practitioners because of both data availability

and specification issues. Instead, they prefer to use the grid method (or sales comparison technique) and are willing to tolerate its subjectivity. Why is this?

The regression method can be described as a “shopping-cart approach.” Each property attribute is thought of as having a price tag, and the total price of a subject property is determined by the quantity of each attribute that it possesses. To implement this technique, the analyst must decide upon a proper specification for the hedonic relationship, which property attributes to include, and then estimate the price adjustment (or regression coefficient, β) for each attribute. It is well-known that the estimation of the β s requires the use of a very large quantity of high-quality data, both in terms of number of properties and attributes. Moreover, the “true” hedonic relationships defining value can vary dramatically across time and location. Such data and specification requirements have prevented the regression technique from being widely used in practice.

The grid method, on the other hand, can be described as a “cut-and-paste” approach. Similar to the regression method, the grid method also assumes that each property attribute has a price tag. However, to implement the grid method, one first refines a set of recently transacted, smaller set of properties that have attributes similar to the subject property. A prospective buyer of the subject property will then sell off (or buy on) those comparable’s attributes s/he finds in excess (or deficit) of those of the subject property. Following this cut-and-paste procedure, the comparable property should be similar to the subject property in terms of attributes, and the price of the subject property (with the attributes the buyer desires) should be the price of the attribute-adjusted comparable property. For several comparable properties, the process may be repeated, and a suitably weighted average of the adjusted comparable prices represents the price that one is willing to pay for the subject property.

It is clear that since the grid method requires a smaller number of property comparables than the regression method, it is subject to greater sampling error. But this may be more than compensated for by the selection of comparables that are “closest” to the subject property in terms of location, age, or other characteristics. This should result in less variability, both in the price of attributes as well as in the “true” specification of the hedonic relationship. However, the analyst must still estimate both the weights for the comparables as well as the price of each attribute. A major drawback of this technique is that there is little guidance on how comparables, weights, and adjustment factors (prices) are to be selected. In other words, there must by necessity be substantial subjectivity in the valuation process.

Thus, when compared to the regression method, the benefit of the grid approach is that an analyst is able to select the comparables used for the evaluation more carefully. Because of the ability to screen in great detail, it is not surprising that the grid method is superior to the regression method under certain circumstances, particularly when properties used in the estimation are not really comparable to each other (leading to unreliably estimated β s), or when only few comparables sufficiently “similar” to the subject property exist.

The purpose of this paper is to develop a valuation methodology for real estate assets that reduces the subjectivity in the valuation process on the one hand, while accommodating the inherent data constraints present on the other. This technique is called the replication method. Specifically, the replication method makes use of large scale data to parameterize any variance-covariance structure in the error term, and then uses this information to improve an “attribute duplication” of the subject property from a smaller, more homogeneous set of comparable properties. Optimality is defined in terms of the weight vector of comparable property attributes that best duplicates the subject property’s attributes.

It should be noted that both the regression and the grid techniques also “duplicate” the subject property by adjusting for attribute differences. Both methods require an estimation of attribute prices, and an implicit or explicit assumption for the hedonic specification.² The proposed replication method differs from these in that it uses error correlations estimated on a large sample to improve on a final stage estimation, in which a smaller attribute-close data set is used. Intuitively, the replication technique retains the benefits of the regression and grid methods, while avoiding the problems associated with each method. Practically, it will be seen that the ability of the analyst to apply detailed knowledge of subject and comparable properties in the replication technique leads to (weakly) improved pricing over both the regression and grid methods.

Section 2 introduces the replication method and provides a theoretical foundation for the approach. Section 3 details the estimation issues. Sections 4 and 5 compare the statistical properties of the replication method with those of the regression and grid methods, respectively. The findings demonstrate that the proposed replication method generally performs better than the regression method, but no worse, and better than the grid method under realistic scenarios. Implementation issues of the replication method are discussed in Section 6. Section 7 concludes.

The Replication Method

The proposed replication method can be demonstrated in two simple steps. The first is to solve for a $1 \times n$ weight vector of comparable properties ω_p such that:

$$\omega_p X_C = X_S, \quad (1)$$

where X_C is an $n \times k$ matrix representing the k property attributes for each of the n comparable properties used in the valuation assignment, and X_S is a $1 \times k$ vector representing the k subject property’s attributes.³ It should be noted that when compared to the grid method, which requires the sum of the weights be equal to unity, there are no restrictions on ω_p for the replication method. Simply put, at this stage, the task of the analyst is to solve a simultaneous set of equations for

the weight vector ω_{rp} that best duplicates⁴ the attributes of the subject property X_S using the attributes of comparable properties X_C .

Once ω_{rp} is determined, the value of the subject property follows from the sales prices of the comparable properties:

$$\hat{V}_{S(rp)} = \omega_{rp} V_C, \quad (2)$$

where V_C is an $n \times 1$ vector representing the sales prices of the n comparable properties and ω_{rp} is the weight vector derived from Equation (1). The estimated value of the subject property using the replication method is defined as $\hat{V}_{S(rp)}$.

