Issues and Results in Housing Price Indices: A Literature Survey

Pei Li
Yong Tu

October 2011
ABS:

Credible housing Price indices play an important role in understanding housing market behavior as well as in simplifying the process of comparing housing price changes over time. For this reason, a number of recent studies have explored various methodologies for the construction of quality-adjusted housing price indices as a means of estimating the rate of inflation in housing prices. However, no single model is free from drawbacks. Also, the conceptual weakness and assumptions in each index are generally different. This paper is in framing these issues, and in specifying a precise metric for comparison. It emphasizes that ideal house price indexes should control for the age effect and impact of illiquidity, should not highly require revision of historical index numbers when data for subsequent periods are added, and should be ease of administration. As implied in this survey, there remain a healthy research agenda before these approaches will be ready to be adopted for widespread use. More attention needs to be paid to where different housing price indexes are aiming and how well these estimators approximate their aims.

We thank the National University of Singapore academic research fund for their financial assistance. Li specially thanks IRAS and Prof Deng for further assistance and his guidance. All errors are ours.

* Corresponding author.
E-mail address: irslp@nus.edu.sg (Pei Li)
1. Introduction

A house price index provides a yardstick of measuring the standard of living and wealth of households and nations. Charting the course of housing prices has been ever more important, as rapid increases in housing prices across a number of countries have led to concerns about housing markets being overvalued and to fears that this might be followed by a sharp correction in prices. This has resulted in a fairly large literature developed on the adequacy of these indexes. Since real estate varies widely in “quality” and transactions are infrequent, the aim for most of the house-price indexes is to track the rate of price appreciation over time for a standard or representative house. These methods vary from the construction of hedonic price indices proposed by Court (1939) to the repeat sales price index originally developed by Bailey et al. (1963) and then by Case and Shiller (1987). With the growing importance of accurate market value appreciation, recent authors have approached this task with a variety of hybrid approaches which combine the hedonic and repeat sales techniques (e.g., Case and Quigley, 1991; Quigley, 1995). In theory, hybrid formulation avoids the inefficiency of using pure repeat sales and the problems of possible misspecification inherent in the hedonic methodology.

However, no single model is shown to be best. Mean drawbacks to the hedonic technique include ignorance of both the functional form of the relation and of the appropriate set of house characteristics to include in the analysis. Since repeat sale model derives from the hedonic approach, all of the apparent weaknesses of the
hedonic model apply equally to the repeat sales models, but are merely hidden from view (Dombrow et al., 1997). Moreover, sample selection bias sensitivity to small sample problem and nonconstancy of implicit housing characteristic prices would appear to plague the repeat-sales approach. In particular, in many case within the literature, age effects, index revision problem and impact of illiquidity have either been ignored or rationalized away. The recognition of such limitations to all indices motivates the following review of the types of problems to which the commonly used index may be vulnerable. Our purpose in illustrating possible drawbacks lying behind each index is not to simplistically claim that one method is right or better than another but rather to see what insights we gain by comparing and contrasting different conceptual weakness and assumption in various indices.

The reminder of the paper is organized as follows. In next section we critically review the previous age-adjusted solution. Section 3 discusses the source and properties of index revision and some empirical implications of this volatility. In Section 4 we describe the primary factors accountable for the observed intertemporal variation in transaction frequency of real estate and current attempts to control the impact of such liquidity. Section 5 discusses the criteria for an ideal housing price index.

2. Adjusting for age effects

A closer look at the economics of housing prices suggests several reasons why the age of a home has a complicated effect on the price. Age effects derived from many possible sources refer to the property’s changes in value when it ages over time,
ceteris paribus (e.g., Palmquist, 1979; Cannaday and Sunderman, 1986; Dixon et al., 1999, Harding et al., 2007). For example, physical deterioration, aesthetic or functional obsolescence, lemon effect and embodied technical change in the construction of new houses may reduce the relative price of an efficiency unit of older housing services, and these factors lead to depreciation. Obviously, this presents a problem, for the failure to distinguish between pure inflation and depreciation on houses, will in large part impede widespread application of indexes of house prices. As such, in the presence of age effects, an ideal housing price index should abstract from the depreciation to isolate the true price trends.

Considering the role of age effects, however, is difficult for several reasons. First, because age effects apply to all transaction data, unlike other occasional quality change such as renovations, we cannot simply drop them from our sample or control them by sample selection (Chau et al., 2005). As a result, only solution is to estimate the age effects explicitly or implicitly. Second, for the challenge of obtaining data, it is difficult to distinguish maintenance from deterioration. This follows from the model’s inability to an explicit treatment of expenditure on repair. Third, as Bailey et al. (1963) indicate, if age is associated with length of tenure, a depreciation adjustment in their formulation will lead to a singularity in the regressor matrix which precludes estimation of the depreciation rate within the repeat-sale model. Thus, in order to adjust the price index for depreciation, it seems that additional information would be needed. Although there are various identification problems associates with this type of
work, how to adjust the change in price indices for depreciation to isolate the pure inflation component becomes underlying foci of several studies. Alternative paths of depreciation, exact multicollinearity in estimating age and time effects, age-related heteroskedasticity, sample selectivity and interpretations of the age coefficients have been the major issues of investigation.

