Multiple Transactions Model: A Panel Data Approach to Estimate Housing Market Indices

Authors Andre H. Gao and George H.K. Wang

Abstract

In this paper, a multiple transactions model with a panel data approach is used to estimate housing market indices. The multiple transactions model keeps the same features of the repeat transactions index model (i.e., tracking the price appreciation of same houses). However, the multiple transactions model overcomes the shortcomings of the repeat transactions model by avoiding the correlated error terms. The indicative empirical analysis on a small sample of actual house transaction data demonstrates that the proposed multiple transactions model is superior to the repeat transactions model in terms of index variance, robustness of estimate, index revision volatility, and out-of-sample prediction of individual house prices.

Both the price levels and trends (or appreciations) of housing markets are important to homebuyers, builders, and mortgage lenders. Among several available methods for monitoring the housing markets, the median house price and repeat transactions index methods are the most widely used.[1] The median house price method provides information on the price levels of housing markets, but can yield unreliable market trends because the quality of houses is not controlled in this method. The possibility of having misleading housing market trends limits the application of the median house price index. On the other hand, the repeat transactions index method can produce better housing market trends because it models the price appreciations of the same houses between pairs of repeat transactions. This is also the reason that the repeat transactions index has been studied by many researchers and applied to mortgage portfolio risk management by mortgage lenders and investors.[2] However, the repeat transactions index gives no clue to the price levels of housing markets. In addition, the repeat transactions method has some statistical issues that have been overlooked.

The repeat transactions model of Bailey, Muth, and Nourse (1963) utilizes the house price data of houses that have repeat transactions. The model computes the same house price appreciations between pairs of consecutive transactions, and then estimates the market appreciation. Thus the model of Bailey et al. can be called the pairs transactions model, to be more precise and descriptive.[3] The pairs transactions model is able to mitigate the quality effect of houses by analyzing the price appreciations of the same houses and thus is able to produce better market trends. Often mentioned problems with the repeat transactions model are: the waste of data, change of housing attributes over time, and sample selection problems. However, these issues can be addressed (see the discussion in the next section). Another disadvantage of this model is the inability of the index to inform the price level of the market. The lack of price level information prevents comparison of the indices of different markets.

Moreover, several shortcomings of the pairs transactions model are overlooked and rarely addressed. First, there are different ways to break multiple transactions of the same houses into pairs of transactions, and the current way of consecutive pairing may not be the best alternative. Second, the pairs transactions method models house price appreciations rather than house prices, and it may not produce the best predictions of individual house prices. Third, when the multiple transactions data are transformed into the price appreciations of pairs of consecutive transactions, the resulting error terms in the pairs transactions model will be correlated. Although there are discussions of the correlated error terms in the repeat transactions model by Bailey et al. (1963) and Palmquist (1982), no viable solution has been proposed.

This paper proposes a multiple transactions model with a panel data approach to overcome the aforementioned shortcomings of the pairs transactions method. The multiple transactions method directly models the house prices and estimates the market indices without breaking the multiple transactions into pairs and modeling the house price appreciations. Thus the multiple transactions model can provide information on the price levels, as well as the market trends of the housing markets. The next section discusses the pairs transactions model. The multiple transactions model is then developed, followed by a description of the data and the empirical results based on the pairs and the multiple transactions models. The final section presents concluding remarks.

Review of Pairs Transactions Model

Model Description

The pairs transactions method for constructing housing market indices is a regression model proposed by Bailey, Muth, and Nourse (1963),[4] where the individual house price appreciation between a pair of transactions follows the market trend such that:

$$r^i_{tt'} = -b_t + b_{t'} + u^i_{tt'}, \tag{1}$$

where $r^i_{tt'}$ is the house price appreciation (in log form) of the ith pair of house transactions between periods t and t', and b_t and $b_{t'}$ are the market indices for period t and t'. The error term $u^i_{tt'}$ represents the deviation of the observed individual house price appreciation from the market appreciation and is assumed to be *iid* with zero mean and variance σ_u^2 for all i, t, and t'.

To estimate the market index b_t, let X^i_j take the value -1 if $j = t$, the value 1 if $j = t'$, and the value zero otherwise for the ith pair. Equation (1) can then be rewritten as follows:

$$r^i_{tt'} = \sum_{j=1}^{T} X^i_j b_j + u^i_{tt'}, \tag{2}$$

where T is the total number of periods covered by the market index. If all the pairs are stacked together, Equation (2) can be put into a matrix form such as:

$$r = Xb + u, \tag{3}$$

where r is the vector of house price appreciation, b is the vector of the market index, X is a matrix whose elements are X^i_j, and u is the vector of the error terms. Thus the least square estimator of the market index b is:

$$\hat{b} = (X'X)^{-1}X'r. \tag{4}$$

Case and Shiller (1987) and Abraham and Schauman (1991) point out that the error terms of house price appreciations between longer intervals tend to have larger variances than those house price appreciations between shorter intervals do. Thus they propose a three-stage regression process to resolve the heteroscedasticity caused by the different time intervals between pairs of transactions.[5]

The three-stage regression method proceeds as follows. First, do an OLS regression on Equation (2). To avoid perfect collinearity among the explanatory variables, it is necessary to set a restriction on one of the market index parameters. Generally, if a period t is chosen as the base period of the index, then $b_\tau = 0$.

Second, do a regression on the residual squares of Equation (2) such that:

$$(u^i_{tt'})^2 = A(t' - t) + B(t' - t)^2 + C + \varepsilon_i, \tag{5}$$

where A, B, and C are model coefficients, and ε_i is the white noise. Thus the predicted values of $(u^i_{tt'})^2$ is obtained based on the estimated parameters in Equation (5). Finally, the following weighted regression is used to estimate the market index:

$$\frac{r^i_{tt'}}{\sqrt{(\hat{u}^i_{tt'})^2}} = \sum_{j=1}^{T} \frac{X^i_j b_j}{\sqrt{(\hat{u}^i_{tt'})^2}} + \frac{u^i_{tt'}}{\sqrt{(\hat{u}^i_{tt'})^2}}. \tag{6}$$

If $s^2 = \hat{A} + \hat{B} + \hat{C}$, i.e., the predicted variance of house price appreciation over one period, then the solution of the market index in Equation (6) can be written as:

$$\hat{b} = (X'\varphi^{-1}X)^{-1}X'\varphi^{-1}r, \tag{7}$$

where the off-diagonal elements of matrix φ are zero and the diagonal elements are $(\hat{u}^i_{tt'})^2/s^2$. The variance of the estimated market index is:

$$Var(\hat{b}) = s^2(X'\varphi^{-1}X)^{-1}. \tag{8}$$

The predicted house price in the pairs transactions model is:

$$\hat{p}^i_t = p^i_s + \hat{b}_t - \hat{b}_s. \tag{9}$$

This shows that a prior transaction price of the house is needed to predict the house's price at a later time period.