In a world of certainty, the replication method is based on one simple assumption, *viz.*, that for sufficiently comparable properties:

$$V = X\beta, \quad (3)$$

stating that the property values can be estimated without error to order $(X - X_0)^2$ by a linear combination of all relevant property attributes and the prices of these attributes. Equation (3) is the first term in a Taylor expansion of $V(X)$ about some “average” property attribute set X_0 , and holds if the quadratic term in $(X - X_0)$ is sufficiently small, *i.e.*, if the properties are sufficiently comparable in all relevant properties.⁵

Under the certainty case, Equation (3) holds for both comparable and subject properties, and given ω_{rp} satisfying Equation (1), Equation (2) can be re-written as:

$$\hat{V}_{S(rp)} = \omega_{rp} V_C = \omega_{rp} X_C \beta = X_S \beta = V_S. \quad (4)$$

Thus, as long as the assumption in Equation (3) holds, the true value of the subject property can always be estimated by the replication method without error. Clearly, this is the most desirable condition possible for any valuation technique under certainty. The method also requires a comparable property that is sufficiently close in attribute space, or the second order term $(X - X_0)V''|_{X_0}(X - X_0)$ is no longer negligible.

In practice, there are two issues to consider. First, in practice, uncertainty exists and property values are estimated with error. This will generally lead to multiple weight vectors ω_{rp} satisfying Equation (1), and the estimated property value based

on Equation (2) will not be unique so that the final valuation is subject to error. This does not necessarily render the proposed method unattractive since other valuation techniques are subject to similar error. Consequently, the relative usefulness of replication will have to be judged relative to regression and the grid method based on established statistical criteria. Later it will be seen that the replication method produces a weakly lower variance in the prediction error for the subject properties value when compared to the traditional techniques.

The second practical issue is that of data constraints. Regression requires a relatively homogeneous data set of comparables that are “close” in attribute space, for the Taylor approximation in Equation (3) to be valid. Given uncertainty, choosing comparables is therefore a trade-off between reducing sampling error by increasing the data set’s cardinality, versus decreasing it to minimize the specification error in Equation (3). In the proposed replication technique, using a large data set regression to estimate a functional form of the variance-covariance of the error is suggested, which is then applied in a second stage replication with a smaller, attribute-close set of comparables.

Estimation

Equation (1) is a system of k simultaneous linear equations for the n unknown components of the attribute weight vector ω_{rp} . The conditions for a unique solution to these equations are that $k = n$ (an equal number of property attributes and comparables), and that X_C be full rank. If there are more property attributes than comparable properties ($k > n$), there is no guarantee that a weight vector can be found that satisfies Equation (1), while if there are more comparable properties than property attributes ($k < n$), there will be an infinite number of possible weight vectors.⁶ In the unique case in which the system is just identified, the analyst simply selects the singular solution as the optimal weight vector. In the two other cases ($k > n$ and $k < n$), the analyst will have to select an optimal weight vector ω_{rp} from the set of alternatives.

Standard Case: When $k < n$

From Equations (1) and (2), for every attribute weight vector ω_{rp} that satisfies Equation (1), one can obtain a value estimate for the subject property using the relationship $V_{S(rp)} = \omega_{rp}V_C$. For the case $k < n$, there is an infinite number of such weight vectors. The analyst therefore has to impose additional criteria to select the optimal weights ω_{rp} . The unbiasedness and minimum variance are proposed as the criteria.

To implement these criteria and to make the comparison with the regression and grid methods meaningful, it is assumed that the “true” values, $V^T = [V_C^T, V_S^T]$,⁷ of the comparable and subject properties can be estimated as a linear function of their attributes X :

$$V = X\beta + \varepsilon, \quad (5)$$

where X is the $(n + 1) \times k$ stacked matrix of comparable and subject property attributes, X_C ($n \times k$) and X_S ($1 \times k$), respectively. The vector β is the $k \times 1$ set of unknown price parameters, and $\varepsilon^T = [\varepsilon_C^T, \varepsilon_S^T]$ is an $(n + 1) \times 1$ vector of error terms. Explicitly separating the comparable and subject properties:

$$V_C = X_C\beta + \varepsilon_C \text{ and} \quad (6)$$

$$V_S = X_S\beta + \varepsilon_S. \quad (7)$$

It is also assumed that the error terms, ε , are zero mean.

For given weight vector ω_{rp} , the prediction error for subject property's value under the replication method, ε_{rp} , is the difference between its true, unknown value, V_S given in Equation (7) and its estimated value, $\hat{V}_{S(rp)} = \omega_{rp}V_C$. Combining Equations (2), (6), and (7):

$$\begin{aligned} \varepsilon_{rp} &\equiv V_S - \hat{V}_{S(rp)} \\ &= X_S\beta + \varepsilon_S - \omega_{rp}(X_C\beta + \varepsilon_C) \\ &= \varepsilon_S - \omega_{rp}\varepsilon_C. \end{aligned} \quad (8)$$

Note that Equation (8) implicitly depends on the price adjustment factors, but, as will be seen, this formulation allows for the possibility of a correlation between the subject and comparable estimation errors.

Given the assumption of zero mean error terms, the expectation of the prediction error is also zero so that replication estimators of the true property value are unbiased:

$$E[\varepsilon_{rp}] = E[\varepsilon_S - \omega_{rp}\varepsilon_C] = 0, \quad (9)$$

where $E[\cdot]$ is the expectations operator. This result is summarized in the following proposition.