2.1 Identification of depreciation in the early studies

Older studies often employ the observed age method or perpetual inventory method to estimate the rate as which structures depreciate (e.g., Hoad, 1942; Weston, 1972; Leigh, 1980 and many others). One drawback of these two methods is that they restrict functional form in a manner which arbitrary imposes a particular depreciation pattern. As an alternative method, regression technique has become the most widely used approach for this type of research. An early study by Chinloy (1977), following the durable goods models used by Hall (1971) and Hotelling (1925), firstly tries to develop a technique for separating depreciation from true price changes. Under Chinloy’s theoretical framework, the relative price definition has been expressed as the product of a depreciation index, a quality index, a characteristics index and an interaction term. As an application of his proposed structure, Chinloy finds that geometric net depreciation exists in the single-family dwellings in suburban London, Ontario, 1967-1975. Later, Chinloy (1979b, 1980) extends his former model to estimate the net depreciation of a house exclusive of maintenance and test the identity of net depreciations among different jurisdictions. He also adds household variables in
regressions for conformity with his assumption that expected selling price must be equal to the discounted net consumption of housing for that household. The results show that maintenance is not dependent on cities but on the age structure in either London or St. Catharines. An important implication is that excluding a model of repair and maintenance in housing where these phenomena are important can lead to erroneous estimates of depreciation rates.

Whereas some assumption made in his underlying models should be tempered, Chinloy (1980) note that more general application may involve a closer examination of these assumptions. In Palmquist’s (1979) comment, he finds the incorrectness of Chinloy’s theoretical derivation. It has been shown in this study that only on the strength of additional data to provide an independent estimate of the rate of depreciation, the true price effects and depreciation effects could be distinguished. However, what “additional data” should be gotten has not been clarified in Palmquist’s (1979) paper (Chinloy, 1979a). Regarding depreciation, Palmquist (1980, 1982) adjusts the price-relative for the depreciation in the years between the sales manually using an independent estimate from the hedonic results. As Palmquist (1982) notes, this method requires the assumption of a stable depreciation rate between areas and ages of houses. This assumption has been the focus of considerable debate in the sequential studies (e.g., Clapp and Giaccotto, 1998; Chau et al., 2005).

2.2 Alternative paths of housing depreciation

Before Hulten and Wykoff’s (1981) work, very few empirical studies have been
undertaken to test whether some alternative path of depreciation is supported by empirical evidence. Hulten and Wykoff (1981, 1996) suggest that the Box-Cox power transformation which contains the functional forms most often discussed in the depreciation literature as special cases is ideally suited for analyzing depreciation. They use a polynomial power series and Box-Cox power transformations to determine the speed and path of depreciation, and find that an alternative path of depreciation that is initially more accelerated than straight-line is supported. Although they do not include residential properties in their study, as a landmark study, it does act as a catalyst for other researches in this field. Cannaday and Sunderman (1986) incorporate the time variable suggested by Bryan and Colwell (1982) and a particular form of the age variable in the generally log-linear hedonic price model to test for alternative paths of depreciation with just one model. The empirical evidence for the Southwood area of Champaign in their paper supports a path of depreciation for single-family houses that is concave.

Malpezzi et al. (1987), using the same hedonic model and representative data from each of 59 metropolitan areas, find the concave depreciations dominated in the empirical results. They also find that depreciation rates, and discounts for age, differ among the targeted 59 metropolitan areas. Regarding different incentives with respect to care for property between tenants and ownership-occupants, tenant occupancy dwellings may depreciate faster (or slower) than owner-occupied units. Leigh (1979) proves that the rental properties depreciate more rapidly than owned properties, so do