Issues with the Pairs Transactions Model

The commonly mentioned issues with the pairs transactions index model are: (1) data is wasted because the data with only a single transaction cannot be used in the model; (2) change of housing attributes between the sales; and (3) the sample of houses with repeat transactions may not represent the entire housing stock in the market. However, these issues can be mitigated. For example, Clapp and Giaccotto (1992) propose the method of using tax assessment values to pair with the sales values in the pairs transactions model so the properties with a single transaction can be used in the model. Even if the housing attributes changed between transactions, the repeat transactions index is the appropriate index for mortgage lenders. That is because the collateral of the mortgage is the property of the mortgage, regardless of home improvement or depreciation. Sample selection issue can be resolved by increasing the sample size. Case, Pollakowski, and Wachter (1991) also propose the weighting method to correct the sample selection problem so different types of houses can be appropriately weighted in the model. In addition, the approach of estimating a separate index for each segment of the market can resolve the sample selection problem.

However, there are still several issues in the pairs transactions model that are not commonly mentioned. The pairs transactions model attempts to obtain a better fit of house price appreciations between pairs of transaction periods. The model does not attempt to fit the house prices directly. Moreover, if the houses have multiple transactions, there can be many ways of pairing up the house prices.[6] Pairing the house prices consecutively will not necessarily produce the best estimation of market appreciation, nor the best prediction of house prices. Furthermore, the variance and covariance matrix of the error terms in the pairs transactions model will not be diagonal when the houses have multiple transactions.

Bailey et al. (1963) notice that, if the houses used in the estimation of the market index have more than two transactions, there will be no unique way of arranging transaction pairs. Ideally, it will be preferable to model the house price directly as:

$$p_t^i = a_t + c^i + v_t^i, \tag{10}$$

where p_t^i is the log of observed price of house i in period t, a_t is the market index (which can be different from the market index in Equation (1)), c^i is the house-specific effect, and the error term v_t^i is *iid* with a zero mean and variance of σ^2. The error term here reflects the deviation of a particular transaction price from the expected price based on the market index and house-specific intercept term. Bailey et al. reason that if the number of houses is large and many houses have multiple transactions, then it will not be computationally feasible to solve Equation (10) because its solution requires inverting a huge matrix. The alternative method proposed by Bailey et al. is to arrange the transaction prices into pairs and model the price appreciations (differences of log prices) of the same house. That is:

$$p_{t'}^i - p_t^i = -a_t + a_{t'} + v_t^i - v_{t'}^i. \tag{11}$$

By comparing Equation (1) with Equation (11), Bailey et al. (1963) find that the error in the pairs transactions model will be correlated if a house has more than two transactions. This is because the second transaction of the first pair is the first transaction of the second pair of the house (other paring arrangements will have the similar problem). Thus Bailey et al. contemplate the idea of using GLS to deal with the correlated error terms in the pairs transactions model. Palmquist (1982) also notices the problem of correlated error terms in the pairs transactions model. The solution proposed by Palmquist is to pre-multiply Equation (1) by the root matrix of inverted covariance matrix of error terms. Then the OLS can be applied to the pretreated price appreciations and the market index can be obtained. The method proposed by Palmquist is essentially the same GLS approach contemplated by Bailey et al. However, the methods proposed by Bailey et al. and Palmquist are not appropriate.

Although the expression in Equation (11) is correct as the log price difference of two transactions of the same house when the house price is modeled in Equation (10), it will be problematic if it is used as a model of house price appreciation. The error term v_t^i in Equation (10) is the deviation of an individual observed house transaction price from the expected price. Thus the combination of the last two terms in Equation (11) is the difference of the two deviations of two observed house transaction prices from the expected prices. On the other hand, the error term in Equation (1) is the deviation of the observed house price appreciation from the market appreciation. Since the two error terms in Equation (11) and the error term in Equation (1) represent two different things, it will not be appropriate to derive the behavior of the error term in the model of the pairs transactions model based on the assumed error structure in the model of individual house transaction prices.[7] Therefore, Bailey et al. (1963) raise the concerns that the error term in Equation (1) is not just simply the difference of two error terms in Equation (10). There may be another extra component, say *w*, that represents the deviation of a particular house's price appreciation from the market appreciation. Without knowing the variances of *v* and *w*, the GLS regression can not be applied. Thus Bailey et al. turn to the pairs transactions model in Equation (1) and argue that when the cases of multiple transactions are few, and the variance of v_t^i is small, Equation (4) will be reasonably efficient for estimating the market index. If $u_{tt'}^i$ has a mean of zero, Equation (4) will still be an unbiased estimator of the market index. Thus, they suggest using consecutive transactions to do the pairing of house prices when multiple transactions occur for individual houses. These limitations of the pairs transactions model are rarely discussed in the pairs transactions model literature.

Multiple Transactions Model

The proposed multiple transactions model follows Bailey, Muth, and Nourse's (1963) specification in Equation (10). However, a panel data approach is employed to overcome the computational difficulties cited above and to obtain solutions to the model. Thus multiple transactions of house prices can be modeled directly without breaking them into pairs of price appreciations.

In the multiple transactions model, the (log) price p^i_{tk} of an individual house i at time t_k can be expressed in terms of the market index α_{tk}, house-specific term q^i, and the *iid* noise term υ^i_{tk}:

$$p^i_{tk} = \alpha_{tk} + q^i + \upsilon^i_{tk}\; k = 1, 2, 3, ..., n_i,\; i = 1, 2, 3, ..., N, \qquad (12)$$

where the market index α_{tk} only changes from period to period; the house-specific term q^i changes from house to house but stays the same for all transactions of the same house; the noise term υ^i_{tk} is both house and transaction period specific and represents the deviation of a particular house transaction price from the expected price; n_i is the number of transactions for house i; and N is the total number of houses.

Equation (12) can be put into a panel data form for all transaction prices of house i as follows:

$$p^i = Y^i\alpha + q^i J^i + \upsilon^i,\; i = 1, 2, ..., N, \qquad (13)$$

where $p^i = (p^i_{t1}, p^i_{t2}, ..., p^i_{tni})'$, $J^i = (1,1,...,1)'_{ni}$, $\alpha^i = (\alpha_1, \alpha_2, ..., \alpha_T)'$, $\upsilon^i = (\upsilon^i_{t1}, \upsilon^i_{t2}, ..., \upsilon^i_{tni})'$, T is the total number of time periods for which the market index will be estimated, and Y^i is an n_i row by T column matrix. In the *kth* row of the Y^i matrix, the t_kth element is 1 and the rest of elements are zero. The noise term has the following property:[8]

$$E(\upsilon^i) = 0,\; E(\upsilon^i \upsilon^{l\prime}) = \begin{cases} \sigma^2_\upsilon I_{ni\times ni},\; i = l \\ 0,\; i \neq l \end{cases}. \qquad (14)$$

The multiple transactions model in Equation (13) differs from the conventional panel data model in two ways. First, the prices of each house are only observed in a few periods, not in all periods from *1* to T. Thus the model is an unbalanced panel data model. Second, the independent variable Y^i is a time period dummy, not the traditional explanatory variable that determines the prices of houses. The coefficient of the time period dummy, α, is the period-specific market index.