Proposition 1. Under the assumption that the true property value is a linear function of the property attributes with a zero mean error term, as specified in

Equation (5), all replication estimators satisfying Equations (1) and (2) are unbiased.

Since any replication estimator using an acceptable weight vector satisfying Equation (1) is unbiased, the minimum variance criteria is then examined in order to select the optimal weight vector, i.e., the expected squared prediction error is minimized:

$$\begin{aligned} \min_{\omega_{rp}} E[\varepsilon_{rp}^2] &= \min_{\omega_{rp}} E[(\varepsilon_S - \omega_{rp}\varepsilon_C)^2] \\ &= \min_{\omega_{rp}} [\sigma_S^2 - 2\Omega_{CS}\omega_{rp}^T + \omega_{rp}\Omega\omega_{rp}^T] \end{aligned} \tag{10}$$

subject to:

$$\omega_{rp}X_C = X_S.$$

Here, σ_S^2 is the variance of ε_S , while:

$$\Omega = E[\varepsilon_C\varepsilon_C^T] \tag{11}$$

is the $n \times n$ variance-covariance matrix of ε_C , and:

$$\Omega_{CS} = E[\varepsilon_S\varepsilon_C^T] \tag{12}$$

is the $1 \times n$ covariance vector between ε_C and ε_S . The optimal solution to Equation (10), given the assumption of zero mean error terms, is:

$$\omega_{rp}^* = \Omega_{CS}\Omega^{-1} + [X_S - \Omega_{CS}\Omega^{-1}X_C](X_C^T\Omega^{-1}X_C)^{-1}X_C^T\Omega^{-1}. \tag{13}$$

Equation (13) is the necessary and sufficient condition that solves Equation (10) since Ω is a positive definite matrix in quadratic mathematical programming. With this choice of ω_{rp} , the value of the subject property is therefore:

$$\begin{aligned}
\hat{V}_{S(rp)} &= \omega_{rp}^* V_C \\
&= \Omega_{CS} \Omega^{-1} V_C + [X_S - \Omega_{CS} \Omega^{-1} X_C] \\
&\quad \times (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1} V_C.
\end{aligned} \tag{14}$$

Defining $\sigma_{(rp)}^2$ as the variance of the prediction error in the replication method, and combining Equation (8) with Equation (13):

$$\begin{aligned}
\sigma_{(rp)}^2 &= \sigma_S^2 + X_S (X_C^T \Omega^{-1} X_C)^{-1} X_S^T - 2X_S (X_C^T \Omega^{-1} X_C)^{-1} (X_C^T \Omega^{-1} \Omega_{CS}^T) \\
&\quad - \Omega_{CS} [\Omega^{-1} - \Omega^{-1} X_C (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1}] \Omega_{CS}^T.
\end{aligned} \tag{15}$$

A comparison of the replication technique to the regression and grid methods of valuation shows that the variance of the prediction error in replication is (weakly) smaller than that of the two traditional techniques, as seen later.

Although Equation (14) provides a general estimation of the property value, the covariance terms in Equation (14) might be difficult to estimate in practice.⁸ In a simpler approach, one might assume that the distributions of error terms ε_C and ε_S are *iid* with zero mean and standard deviation σ , i.e., $\Omega = \sigma^2 I$, $\sigma_S = \sigma$ and Ω_{CS} a null vector. In this case, Equation (10) may be rewritten as:

$$\min_{\omega_{rp}} [\sigma^2 + E(\omega_{rp} \varepsilon_C)^2] = \sigma^2 \min_{\omega_{rp}} [1 + (\omega_{rp} \omega_{rp}^T)] \tag{16}$$

subject to:

$$\omega_{rp} X_C = X_S.$$

The optimal solution to Equation (16) can be obtained from Equation (13):

$$\omega_{(rp,iid)}^* = X_S (X_C^T X_C)^{-1} X_C^T \tag{17}$$

and the estimated value of the subject property then becomes:

$$\hat{V}_{S(rp,iid)} = \omega_{(rp,iid)}^* V_C = X_S(X_C^T X_C)^{-1} X_C^T V_C. \tag{18}$$

It has been shown that, when ε_C is iid and uncorrelated with ε_S , Equation (14) collapses to Equation (18). Defining $\sigma_{(rp,iid)}^{2*}$ as the minimum variance of the prediction errors by the replication method under the iid assumption, observe from Equation (15) with $\Omega = \sigma^2 I$ and $\Omega_{CS} = 0$ that:

$$\begin{aligned} \sigma_{(rp,iid)}^{2*} &= \sigma^2 + [\omega_{(rp,iid)}^* \omega_{(rp,iid)}^{*T}] \sigma^2 \\ &= \sigma^2 + \sigma^2 X_S(X_C^T X_C)^{-1} X_C^T X_C(X_C^T X_C)^{-1} X_S^T \\ &= \sigma^2 [1 + X_S(X_C^T X_C)^{-1} X_S^T]. \end{aligned} \tag{19}$$

Also note that Equations (18) and (19) correspond to the standard OLS regression results. The findings of this sub-section are summarized in the following proposition.