2.3 Separating the effects of age and time

Many attempts as solving the multicollinearity problem focus mainly on separating the effects of age and time. Clapp and Giaccotto (1998) attempt to introduce a nonlinear weighting of at the times of the two sales, i.e. separate time and age by interacting age with time dummies. To avoid the singularity of explanatory variables, Cannaday et al. (2005) break down the age variable into a series of age dummies and then drop two age variables and one time variable. The age-adjusting price index derived from their multivariate repeat-sales model used to disentangle the effects of age and time is shown to be an extended Case-Shiller index. However, both of above methods have been proved to not be free from exact multicollinearity by Chau et al. (2005). Under the assumption of a non-constant deprecation rate over time, they suggest that the non-linearity of age effects has been approximated by highly flexible Box-Cox functions. They prove that the repeat-sales model based on a two-stage estimation procedure employed to estimate the coefficients can effectively disentangle the age and time effects. While the Box-Cox transformations allows patterns in the data to emerge rather than forcing the patterns into a predefined functional form, the interpretation of results also become more difficult.
In other cases, procedures have been introduced which combine the information from repeat sales transactions with the hedonic model estimates in an effort to control for the effects of age. Hill et al. (1997) jointly estimate typical hedonic and repeat sales models via maximum-likelihood procedures to capture depreciation within the repeat sales model and account for serial correlation in hedonic data. The improvement in precision obtained by estimating the joint model is illustrated by smaller standard errors and confirmed by the results of empirical example and simulation that illustrate the properties of the maximum-likelihood estimation procedure. Harding et al. (2007) include the log of house age to avoid the perfect collinearity issue. In such expression, the age coefficient refers to the elasticity of house price appreciation with respect to the change in age between sale dates.

2.4 Age-related heteroskedasticity

The age-related heteroskedasticity has been also one major issue of investigation. As dwelling age, the probability that they have renovations of different vintages increases. In addition, better quality structures have been proved to be more likely to survive for a long period of time and to be renovated (Randolph, 1988b). However, the major home improvements are usually not recorded in publicly available data sets. Consequently, lack of data on such renovations in the hedonic specification will contribute to the dwelling age-related heteroskedasticity. To accommodate heteroskedasticity by explicitly modeling the residual variance, Goodman and Thibodeau (1995) introduce the iterative generalized least squares (GLS) procedure,
an econometric procedure described by Davidian and Carroll (1987), that models the relationship between dwelling age and the residual variance for a large sample of housing transactions. They demonstrate that residual variance in hedonic house price equation is frequently related to dwelling age and GLS procedure increase efficiency by reducing estimated standard errors. In a subsequent paper, Goodman and Thibodeau (1997) extend their 1995 research by incorporating additional housing structure characteristics into the hedonic specification and verify compelling evidence of dwelling age-induced heteroskedasticity for half the submarkets examined and for all areas combined.

As Case and Shiller (1987) indicate, inherent heteroskedasticity may be also pervasive in repeat sales model. However, although the heteroskedasticity in repeat sales indices has usually been corrected related to holding periods (e.g., Case and Shiller, 1987, 1989; Quigley, 1995), addressing dwelling age related heteroskedasticity that is likely to occur in the repeat sales model has been ignored. Following their previous works, Goodman and Thibodeau (1998) employ the GLS procedure in repeat sales equations and identify the determinants of heteroskedasticity related both to dwelling age and to time between sales. Using data for nearly 2,000 repeat sales in Dallas, they conclude that repeat sales heteroskedasticity is a function of both dwelling age and time between sales. The results show that after controlling the dwelling-age-related heteroskedasticity by iterative GLS procedure more robust parameter estimates have been obtained.
2.5 An evaluation of sample selectivity

Change of depreciation rate could be caused either by the changing depreciation of the population from which the sample is drawn, or a change in age of the sample itself. Hence, as Dixon et al. (1999) suggest, a “static” sample should be substituted by a “moving” sample for explaining differences in the pace of the depreciation rather than differences in age profile of the sample. However, the problem is that although limiting the analysis to dwellings built during some specific periods or regions may ensure comparability of the data, how to choose the ideal periods or regions can obviously have no unique claim on correctness (Shilling et al. 1991). Moreover, if we take the need to allow for housing retirement into consideration, any sample of used house prices will be “censored” by the retirement process, since houses taken out of service are no longer available for sampling. As defective properties being atypical of the whole portfolio or “lemons”, it may also well be an issue for a building depreciation study (see Case et al., 1991).

Obviously, such potential sampling biases will yield a distorted picture of the average depreciation experience of the cohort. Based on Davies’s (1977) estimation on the rate of conversion and demolition for London, Chinloy (1979b, 1980) additionally use a sample randomly drawn from the house population to avoid the sample selection problem. To correct this bias, Hulten and Wykoff (1981, 1996) employ the survival probability approach. Due to lack of information on net scrap values, they assume that a nil value for non-surviving assets so the latter value is nil. The results indicate that
the actual estimates are sensitive to the form of the retirement distribution which has always been arbitrarily chosen. Lee et al. (2005) avoid the sampling bias by restricting their sample to the units whose age in no greater than the age of the oldest apartment unit or complex pursuing redevelopment in their data set. Unfortunately, this method only approximately allows the use of a random sample to estimate the probability that each observation has complete data. Additionally, as McMillen and Thorsnes (2008) and McMillen (2008) show, the quantile approach could reduce the sample selection issue that has plagued the repeated sales estimators. Further, with better access to data, the censored likelihood approach proposed by Heckman (1979) may serve as a possible useful direction for further research.1