The multiple transactions model in Equation (13) can take either the fixed or random effect model specifications,[9] depending on the assumed behavior of house-specific terms, q^is. In the following subsections, these two model specifications are discussed, along with the test of model selection between the two specifications.

Fixed Effect Model Specification

In the fixed effect model specification, the house-specific terms q^is are assumed to be fixed and to be estimated. In this model specification, the model error in Equation (13) is v^i. Because of the collinearity problem, the house-specific terms q^is and the market index α can not be independently determined. Thus there needs to be a restriction on either q^is or α. With no restriction on the market index α, but a restriction on the house-specific terms q^is, gives:[10]

$$q^1 = -\sum_{i=2}^{N} q^i. \tag{15}$$

In order to obtain the OLS estimate of q^is and α, the sum of error squares of the model in Equation (13) is such that:[11]

$$S = \sum_{i=1}^{N} v^{i\prime} v^i = \left(p^1 - Y^1\alpha + \sum_{l=2}^{N} q^l J^1\right)' \left(p^1 - Y^1\alpha + \sum_{k=2}^{N} q^k J^1\right) + \sum_{i=2}^{N} (p^i - Y^i\alpha - q^i J^i)'(p^i - Y^i\alpha - q^i J^i). \tag{16}$$

Minimizing S with respect to q^is ($i = 2, 3, \ldots, N$) will yield:

$$\hat{q}^i = \overline{p}^i - \overline{Y}^i\alpha - \sum_{l=1}^{N} (\overline{p}^l - \overline{Y}^l\alpha) / \left(n_i \sum_{k=1}^{N} \frac{1}{n_k}\right), \; i = 1, \ldots, N. \tag{17.a}$$

$$\overline{p}^i = \frac{1}{n_i} J^{i\prime} p^i. \tag{17.b}$$

$$\overline{Y}^i = \frac{1}{n_i} J^{i\prime} Y^i. \tag{17.c}$$

If the result in Equation (17) is substituted into Equation (16) and S is minimized with respect to α, then:[12]

$$\hat{\alpha} = \left[\sum_{i=1}^{N} YM^{i\prime}YM^{i}\right]^{-1} \sum_{l=1}^{N} YM^{l\prime}PM^{l}. \tag{18.a}$$

$$YM^{i} = Y^{i} - J^{i}\overline{Y}^{i} + J^{i}\sum_{l=1}^{N}\overline{Y}^{l}/\left(n_i \sum_{k=1}^{N}\frac{1}{n_k}\right). \tag{18.b}$$

$$PM^{i} = p^{i} - \overline{p}^{i}J^{i} + J^{i}\sum_{l=1}^{N}\overline{p}^{l}/\left(n_i \sum_{k=1}^{N}\frac{1}{n_k}\right). \tag{18.c}$$

The variance of the estimated market index is:

$$Var(\hat{\alpha}) = \sigma_v^2 \left[\sum_{i=1}^{N} YM^{i\prime}YM^{i}\right]^{-1}. \tag{19}$$

The variance of the error terms can be estimated by:

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N}[PM^{i} - YM^{i}\hat{\alpha}]'/[PM^{i} - YM^{i}\hat{\alpha}]}{\sum_{i=1}^{N} n_i - T - N + 1}. \tag{20}$$

The predicted house price is given by:

$$\hat{p}_t^i = \hat{\alpha}_t + \hat{q}^i. \tag{21}$$

If a house has no previous observed transaction price, then its house intercept term q^i cannot be estimated from Equation (17). Thus the price for such a house cannot be predicted.

Random Effect Model Specification

If the house-specific terms are treated as random, then the multiple transactions model has the random effect model specification. In this specification, the house-specific term q^i represents the common deviation of all transaction prices of the same house from the market index. In addition, q^i is independent of the noise term v^i that stems from the individual transaction prices and has the following behavior:

$$E(q^i) = 0,\ E(q^i v^l) = 0,\ E(q^i q^l) = \begin{cases} \sigma_q^2, & i = l \\ 0, & i \neq l \end{cases}. \tag{22}$$

Since q^is are random, the error terms of the random effect model specification in Equation (13) are the combination of the house-specific noise and transaction-specific noise, and can be expressed as the following:

$$z^i = q^i J^i + v^i. \tag{23.a}$$

$$E(z^i) = 0,\ E(z^i z^{l\prime}) = \begin{cases} \sigma_v^2 I_{ni\times ni} + \sigma_q^2 J^i J^{i\prime}, & i = l \\ 0, & i \neq l \end{cases}. \tag{23.b}$$

Thus Equation (13) can be rewritten as:

$$p^i = Y^i \alpha + z^i. \tag{24}$$

If V_i denotes the variance matrix of error term z^i, then the inverse of V_i can be computed as:

$$V_i^{-1} = \frac{1}{\sigma_v^2} H^i. \tag{25.a}$$

$$H^i = I_{ni\times ni} - \frac{\sigma_q^2}{\sigma_v^2 + n_i \sigma_q^2} J^i J^{i\prime}. \tag{25.b}$$

Thus the solution for the GLS estimator of the market index in Equation (24) is:

$$\tilde{\alpha} = \left[\sum_{i=1}^{N} Y^{i\prime} H^i Y^i \right]^{-1} \sum_{l=1}^{N} Y^{l\prime} H^l p^l. \tag{26}$$

The variance of the estimated market index is:

$$Var(\tilde{\alpha}) = \sigma_v^2 \left[\sum_{i=1}^{N} Y^{i\prime} H^i Y^i \right]^{-1}. \tag{27}$$

In general, the variances σ_v^2 and σ_q^2 are unknown. Thus they have to be estimated before a GLS estimator of the market index can be obtained. The first variance σ_v^2 can be estimated by Equation (20), and the second variance can be estimated by:[13]

$$\hat{\sigma}_q^2 = \frac{\sum_{i=1}^{N} \left[(\overline{p}^i - \overline{Y}^i \breve{\alpha}) - \frac{1}{N} \left(\sum_{l=1}^{N} (\overline{p}^l - \overline{Y}^l \breve{\alpha}) \right) \right]^2}{N - T} - \frac{1}{N} \sum_{k=1}^{N} \frac{1}{n_k} \hat{\sigma}_v^2, \tag{28}$$

where $\overline{p}^i$ and $\overline{Y}^i$ are defined in Equation (17). The estimator $\breve{\alpha}$ in Equation (18) can be approximated by:

$$\breve{\alpha} = \left[\sum_{i=1}^{N} \overline{Y}^{i\prime} \overline{Y}^i \right]^{-1} \sum_{l=1}^{N} \overline{Y}^{l\prime} \overline{p}^l. \tag{29}$$

The predicted price of house i at time t in the random effect model will be:

$$\tilde{p}_t^i = \tilde{\alpha}_t. \tag{30}$$

The variance of predicted house price is $\sigma_v^2 + \sigma_q^2$. The random deviation of house price from the market index consists of two parts: the house-specific deviation q^i and the noise term v_t^i. If a house has a previous observed price (or prices), then the house deviation term q^i can be computed as:

$$\tilde{q}^i = \frac{1}{n_i} J^{i\prime} (p^i - Y^i \tilde{\alpha}). \tag{31}$$

If the house deviation is added to the predicted house price, then:

$$\tilde{p}_t^i = \tilde{\alpha}_t + \tilde{q}^i. \tag{32}$$

The variance of the predicted house price in Equation (32) can be reduced to σ_v^2. Therefore, the accuracy of the predicted house price can be improved by using the house-specific deviation.