Proposition 2. Assume that the true property values are a linear function of their property attributes, as specified in Equation (5), and zero mean error structure as given in Equations (11) and (12). If the number of comparable properties is more than the number of property attributes used in the valuation, the optimal weight set of the replication estimator is given by Equation (13):

$$\omega_{rp}^* = \Omega_{CS} \Omega^{-1} + [X_S - \Omega_{CS} \Omega^{-1} X_C] (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1},$$

the estimated value of the subject property is given by Equation (14):

$$\hat{V}_{S(rp)} = \Omega_{CS} \Omega^{-1} V_C + [X_S - \Omega_{CS} \Omega^{-1} X_C] \times (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1} V_C.$$

and the (minimum) variance of the prediction error of the replication estimator is given by Equation (15):

$$\begin{aligned} \sigma_{(rp)}^2 &= \sigma_S^2 + X_S(X_C^T \Omega^{-1} X_C)^{-1} X_S^T - 2X_S(X_C^T \Omega^{-1} X_C)^{-1} (X_C^T \Omega^{-1} \Omega_{CS}^T) \\ &\quad - \Omega_{CS} [\Omega^{-1} - \Omega^{-1} X_C(X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1}] \Omega_{CS}^T. \end{aligned}$$

In the case of *iid* error terms, $\Omega = \sigma^2 I_n$, $\sigma_S = \sigma$ and $\Omega_{CS} = 0$, the corresponding replication method results collapse to the standard OLS results:

$$\begin{aligned}\omega_{rp,iid}^* &= X_S(X_C^T X_C)^{-1} X_C^T, \\ \hat{V}_{S(rp,iid)} &= X_S(X_C^T X_C)^{-1} X_C^T V_C, \\ \sigma_{rp,iid}^{2*} &= \sigma^2 [1 + \sigma^2 X_S(X_C^T X_C)^{-1} X_S^T].\end{aligned}$$

Special Cases: When $k = n$ and $k > n$

When the number of properties equals the number of property attributes ($n = k$) and X_C is a full rank matrix, there exists only a unique solution for Equation (1). Under this circumstance, the analyst would simply use the given weight vector to derive the value of the subject property. When there are more property attributes (equations) than the number of comparable properties (variables) (i.e., when $k > n$), however, there is no solution to Equation (1) in general.⁹ In other words, there is no guarantee that one is able to replicate an arbitrary subject property using the limited number of comparable properties.¹⁰

While $k > n$ implies that there is no exact solution for Equation (1), the analyst can still solve for an optimal weight vector for estimating property values using an appropriate objective function. The idea is to synthesize a property (using the limited comparable attributes) that is most similar to the subject property's attributes. While the subject attributes cannot be replicated exactly, a minimum sum of squared errors optimality criterion can be used to come close to them:

$$\min_{\omega_{rp}} (X_S - \omega_{rp} X_C)(X_S - \omega_{rp} X_C)^T. \quad (20)$$

The unique set of weights that solves Equation (20) is:

$$\omega_{rp}^* = X_S X_C^T (X_C X_C^T)^{-1}, \quad (21)$$

and gives the value of the subject property as:

$$\hat{V}_{S(rp)} = \omega_{rp}^* V_C = X_S X_C^T (X_C X_C^T)^{-1} V_C. \quad (22)$$

Although ω_{rp}^* in Equation (21) is not an exact solution to Equation (1), $\omega_{rp}^* X_C$ best “mimics” the attributes of the subject property in the sense of minimum mean squared error. In other words, while an exact answer is not possible because of the constrained number of comparables, the assignment becomes one of identifying the “best” answer as the solution to the valuation problem.¹¹

Comparison Issues

While all three possible scenarios ($k > n$, $k = n$, and $k < n$) have been discussed in this section, the scenario $k < n$ is most relevant from the standpoint of comparisons between the replication method and the traditional regression and grid methods. Clearly, when the number of comparables is less than or equal to the number of property attributes ($n \leq k$), it is not possible to estimate hedonic price vector $\hat{\beta}$ via regression. Since $\hat{\beta}$ cannot be estimated, it is not feasible to apply either the regression method or the grid method.¹² While the replication method might be used, it is not possible to ascertain how the estimator performs when compared to the other two methods. Given these issues, the following sections will concentrate on the case of $k < n$.

Comparison to the Regression Method

In this section, the replication estimator is compared with the estimators derived from the traditional regression method under the same set of conditions. The comparison to the grid method follows in the next section.

A maximum likelihood estimation of the hedonic prices or adjustment factors of property attributes, β , on the comparable properties, Equation (6), leads to the familiar generalized least squares (GLS) estimator:

$$\hat{\beta}_{GLS} = (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1} V_C, \tag{23}$$

where X_C and V_C are the comparable property attributes and values, respectively, and the variance-covariance matrix of errors, Ω , is defined in Equation (11). The estimator $\hat{\beta}_{GLS}$ is unbiased,¹³ and has error variance:

$$E[(\beta - \hat{\beta}_{GLS})(\beta - \hat{\beta}_{GLS})^T] = (X_C^T \Omega^{-1} X_C)^{-1}. \tag{24}$$

From Equation (7), the GLS regression estimate of the subject property value is:

$$\hat{V}_{s, GLS} = X_s \hat{\beta}_{GLS} = X_s (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1} V_C. \tag{25}$$

This result collapses to the standard ordinary least squares (OLS) estimate of the subject value in the case that the errors in the comparable valuation are homoscedastic, $\Omega = \sigma I_n$.¹⁴

$$\hat{V}_{S,OLS} = X_S \hat{\beta}_{OLS} = X_S (X_C^T X_C)^{-1} X_C^T V_C. \quad (26)$$

A comparison of Equations (25) and (26) with the replication estimator, Equation (14), shows that \hat{V}_{rp} coincides with the GLS estimator $\hat{V}_{S,GLS}$ if the correlation between the error in the subject and comparable property values is zero ($\Omega_{CS} = 0$), and with the OLS estimator $\hat{V}_{S,OLS}$ if both $\Omega_{CS} = 0$ and $\Omega = \sigma I_n$ hold (i.e., the errors are *iid*).