Three important features of the multiple transactions model emerge here. First, the houses with only one transaction will still have an impact on the market index in both the fixed and random effect model specifications of the multiple transactions model.[14] This can be seen from Equations (16), (17), (18), (25), and (26). By contrast, the houses with only one transaction will not have any impact on the market index in the pairs transactions model because these houses will not enter the equations of the model.[15] Second, the market indices estimated by the fixed and random effect specifications of the multiple transactions model do not need to be based in a specific time period. The market index estimated by the multiple transactions model reflects both the price level and the trend of the housing market. Thus the index estimated by the multiple transactions model has the features of the indices by the pairs transactions model and the median house price model, because the multiple transactions method models house prices directly while controlling for the quality of houses by using the house-specific terms. Third, by using the sums of smaller matrixes in Equations (18) and (26) to obtain the market index, the multiple transactions model is computationally efficient and avoids the inversion of huge matrixes.

Model Specification Test

The treatment of the house-specific terms, whether as fixed or random, appears to be arbitrary. If the house-specific terms is treated as fixed, the loss of degrees of freedom can be costly, especially when the number of houses is large and the transactions of each house are few. Thus the random effect model specification sounds more appealing. However, if the house-specific terms are correlated with the market index, the random effect model specification can generate inconsistency because of the omitted variable problem. Therefore, the choice of model specification will be based on the test of orthogonality between the house-specific terms and the market index, which can be done by applying Hausman's (1978) method.

Here is the idea of the Hausman test. Under the null hypothesis of no correlation between the house-specific terms and the market index, both the fixed and random effect estimators of the market index are consistent, but the fixed effect estimator is not efficient. However, under the alternative hypothesis, the fixed effect estimator is consistent, but the random effect estimator is not. Therefore, under the null hypothesis, the two estimators should not differ significantly. The chi-squared test of the difference of two estimators is based on the Wald criterion:

$$W = \chi^2(T) = (\hat{\alpha} - \tilde{\alpha})'[Var(\hat{\alpha} - \tilde{\alpha})]^{-1} (\hat{\alpha} - \tilde{\alpha}), \quad (33)$$

where $\hat{\alpha}$ and $\tilde{\alpha}$ are respectively the market indices based on the fixed and random effect specifications. The computation of the variance term in Equation (33) can be simplified by using Hausman's (1978) results showing that the covariance of an efficient estimator with its difference from an inefficient estimator is zero. Thus the variance term in Equation (33) can be reduced to:

$$Var(\hat{\alpha} - \tilde{\alpha}) = Var(\hat{\alpha}) - Var(\tilde{\alpha}). \quad (34)$$

Under the null hypothesis that the house-specific terms and the market index are uncorrelated, the test statistic W in Equation (33) is asymptotically distributed as chi-squared with T degrees of freedom. When the null hypothesis is satisfied, the random effect model specification should be used for the market index estimation. Otherwise, the fixed effect model specification should be applied.

The Data and Test of Model Specification

The house transaction data used in the analysis are from Howard County, Maryland. The data were collected from the county real estate property tax records, which have information on house sales (arms length) transactions.

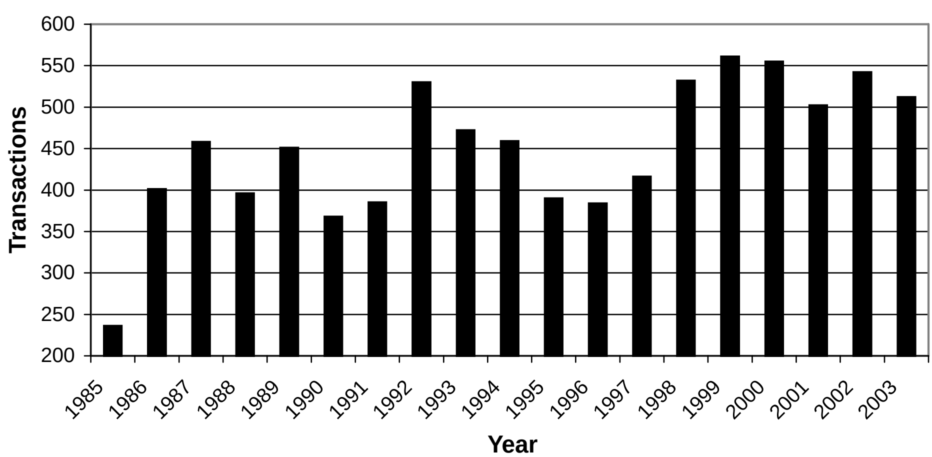
The data for 5,000 houses with 8,550 transactions were collected from five locations (ZIP Codes) in Howard County. One thousand houses were randomly selected from each location. This sample size is comparable to those used in other empirical studies.[16] The dataset contains house transaction prices from 1985 to 2003. Thus, there are 76 quarters of house transaction data. Exhibit 1 lists the frequency of transactions per house for our data. About 48% of the houses have repeat transactions, and about 18% of the houses have more than two transactions. The number of house transactions over time is shown in Exhibit 2. As can be seen, the volume of house transactions peaked during the years 1992 and 1999.

As discussed in the last section, there are two specifications of the multiple transactions model. The first step in the empirical analysis is to determine which model specification should be used for the data. Thus the Hausman model specification test described in the last section was applied to the house transaction data. The chi-squared values of the Hausman test are shown in Exhibit 3 for all locations together (the aggregate market) and for each individual location. The chi-squared statistics are significant for all locations and the aggregate market, all

Exhibit 1 | Frequency of House Transactions

# of Transactions per Each House	# of Houses	% of Houses	Total Number of Transactions	Percentage of Transactions
1	2,596	51.9%	2,596	30.4%
2	1,528	30.6%	3,056	35.7%
3	659	13.2%	1,977	23.1%
4	170	3.4%	680	8.0%
5	41	0.8%	205	2.4%
6	6	0.1%	36	0.4%
Total	5,000		8,550	

Exhibit 2 | Number of House Transactions in Each Year



houses are used or only those with repeat transactions. The exception is when only the houses with repeat transactions are used in location 2. Thus, the fixed effect specification of the multiple transactions model should be used for the house transaction data in this study.

Empirical Results

In the following, the data described in the last section is used to investigate the performance of the pairs transactions model and the multiple transactions model

Exhibit 3 | Hausman's Chi-squared Model Specification Test

	Using All Houses	Using Houses with Repeat Transactions
Aggregate Market	105.6**	104.4**
Location 1	108.2***	132.9***
Location 2	106.0**	90.8
Location 3	147.4***	175.3***
Location 4	122.5***	134.4***
Location 5	243.5***	229.7***

Notes:
** Significant at the 5% level.
*** Significant at the1% level.

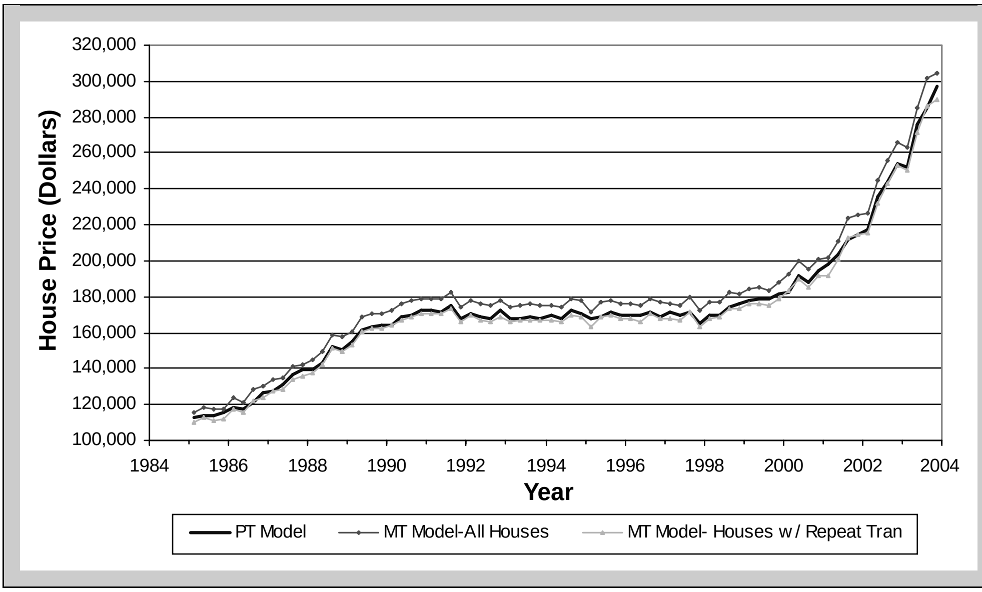
with the fixed effect specification. First, the estimated indices and their variances are compared using the two index models. Then there is an examination of which model does better in terms of out-of-sample prediction accuracy on individual house transaction prices. The robustness of index estimates is also examined in terms of the difference between the full sample index and the sub-sample index. Because the indices from both models are subject to revision when house transaction data in a new period arrives, the index revision volatility is also a measure for determining the desirability of index methodologies.[17] Thus the revision volatility of the two index models is also examined.

Estimated Market Indices

Two indices can be estimated by the multiple transactions model: one by using only the houses with repeat transactions, the other by using all houses, whether they have single or repeat transactions. The market indices for all locations together can be estimated (the aggregate market), along with each individual location.

First, the indices of the aggregate market are computed. Examination of the indices estimated by the different models in Exhibit 4 reveals that the two models are quite similar for most of the periods. For the periods where the indices are different, the resulting quarterly growth rates of the market indices can be quite different. The two indices estimated by the multiple transactions model look alike. The inclusion of the houses with only a single transaction in the multiple transactions model produces a parallel shift of the market index from the index based on the houses with repeat transactions. If one index is rebased in the first period to be the same as the other index, then the two indices produced by the multiple transactions model will be the same for all periods.

Exhibit 4 | Comparison of Aggregate Market Indices Across Models



An exponential function is applied to the log form of the market indices. The index of the pairs transactions (PT) model is rebased such that it equals the average of the two indices of the multiple transactions (MT) model in the first period.

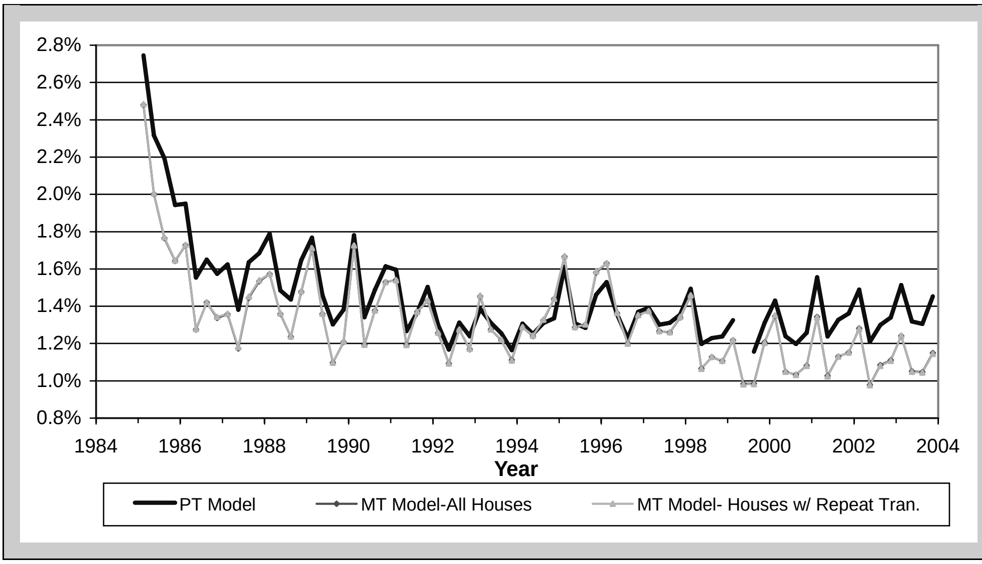
Exhibit 5 shows that the standard errors of the market index of the pairs transactions model are higher than those of the multiple transactions model in almost all periods. It also shows that the two indices produced by the multiple transactions model have nearly the same standard errors for all periods.

Because the multiple transactions model can produce price level information on housing markets, the market indices can be compared across locations. The results in Exhibit 6 show that the level of the market index in Location 1 is higher than the level of the market index in other locations. The index in Location 1 also grew faster over the last eighteen years. The indices of the other four locations are very close to the aggregate market index. The index in Location 4 has the lowest level.

Predicted House Prices

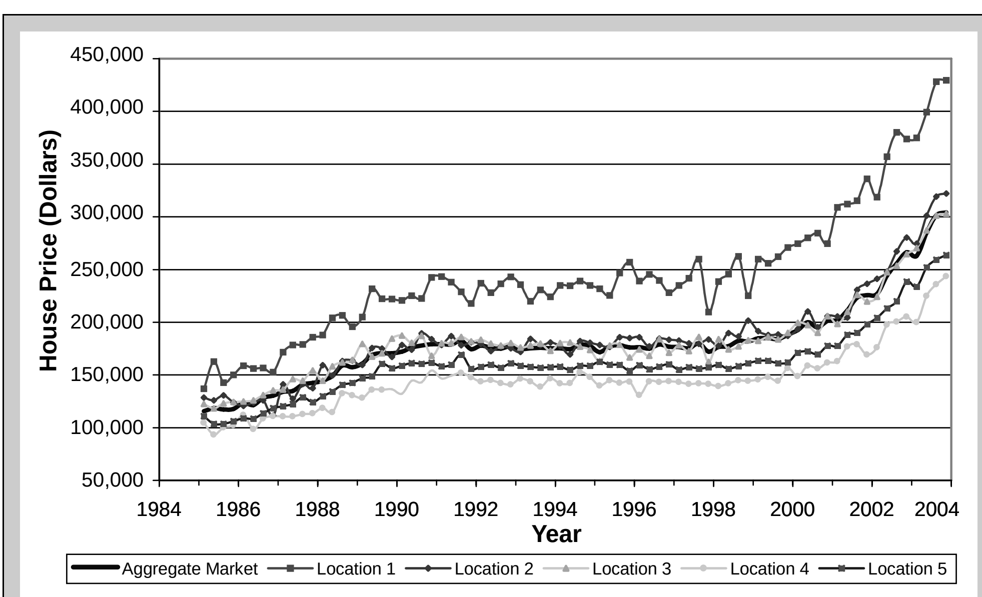
Now the accuracy of predicted house prices based on the pairs and multiple transactions models is compared, using the out-of-sample test technique. The test is based on the aggregate market indices and model parameters. The procedure of the test is the rotation of estimation and holdout samples. First, the entire set of observations of house transactions is randomly divided into ten groups with roughly the same number of observations in each group with no house having more than one transaction in each group. One group of data is the holdout sample,