The variance of the GLS prediction error for the subject property is:

$$\sigma_{GLS}^2 = E[(V_S - \hat{V}_{S,GLS})^2] = E[(X_S(\beta - \hat{\beta}_{GLS}) + \varepsilon_s)^2], \quad (27)$$

which may be evaluated as:¹⁵

$$\begin{aligned} \sigma_{GLS}^2 &= \sigma_s^2 + X_S (X_C^T \Omega^{-1} X_C)^{-1} X_S^T \\ &\quad - 2X_S (X_C^T \Omega^{-1} X_C)^{-1} X_C^T \Omega^{-1} \Omega_{CS}^T. \end{aligned} \quad (28)$$

Note that for *iid* error terms, $\Omega = \sigma^2 I_n$ and $\Omega_{CS} = 0$, Equation (28) collapses to the ordinary least squares (OLS) result:

$$\sigma_{OLS}^2 = \sigma^2 [1 + X_S (X_C^T X_C)^{-1} X_S^T]. \quad (29)$$

which is identical to the replication error variance under the *iid* assumption, Equation (19).

How does the variance of the GLS prediction error σ_{GLS} (Equation (28)) compare to that of the replication technique, σ_{rp} (Equation (15))? First, note that Ω is positive definite and symmetric, so that Ω^{-1} is positive definite and symmetric also, and has a Cholesky decomposition $\Omega^{-1} = P^T P$. Then, after some algebra, the difference between Equations (15) and (28) may be shown to have the quadratic form:

$$\sigma_{GLS}^2 - \sigma_{rp}^2 = x^T B x, \tag{30}$$

where:

$$x = P \Omega_{CS}^T, n \times 1, \tag{31}$$

$$B = I_n - P X_C (X_C^T \Omega^{-1} X_C)^{-1} X_C^T P^T, n \times n. \tag{32}$$

Since B is idempotent¹⁶ and symmetric:¹⁷

$$B = D \Lambda D^T, \tag{33}$$

where D is orthogonal ($DD^T = I_n$) and Λ is a diagonal matrix of ones and/or zeros.¹⁸ Therefore:

$$\sigma_{GLS}^2 - \sigma_{rp}^2 = x^T D \Lambda D^T x \equiv y^T \Lambda y = \sum_{i=1}^J y_i^2 \geq 0, \tag{34}$$

where $y = D^T x = D^T P \Omega_{CS}^T$ and $0 \leq J \leq n$ is the rank of B. Thus, $\sigma_{rp} \leq \sigma_{GLS}$, as claimed earlier.

A strict improvement in the prediction error follows if (a) y is non-zero, and (b) $J > 0$. Several cases are outlined in which the equality of prediction error variance obtains:

1. $n = k$, the number of property attributes equals the comparable property count: In this case, $J = 0$ and all eigenvalues of B are zero. This follows because $\text{tr} (P X_C (X_C^T \Omega^{-1} X_C)^{-1} X_C^T P^T) = \text{tr} (X_C^T \Omega^{-1} X_C)^{-1} (X_C^T P^T P X_C) = \text{tr} I_n = n$. Therefore $\text{tr} B = 0$, so that $\text{rank} B = J = 0$, and the RHS of Equation (34) is identically zero.
2. $\Omega_{CS} = 0$, the subject and comparable property errors are uncorrelated: In this case, the vector y in Equation (34) is the zero vector. Notice that in this case the replication estimator of the property value, Equation (14), also coincides with the GLS estimator, Equation (25).
3. $\Omega_{CS} = 0$ and $\Omega = \sigma^2 I_n$: if additionally the comparables' error structure is homoscedastic, replication and OLS (Equations (26) and (29)) results are identical.

An estimation of the comparables variance-covariance structure, Ω , is typically accomplished relatively easily using, say, the standard White estimator. The correlation between the subject and comparables' error terms, Ω_{CS} , on the other hand, requires some further discussion. The issue, of course, is that the subject valuation error is unknown.