Exhibit 5 | Standard Error of Aggregate Market Indices Across Models



Standard error is computed by applying the square root of variance of the market index (in log form). The second quarter of 1999 has the most transactions, thus is set as the base period for the pairs transactions (PT) model.

Exhibit 6 | Market Indices Across Locations by The Multiple Transactions Model



An exponential function is applied to the log form of the market indices. All houses are used in the estimation of the market indices.

and the remaining nine groups are the estimation sample. The market indices of both models are estimated based on the estimation sample. The house-specific intercept terms in the multiple transactions model are also estimated based on the estimation sample. Thus, the predicted house prices can be computed based on the estimation sample. The house prices in the holdout sample are used for comparison to the predicted house prices and derivation of the prediction errors. The holdout sample is rotated throughout all ten groups. For each rotation of the holdout sample, the market indices and prediction errors are computed for both the pairs and multiple transactions models. The design of out-of-sample prediction analysis is different from Clapp and Giaccotto's (2002), where the house transaction data in earlier periods are used as the estimation sample and the house transaction data in the last six quarters are used as the holdout sample. In Clapp and Giaccotto's test design, the predicted house prices will likely have time lag, especially in the rapid moving housing markets.

The predicted house prices are based on Equation (9) for the pairs transactions model and Equation (21) for the fixed effect specification of the multiple transactions model. The prediction errors are defined as:

$$pe_t^i = \hat{p}_t^i - p_t^i, \tag{35}$$

where $\hat{p}_t^i$ is the log of the predicted house price base on the estimation sample and p_t^i is the log of the observed house price from the holdout sample.

If one transaction of a house is used as the observed price in the holdout sample, then at least one other transaction of the house will be needed in the estimation sample to compute the predicted price of the house; thus, only the houses with repeat transactions are used. Besides, the houses with only one transaction will not have an impact on the pairs transactions model, and they will not affect the market trend or the predicted house prices in the multiple transactions model.

The test results are divided into two groups: one including the houses that have single transaction in the estimation sample, and the other including the houses that have multiple transactions in the estimation sample. The results of the out-of-sample prediction errors test are shown in Exhibit 7. Measured by the standard deviations of prediction errors, the multiple transactions model outperforms the pairs transactions model. The multiple transactions model has a larger improvement for the houses with multiple transactions in the estimation sample compared to the houses with only one transaction in the estimation sample. This indicates that when multiple transactions of houses are used to estimate the house-specific intercept terms, the predicted house prices can be more accurate. Overall, the analysis on the predicted house prices shows that the multiple transactions model is better than the pairs transactions model.

Exhibit 7 | Prediction Errors of Individual House Transaction Prices

	With Multiple Transactions in Estimation Sample			With One Transaction in Estimation Sample		
	# of Obs.	Mean of Prediction Errors	Std. Dev. of Prediction Errors	# of Obs.	Mean of Prediction Errors	Std. Dev. of Prediction Errors
Pairs Transactions Model	2,898	−0.431%	10.57%	3,056	0.005%	12.79%
Multiple Transactions Model	2,898	0.011%	10.14%	3,056	0.012%	12.78%

Robustness of Two Index Estimates

The robustness of the two models can be checked by rotating the estimation and holdout samples as described in the last sub-section. The measure of robustness is the difference of the sub-sample index from the full sample index. Specifically, if b_t^f is the index estimated based on the full sample (all ten groups), b_t^k is the index based on the sub-sample when the *k*th group of observations is used as the holdout sample while the remaining nine groups are used as the estimation sample, then the difference of the sub-sample index from the full sample index is defined as:

$$d_t^k = b_t^k - b_t^f, k = 1, 2,..., 10, t = 1, 2, 3,..., 76. \quad (36)$$

The quarterly growth rate difference of the sub-sample index from that of the full sample index is defined as:

$$dgr_t^k = (b_t^k - b_{t-1}^k) - (b_t^f - b_{t-1}^f),$$
$$k = 1, 2,..., 10, t = 2, 3,..., 76. \quad (37)$$

The standard deviations of index difference and quarterly growth rate difference are computed for all *k*s and *t*s.

The results in Exhibit 8 show that the standard deviation of the index difference of the multiple transactions model is smaller than that of the pairs transactions model. The standard deviation of the quarterly growth rate of the index of the multiple transactions model is also smaller than that of the pairs transactions

Exhibit 8 | Difference of Sub-sample Index and Full Sample Index

	Std. Dev. of Index Difference	Std. Dev. of Quarterly Growth Rate Difference
Pairs Transactions Model	0.73%	0.86%
Multiple Transactions Model	0.59%	0.80%

model. Thus the two measures of robustness show that the multiple transactions model is more robust and superior to the pairs transactions model.

Index Revision Analysis

The last test on the performance of the pairs transactions model and the multiple transactions model is the test of index revision. It is well-known that when house transaction datasets are updated with observations for a newly reported time period, the re-estimated market index for the earlier time periods can be revised. Furthermore, the amount of revision for the most recent prior period will be larger than that of the earlier periods. The analysis of index reversion starts with the estimation of the aggregate market index using the house transaction data through the first 48 quarters to analyze the revision of the market index. The market index is re-estimated as each additional quarter's house transaction data is added to the dataset. The index revision amount is defined as the difference in the quarterly growth rate of the market index for the same period before and after adding one more quarter's house transaction data. This process is repeated until all 76 quarters' house transaction data is included in the estimation of the market index.

Explicitly, the revision amount can be expressed as:

$$rev_{T-k}^{T} = (b_{T-k}^{T} - b_{T-k-1}^{T}) - (b_{T-k}^{T-1} - b_{T-k-1}^{T-1}),$$
$$T = 49, 50, \ldots, 76,\ k = 1, 2, 3, 4, \qquad (38)$$

where b_{T-k}^{T} is the market index for time period $T - k$ by using the house transaction data up to period T. Here k indicates how far back the revision goes, and it is called the vantage of revision. The index revision is examined up to four quarters back from the current period. Then the standard deviation of the revision amount can be computed for each vantage by using the revision amounts at T = 49, 50,..., 76, for total of 28 periods. Since the quarterly growth rate of the market

Exhibit 9 | Revision of Quarterly Growth Rate of the Market Index

	# of Obs.	Mean	Std. Dev.	Minimum	Maximum
Panel A: One Quarter Vantage					
Pairs Transactions Model	28	0.20%	0.52%	−0.97%	1.18%
Multiple Transactions Model	28	−0.04%	0.36%	−0.90%	0.52%
Panel B: Two Quarters Vantage					
Pairs Transactions Model	28	0.20%	0.27%	−0.32%	0.78%
Multiple Transactions Model	28	−0.05%	0.25%	−0.44%	0.54%
Panel C: Three Quarters Vantage					
Pairs Transactions Model	28	0.18%	0.28%	−0.76%	0.87%
Multiple Transactions Model	28	−0.07%	0.12%	−0.40%	0.30%
Panel D: Four Quarters Vantage					
Pairs Transactions Model	28	0.16%	0.27%	−0.25%	0.94%
Multiple Transactions Model	28	−0.07%	0.11%	−0.31%	0.07%

index is being analyzed, only the houses with repeat transactions are used, because the houses with single transactions will not affect the market index appreciation for either the pairs transactions model or the multiple transactions model.

Exhibit 9 summarizes the amounts of the index revisions for the most recent four quarters of vantages. The average revision of the multiple transactions model is nearly zero while the index of the pairs transactions model tends to be revised upward. In addition, the standard deviation of the index revision amount of the multiple transactions model is smaller than that of the pairs transactions model. Thus, based on the revision analysis of the estimated market index, the multiple transactions model is better than the pairs transactions model.

Conclusion

The multiple transactions method proposed in this paper models house prices directly without breaking them into pairs of transactions. A panel data approach is used to resolve the computational difficulties confronted by Bailey et al. (1963). The multiple transactions model can avoid the problem of correlated errors in the pairs transactions model, and produce the market indices that reflect both the level and the trend of the housing markets.

The multiple transactions model is studied empirically with sales transaction data for 5,000 houses in Howard County, Maryland. The Hausman (1978) test shows that the fixed effect specification of the multiple transactions model should be applied to the data. The empirical results reveal that the variance of the estimated market index of the multiple transactions model is smaller than that of the pairs transactions model. The out-of-sample test on the prediction errors of individual house transaction prices indicates that the multiple transactions model is more accurate than the pairs transactions model. When the deviation of the sub-sample index from the full sample index is examined, the findings reveal that the multiple transactions model is more robust than the pairs transactions model. Finally, the study of index revision demonstrates that the multiple transactions model produces a market index with less revision volatility than the pairs transactions model does.

The multiple transactions model can overcome the shortcomings of the repeat transactions model and performs better on many measures based on the empirical data. Researchers and real estate practitioners should consider using the multiple transactions model for constructing housing market indices, monitoring housing market trends, managing mortgage portfolio risks, and marking house prices to market. Future research should include more empirical study of the multiple transactions model on sample data from other geographic areas. Some more recent advanced research work on panel data model can also be applied in the future extension of the multiple transactions model.[18]

Endnotes

1 For example, the median house price published by National Association of Realtors (NAR) and the repeat transactions index published by the Office of Federal Housing Enterprise Oversight (OFHEO) are widely followed by economic and financial reporters. These two types of index are available for most metropolitan areas and states in the United States. The other frequently researched house price index method is the hedonic index model. However, because of the issues of omitted variables, model mis-specification, and more importantly, data availability, the hedonic index model has been applied to only a handful of local housing markets. For more discussions of the hedonic index method, see Musgrave (1969), Palmquist (1980), Meese and Wallace (1991), Case and Quigley (1991), for example.

2 The research of Zhou (1997) is one of the few exceptions that study the time series of median house price. On the other hand, the studies on the time series of repeat transactions index are numerous (see Case and Shiller, 1989; Nothaft, Wang, and Gao, 1995; Cho, 1996; Gu, 2002; Jud and Winkler, 2002; and Crawford and Fratantoni, 2003). The repeat transactions index by OFHEO is also used by government agencies and mortgage lenders to assess the mortgage risks.

3 Meese and Wallace (1991) use similar terminology, the ''paired sales technique,'' to denote the method of Bailey et al. (1963).

4 An early attempt to use the repeat transactions data for constructing housing market indices is the multiplicative chain (or bootstrap) method proposed by Wyngarden (1927) and enhanced by Wenzlik (1952). The chain method takes the average of relative prices

of the houses that have pairs of transactions in the base (zero-th) period and the first period, and obtains the index for the first period. The relative prices of the houses that have pairs of transactions starting from the first period are adjusted by the index of the first period. Then the index of the second period is constructed by taking the average of relative prices of the houses that have pairs of transactions between the base period and the second period or between the first period and the second period. The process is replicated for the third period and so on until the indices of all periods are constructed.

5 An interesting work by Evans and Kolbe (2005) analyzes the heteroscedasticity and abnormal returns associated with selection of real estate agent by using the pairs transactions model.

6 In the pairs transactions model where heteroscedasticity presents, different indices will be produced by different pairing arrangements.

7 See Englund, Quigley, and Redfearn (1998) for an example of identifying the variance-covariance matrix of disturbances based on the assumed error structure.

8 In an interesting work by Clapp and Giaccotto (1992), the sales transactions are paired with the tax assessment values to obtain the price appreciation from the time of tax assessment to the time of sales transaction. In this setup, if a house has two sales transactions, there would be two pairs of price appreciations. Then the error terms of these two price appreciations would be positively correlated. The correlation of the error terms in this model is due to the potential errors in the tax assessment value, which pairs with both sales transactions, and thus the same error in the tax assessment value can be introduced to the price appreciations of both pairs. On the other hand, the error terms of individual transaction prices are still uncorrelated.

9 For an overview of the fixed and random effect panel data method, see Hsiao (1976).

10 Alternatively, there can be a restriction on the market index with the index set in one period as fixed, say zero. Then the house-specific terms can be freely determined. If this alternative is used, the estimated market index will not be impacted by the houses with only one transaction. In addition, the level of the market index will not reflect the price level of the housing market. The only difference between the indices of the two approaches is the level, not the trend of the market indices. Thus the houses with a single transaction will not affect the trend of market index produced by the fixed effect specification of the multiple transactions model.

11 Judge, Hill, Griffiths, Lutkepohl, and Lee's (1988) partitioned matrix inversion method can also provide a solution to the model.

12 By expressing the solution of $\hat{q}^i$s as the function of α and substituting it into Equation (16), the expression becomes the reduced form the minimization problem. The final solution for $\hat{q}^i$s and $\hat{\alpha}$ will have the sum of error squares minimized with respect to both parameters.

13 See Hsiao (1976) for the general idea of deriving the estimators of these two variances.

14 The houses with only one transaction will have an impact on the level, but not the trend of the market index in the fixed effect model specification (see Endnote 10).