In essence, the replication method relies on a standard regression estimation being improved upon by adding knowledge of the covariance between the subject and comparable error terms, *viz.*, Ω_{CS} , Equation (12). In the standard spatial analysis, this may be accomplished by estimating a variogram model on the relative spatial distribution of the errors in the comparables set (see, for example, Isaaks and Srivastava, 1990). In the simplest case, the model would depend on only the distance between clusters of similar homes (isotropy), but more generally would depend on the analyst's detailed knowledge of the subject area. Additionally, the method naturally lends itself to estimating error covariances in age differences; the underlying rationale is that construction features and functionality of structures built at the same time might produce similar price deviations about the hedonic pricing mean. More generally, a continuously labeled analysis of the error term may be extended to other characteristics, which may be discrete¹⁹ or quasi-continuous.²⁰

When compared to a standard regression analysis, therefore, the replication method is preferable if the analyst finds a statistically significant correlation in the pricing errors, particularly in the dimension of location and age, and in products of these with discrete property characteristics. Note, however, that such correlations may only be discernable if the comparables data set is sufficiently large. In applying replication, the analyst would therefore extract the structure from a first stage regression on the larger, attribute-heterogeneous sales data. In the second stage, replication incorporates this information Ω_{CS} to reduce the sampling error, as shown in this section and summarized by the following proposition:

Proposition 3. The mean squared prediction error for the subject property value using the replication method, $\sigma_{(rp)}^2$ (Equation (15)), is weakly smaller than corresponding prediction error from a standard regression estimate, σ_{GLS}^2 (Equation (28)). Equality obtains if the number of attributes k equals the number of comparable properties, or if the subject property's error is uncorrelated with the comparables' error term ($\Omega_{CS} = 0$). With homoscedasticity ($\Omega = \sigma^2 I_{n+1}$), the replication method collapses to a standard OLS regression.

Comparison to the Grid Method

The grid method also uses a weighted average of the observed comparable property prices (after adjusting for differences in property attributes) to estimate the value of a subject property. Defining $V_{S(grid)}$ as the estimated value of the subject property using the grid method, the estimation equation of the grid method can be written as:

$$\hat{V}_{S(grid)} = \omega_{grid}[V_C + (eX_S - X_C)\hat{\beta}] \tag{35}$$

where:

$$\omega_{grid}e = 1, \tag{36}$$

and e is an $n \times 1$ vector of ones,²¹ ω_{grid} is the $1 \times n$ weight vector derived under the grid method, and $\hat{\beta}$ is the adjustment factor (i.e., the hedonic prices of the property attributes). When implementing the grid method, an analyst’s assignment is to select the appropriate number of comparables n and to estimate both ω_{grid} and $\hat{\beta}$. The intuition for Equation (35) is clear: The analyst adjusts the price of each comparable by the difference between its attributes and the subject’s, weighted by the price of the corresponding attribute. As a final step, a weighted average of adjusted comparable prices is used to estimate the subject property’s value.

While there are several methods used to guide the estimation of the weight vector ω_{grid} ,²² it is commonly accepted that the hedonic coefficients derived from the regression method present a defensible case for representing the “best” unbiased estimates of the adjustment factors $\hat{\beta}$.

Under the same assumption that the true property value is a linear function of its property attributes with a zero mean error term (see Equation (5)), the prediction error of a grid estimator ε_{grid} can be specified as:

$$\begin{aligned} \varepsilon_{grid} &= V_S - \hat{V}_{S(grid)} \\ &= (X_S - \omega_{grid}X_C)(\beta - \hat{\beta}) + \varepsilon_S - \omega_{grid}\varepsilon_C, \end{aligned} \tag{37}$$

where Equations (7), (35), and constraint (36) have been used. Using the GLS estimator for β , Equations (23) and (24), tedious but straightforward algebra leads to the following expression for the variance of the prediction error, Equation (37):

$$\begin{aligned} \sigma_{grid}^2 = E[\varepsilon_{grid}^2] &= \sigma_S^2 + X_S(X_C^T\Omega^{-1}X_C)^{-1}X_S^T \\ &\quad + \omega_{grid}[\Omega - X_C(X_C^T\Omega^{-1}X_C)^{-1}X_C^T]\omega_{grid}^T \\ &\quad - 2\omega_{grid}[\Omega - X_C(X_C^T\Omega^{-1}X_C)^{-1}X_C^T]\Omega^{-1}\Omega_{CS}^T \\ &\quad - 2X_S(X_C^T\Omega^{-1}X_C)^{-1}X_C^T\Omega^{-1}\Omega_{CS}^T, \end{aligned} \tag{38}$$

In analogy to the discussion leading up to Equation (30), the difference between the variance of the prediction error of the grid estimator (Equation (38)) and the corresponding value for the replication method (Equation (15)) can be written as a quadratic form:

$$\sigma_{grid}^2 - \sigma_{rp}^2 = x'^T B x', \quad (39)$$

where B is given by Equation (32), but now:

$$x' = P(\Omega_{CS} - \omega_{grid}\Omega)^T. \quad (40)$$

It can therefore be concluded that $\sigma_{rp}^2 \leq \sigma_{grid}^2$, with equality holding when $n = k$. On the other hand, equality now no longer follows if $\Omega_{CS} = 0$, and/or $\Omega = I_n \sigma^2$: For unbiased estimates, $\omega_{grid}e = 1$ and $\Omega_{CS} - \omega_{grid}\Omega = 0$, i.e., only in the degenerate case that $\Omega_{CS}\Omega^{-1}e = 1$ will the prediction error variances coincide. Thus, the preference for the replication method versus the grid method is reinforced if the unbiasedness criterion is taken into consideration.

The following proposition summarizes the findings of this section.

Proposition 4. Under the assumption that the true property values are a linear function of their property attributes, the variance of the prediction error derived from the replication method is less than that derived from the grid method, as long as the number of comparable properties n exceeds the number of attributes k used in the valuation. Equality holds if $n = k$ or $\Omega_{CS}\Omega^{-1}e = 1$.

Conclusion

All appraisal methods are established with certain assumptions, some of which hold better than others in a practical setting. Consequently, the proper use of appraisal methods should depend on the nature of the individual appraisal assignment. Given this, when is it best for the analyst to use the replication method?

In practice, an attribute-close comparable data set typically has low cardinality, leading to large sampling errors. Using a large data set, on the other hand, may lead to a breakdown of the linear hedonic specification. A trade-off between sampling error and using a valid specification thus seems inevitable, but traditional regression would seem to work best when a large number of comparable properties are attribute-homogeneous; however, such a large, high quality data set is rarely

available in practice. The grid method, on the other hand, results in smaller estimation errors when the market is more heterogeneous with a possibly uncertain set of amenity variables. The analyst sifts sales for attribute-close comparables and is able to bypass specification errors by using this (smaller) data set. However, under uncertainty, the method suffers from larger sampling errors when compared to regression. Additionally, in practical applications of the grid method, there may be significant subjectivity in the choice of adjustment factors, leaving open the possibility of systematic over- or undervaluation of the subject property.

The replication technique presented here proposes reducing specification issues by duplicating the subject property using a relatively small attribute-close set of comparable properties, but reducing the attendant sampling error by including the error term's covariance structure, as adduced from the larger, more heterogeneous sales data.

When the replication method is used and analysts are not confident of the true model (or when there is a probability of including irrelevant—or omitting relevant—variables in the estimation equation), the best course may be to include more (rather than less) property attributes in the estimation equation (as long as $n > k$). For example, when in doubt, analysts could add the squares or the cross products of property attributes to the model characteristic set. In general, adding such terms will result in a better prediction than from solely linear models. It seems that the cost of including irrelevant variables (with a higher variance) should be much less than the cost of omitting relevant variables (with biased estimators). Given this assumed loss function, it is preferable that analysts use all the reasonable property attributes they can find in the valuation process.

The development of techniques that can remove (or, at least, substantially reduce) subjectivity in the appraisal process is needed in valuing real estate assets. Without such a method, real estate valuation will always remain an art, subject too frequently to ad hoc assumptions, rather than a precise methodology. At the same time, it is also understood that the methods employed for valuing real estate assets must differ from those used for financial assets because of heterogeneity and data limitations.

This paper develops a real estate valuation model that substantially reduces the degree of subjectivity in the valuation process and performs well under a realistic set of data constraints. It is also shown that the proposed replication method should perform better than the traditional grid and regression methods under most scenarios. In the worst case, the replication method performs identically well when compared to the other two methods. However, while this model seems to perform well on theoretical grounds, it is important to evaluate its performance under actual field conditions. Given this, the next logical step is an empirical study that compares the performance among the three methods (replication, grid, and regression) using a large data set.

Endnotes

- ¹ Of course, the selection of input parameters still depends upon individual judgments.
- ² Hedonic estimation is explicit in the case of the regression method. In the case of the grid method, it is assumed either (1) a hedonic estimation has been undertaken in the background to produce the adjustment factors (an assumption made throughout permit comparisons among the methodologies), or (2) that some other statistical approach, such as matched pairs, the cost approach, etc., has been employed. In both cases, the adjustment factors (prices) are estimated with error.
- ³ The subscript “rp” denotes the “replication” method. For any $1 \times n$ vector G and any $k \times n$ matrix Φ with the same rank k as X_S , all solutions for Equation (1) must be in the form of $\omega_{rp} = G + [X_S - GX_C](\Phi X_C)^{-1}\Phi$. For example, Equation (13) sets $G = \Omega_{CS}\Omega^{-1}$ and $\Phi = X_C^T\Omega^{-1}$.
- ⁴ If $k > n$.
- ⁵ The zero-order term $V(X_0)$ is included as a constant term on the RHS.
- ⁶ Since every set of k comparables can produce an exact solution, there will be $n!/[(n - k)!k!]$ weight vectors that can satisfy Equation (1) if there are n comparables. For example, if there are five equations and four unknowns, there will be five possible solutions when solving this system of equations. Since linear combinations of those $n!/[(n - k)!k!]$ weight vectors also satisfy Equation (1), the number of solutions are infinite. To see this, let ω_1 and ω_2 be the two solutions satisfying Equation (1). Then any convex combination $\alpha\omega_1 + (1 - \alpha)\omega_2$ is also a solution because $[\alpha\omega_1 + (1 - \alpha)\omega_2]X_C = \alpha X_S + (1 - \alpha)X_S = X_S$.
- ⁷ A superscript “T” is used to denote the transpose of a vector or matrix.
- ⁸ The covariance terms in Equation (14) are difficult to estimate without a sufficient number of comparables. This is particularly true of the term Ω_{CS} . Note, however, that if and when the covariance terms can be estimated without difficulty, the property value estimator derived from the replication method (via Equation (13)) is unbiased and with minimum variance.
- ⁹ Unless a sufficiently large set of unknowns are linearly dependent.
- ¹⁰ While it is possible that in specific situations when $k > n$ the subject property could be duplicated, the solution would then be the same as that discussed in the $k < n$ or $k = n$ scenarios. Therefore, this possibility is ignored.
- ¹¹ Applying the arbitrage pricing model (APT) to real data, Huberman, Kandel, and Stambaugh (1987) demonstrate that it is feasible to use mimicking portfolios to substitute for the k factors in an exact arbitrage pricing model. Ang and Lai (1998) show that the factor loading in the APT can be replicated, and thus there are $k + 1$ risk premia under the k -factor model in a finite economy. In this regard, the approach here is similar to their approach.
- ¹² It might be feasible to use other subjective methods to derive the adjustment factors for the grid method. However, in such a case, meaningful comparisons with the regression or the replication methods are not possible, since such subjective methods would not be expected to contribute to a closed-form solution in general.
- ¹³ Provided there are no specification issues.
- ¹⁴ I_n denotes the $n \times n$ identity matrix.