15 A related discussion might be the effect of an artificially observed higher sales price on both the pairs transactions model and the multiple transactions model. For example, if the sales price of a house's first transaction is artificially higher, the index of the pairs transactions model can be impacted because the price appreciation will be lower between the first transaction and the second transaction of the house. Likewise, this will also impact the index of the multiple transactions model because the individual transaction

prices of all houses are modeled directly. However, if all transactions of a house are artificially higher by the same proportion (which is very unlikely), then the index of the pairs transactions model will not be impacted because the price appreciation of the house will not change. In this case, the index of the multiple transactions model will still be impacted. However, while the index levels of both the fixed and random effect specifications will affected, the trend of the index produced by the fixed effect specification will not be affected. In both cases of one or all transaction prices being artificially higher, the house-specific term q^i will be impacted more and the index will be impacted to a lesser degree. Because just as the price of a house might be artificially higher, it is equally possible that the price of another house might be artificially lower. Thus the overall net effect will be that the opposite forces cancel each other out and the index of the multiple transactions model will not be biased one way or the other.

[16] For example, Bailey et al. (1963) use a dataset with 1,512 transaction pairs; Case and Shiller (1987) have pairs of transactions ranging from 6,669 to 15,530 for four metropolitan areas; Palmquist (1980) has 1,613 pairs of transactions; Case, Pollakowski, and Wachter (1991) have 1,765 pairs of transactions; Meese and Wallace (1991) study 16 municipalities with transaction pairs ranging from about 2,000 to 16,000; and Clapp and Giaccotto (1999) have 5,510 and 9,351 pairs of transactions for each of the two counties studied.

[17] See discussions of Shiller (1993), Clapp and Giaccotto (1999), Clapham, Englund, Quigley, Redfearn (2004), and Butler, Chang, and Cutts (2005).

[18] For example, Kezdi (2004) analyzes the case when the error terms of the panel data model are serially correlated.

References

Abraham, J. and W. Schauman. New Evidence on Home Prices from Freddie Mac Repeat Sales. *Journal of the American Real Estate and Urban Economics Association*, 1991, 19: 3, 333–52.

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

Butler, J.S., Y. Chang, and A.C. Cutts. Revision Bias in Repeat-Sales Home Price Indices. Freddie Mac Working Paper, 2005, #05-03.

Case, B., H.O. Pollakowski, and S.M. Wachter. On Choosing Among House Price Index Methodologies. *Journal of the American Real Estate and Urban Economics Association*, 1991, 19:3, 286–307.

Case, B. and J. Quigley. The Dynamics of Real Estate Prices. *Review of Economics and Statistics*, 1991, 73:1, 50–8.

Case, K. and R. Shiller. Prices of Single-Family Homes Since 1970: New Indexes for Four Cities. *New England Economics Review*, 1987, September-October, 45–56.

——. The Efficiency of the Market for Single Family Homes. *American Economic Review*, 1989, 79:1, 125–37.

Cho, M. House Price Dynamics: A Survey of Theoretical and Empirical Issues. *Journal of Housing Research*, 1996, 7:2, 145–72.

Clapham, E., P. Englund, J.M. Quigley, and C.L. Redfearn. Revisiting the Past: Revision in Repeat Sales and Hedonic Indices of House Prices. Working Paper. Institute of Business and Economic Research, University of California–Berkeley, 2004.

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.

——. Revisions in Repeat Sales Price Indexes: Here Today, Gone Tomorrow? *Real Estate Economics*, 1999, 27:1, 79–104.

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

Crawford, G.W. and M.C. Fratantoni. Assessing the Forecasting Performance of Region Switching, ARIMA and GARCH Models of House Prices. *Real Estate Economics*, 2003, 31:2, 223–44.

Englund, P., J.M. Quigley, and C.L. Redfearn. Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices. *Journal of Urban Economics*, 1998, 44, 171–96.

Evans, R.D. and P.T. Kolbe. Homeowner's Repeat-Sale Gains, Dual Agency and Repeated Use of the Same Agent. *Journal of Real Estate Research*, 2005, 27:3, 267–92.

Gu, A.Y. The Predictability of House Prices. *Journal of Real Estate Research*, 2002, 24:3, 213–33.

Hausman, J.A. Specification Test in Econometrics. *Econometrica*, 1978, 46:6, 1251–71.

Hsiao, C. *Analysis of Panel Data*. Cambridge: Cambridge University Press, 1986, 29–32.

Jud, G.D. and D.T. Winkler. The Dynamics of Metropolitan Housing Prices. *Journal of Real Estate Research*, 2002, 23:1/2, 29–45.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lutkepohl, and T.C. Lee. *Introduction to the Theory and Practice of Econometrics.* New York City, NY: John Wiley & Sons, 1988, 468–72.

Kezdi, G. Robust Standard Error Estimation in Fixed-Effect Panel Models. *Hungarian Statistical Review,* 2004, 9, 95–116.

Meese, R. and N. Wallace. Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Housing Price Indices. *Journal of the American Real Estate and Urban Economics Association*, 1991, 19:3, 308–32.

Musgrave, J.C. The Measurement of Price Changes in Construction. *Journal of the American Statistical Association*, 1969, 64, 771–86.

Nothaft, F., G.H.K. Wang, and A.H. Gao. The Stochastic Behavior of the FREDDIE-FANNIE Conventional Mortgage Home Price Index. Annual Meeting of the American Real Estate and Urban Economics Association, Washington, DC, January, 6–8, 1995.

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

——. Measuring Environmental Effects on Property Values Without Hedonic Regression. *Journal of Urban Economics*, 1982, 11, 333–47.

Shiller, R.J. Measuring Asset Value for Cash Settlement in Derivative Markets: Hedonic and Repeated Measures and Perpetual Futures. *Journal of Finance*, 1993, 48:3, 911–31.

Wenzlik, R. As I See the Fluctuations in Selling Prices of Single-Family Residences. The Real Estate Analysis XXI, 1952, 24, 541–48.

Wyngarden, H. An Index of Local Real Estate Prices. Michigan Business Studies, University of Michigan, Ann Arbor, 1927, 1:2.

Zhou, Z. Forecasting Sales and Price for Existing Single-Family Homes: A VAR Model with Error Correction. *Journal of Real Estate Research*, 1997, 14:1/2, 155–67.

The authors wish to thank Bradford Case, Mark An, Chionglong Kuo, Yevgeny Yuzefovich, Amy Crews Cutts, Tom Lutton, Karl Snow, Dean Gatzlaff, and the participants at the 2005 AREUEA mid year meeting, Joint Statistical Meetings, Financial Management Association Meetings, Washington Area Finance Association Meetings, and the 2006 AREUEA annual meeting for helpful comments. Michael Goodrum provided valuable editing work. The reviews and comments by Ko Wang and four anonymous referees helped to substantially enhance the quality of the paper. The views expressed in the paper are entirely the authors and do not necessarily reflect the views of Fannie Mae.

Andre H. Gao, Fannie Mae, Washington, DC 20016 or andre_h_gao@fanniemae.com.

George H.K. Wang, George Mason University, Fairfax, VA 22030 or gwang2@ gmu.edu.