- ¹⁵ Using Equations (6), (7), (11), (12), and (23) through (25), along with $E[\varepsilon_s] = 0$ and $E[\varepsilon_s^2] = \sigma_s^2$.
- ¹⁶ Since, $[PX_C(X_C^T\Omega^{-1}X_C)^{-1}X_C^TP^T] [PX_C(X_C^T\Omega^{-1}X_C)^{-1}X_C^TP^T] = PX_C(X_C^T\Omega^{-1}X_C)^{-1}(X_C^T\Omega^{-1}X_C)(X_C^T\Omega^{-1}X_C)^{-1}X_C^TP^T = PX_C(X_C^T\Omega^{-1}X_C)^{-1}X_C^TP^T$, so that B is idempotent also.
- ¹⁷ Since $(XY)^T = Y^TX^T$ and $(X^T)^{-1} = (X^{-1})^T$.
- ¹⁸ Since D orthogonal and B idempotent, $B^2 = DAD^TDAD = DA^2D^T = B = DAD^T$. Thus $\Lambda^2 = \Lambda$, implying that each eigenvalue λ_i is 1 or 0. Note that the λ_i are real since B is symmetric.
- ¹⁹ For example, construction style: “ranch,” “colonial,” etc.
- ²⁰ For example, an analyst observed “quality scale.”
- ²¹ In order to obtain an unbiased estimator of the property value, it is necessary to restrict $\omega_{grid}e = 1$. It should be noted, however, that the elements of $\omega_{grid}e$ need not be restricted to be positive.
- ²² Practitioners insist that one should give more weight to the “most comparable” properties, but they have a difficult time defining what are the most comparable properties, or giving a precise method of determining the weights. There is a large body of academic literature that derives sets of optimal weights under specific conditions, but those rules are not applied broadly by practitioners [see, for example, Colwell, Canaday, and Wu (1983), Isakson (1986, 2002), Vandell (1991), Gau, Lai, and Wang (1992, 1994), Green (1994), Pace and Gilley (1997, 1998), and Pace (1998) for alternative estimation methods for ω_{grid}].

References

- Ang, J. and T-Y. Lai. Mutual Funds and Asset Pricing Models in a Finite Economy. *Research in Finance*, 1998, 16, 107–25.
- Colwell, P., R. Canaday, and C. Wu. The Analytical Foundations of Adjustment Grid Methods. *Journal of the American Real Estate and Urban Economics Association*, 1983, 11, 11–29.
- Gau, G., T-Y. Lai, and K. Wang. Optimal Comparable Selection and Weighting in Real Property Valuation: An Extension. *Journal of the American Real Estate and Urban Economics Association*, 1992, 20, 107–23.
- . A Further Discussion of Optimal Comparable Selection and Weighting, and A Response to Green. *Journal of the American Real Estate and Urban Economics Association*, 1994, 22, 655–63.
- Green, R. Optimal Comparable Weighting and Selection: A Comment. *Journal of the American Real Estate and Urban Economics Association*, 1994, 22, 647–54.
- Issaks, E.H. and R.M. Srivastava. *An Introduction to Applied Geostatistics*. New York: Oxford University Press, 1990.
- Isakson, H. The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Method. *Journal of the American Real Estate and Urban Economics Association*, 1986, 14, 274–86.
- . The Linear Algebra of the Sales Comparison Approach. *Journal of Real Estate Research*, 2002, 24, 117–28.
- Huberman, G., S. Kandel, and R.F. Stambaugh. Mimicking Portfolios and Exact Arbitrage Pricing. *Journal of Finance*, 1987, 42, 1–9.

Pace, K. Total Grid Estimation. *Journal of Real Estate Research*, 1998, 15, 101–14.

Pace, K. and O. Gilley. Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics*, 1997, 14, 333–40.

_____. Optimally Combining OLS and the Grid Estimator. *Real Estate Economics*, 1998, 26, 331–47.

Vandell, K. Optimal Comparable Selection and Weighting in Real Property Valuation. *Journal of the American Real Estate and Urban Economics Association*, 1991, 19, 213–39.

Tsong-Yue Lai, California State University, Fullerton, CA 92834 or tylai@fullerton.edu.

Kerry Vandell, University of California, Irvine, CA 92697 or kvandell@merage.uci.edu.

Ko Wang, Baruch College, New York, NY 10010 or ko.wang@baruch.cuny.edu.

Gerd Welke, Baruch College, New York, NY 10010 or gerd.welke@baruch.cuny.edu.