2.6 Vintage effects

Vintage effects are inextricably tied to the age of a home. Such effects due to some unmeasured housing characteristic correlated with the year that a dwelling was built undoubtedly tend to work in such a way that the effect of age alone is often obscured in casual observation. For example, if new technology is superior then the arrival of new and better vintages depresses the price of old vintages lacking the enhancement (Dixon et al., 1999). Under such circumstances, even if there were no depreciation, the price of older houses built worse than today’s houses would be lower (Margolis, 1982). Therefore, the change in the price of a house should be adjusted for vintage effects to isolate the true price trend. This is unfortunate because very few efforts have been undertaken to indicate problems of misspecification if vintage effects is ignored.
Consequently, if construction quality could not be controlled, depreciation may be overstated.

In the early studies, independent effects of age, date and vintage had been doubted to be separately identified econometrically (e.g., Hall, 1968). This identification problem exists in part because age and vintage are collinear in a cross section. Randolph (1988b) indicates that under the assumption of stable unobserved quality and with right measures of neighborhood characteristics in the estimated regression, it is possible to identify depreciation. Within the context of a specific application to rental housing data from the Annual Housing Survey for the Detroit SMSA, he shows the possibility of identifying depreciation if it is assumed either those vintage effects are negligible or that unmeasured quality is stable. The disadvantage of this method is that if the restrictions are incorrect all the estimated coefficients will become bias.

Englund et al. (1998) propose an improved methodology for distinguishing between vintage and depreciation in their effects upon housing prices. This method is reminiscent of the recently emerged hybrid techniques, which allows one to take advantage of the information contained in transactions which occur only once (Case and Quigley, 1991; Quigley, 1995). In contrast to the hybrid method, the estimator proposed by them incorporates a conventional autoregressive structure rather than the more cumbersome assumption of a random walk in housing prices. They argue that this error structure is sufficient to provide an estimate of the independent effects of age. The empirical findings based on a complete and detailed body of housing market
information about Sweden indicate that there is a small positive vintage effect and the unmeasured vintage-related aspects of dwellings in their sample are probably quite subtle. Fletcher et al. (2000) and Smith (2004) model the vintage effects by including dummy variables for properties of particular vintages in their hedonic models. Although the authors mention that the cutoff date is not selected arbitrarily, in which case the extreme categories would constitute a systematically smaller proportion of the sample than would result from a more objective measure.

2.7 Effect of maintenance

Constant housing price indexes should track the rate of price appreciation free from depreciation and maintenance effects that are concurrent. Unfortunately, for the challenge of obtaining data on home maintenance, most of the former studies exclude major property improvement due to ordinary and reasonable renovation. Yet relatively little has been done to measuring the existence and importance of the connections between maintenance expenditures and depreciation rates. By examining maintenance separately, Chinloy (1980) incorporates an explicit treatment of expenditure on repair and shows the erroneous estimates of depreciation rates from failure to include a model of maintenance for housing where these phenomena are important. However, his model and results depend on various assumptions about housing market. A study by Knight and Sirmans (1996) is among the first to employ a time-series of data that permits us to explore not only the impact of maintenance on individual house value, but also the extent to which extraordinary maintenance may affect the precision of
house price index estimation. Based on the broker remarks, they assign each transaction to one of four maintenance categories which serve as their proxy for the level of maintenance characterizing the home. The results show that maintenance activity is important in explaining individual house price, but little effect is seen at the aggregate level when maintenance is omitted from the variable specification for hedonic price indexes both with the data from a large neighborhood in Baton Rouge, Louisiana over the period 1985-1993 and simulations. Due to the possible noise in their maintenance variable that results from different perceptions among agents in assigning remarks about the condition of property, they suggest that further investigation performed with better data should be warranted.

Blessed with a panel of single family home sales from the American Housing Survey, “the only major survey that provides detailed information on home maintenance”, Harding et al. (2007) estimate the rate at which housing capital depreciates gross of maintenance by three-stage squares (3SLS) treating each of the maintenance terms as endogenous and using structural attributes of housing as instruments. Within their model, the annual rate at which a portfolio of homes appreciate is equal to the rate of house price inflation less age-related depreciation plus the annual contribution of maintenance to the rate of house price appreciation. The result shows that the tendency of OLS to understate the role of depreciation more than offsets the impact of understating the role of maintenance. This causes the 3SLS inflation index to rise at a faster rate than the OLS inflation index.
A useful approach, quantile regression, for treating the bias in repeat-sales price indexes from unobserved renovation or remodels between sales has been developed by McMillen and Thorsnes (2006). As renovations generally increase a home’s market value, the dependent variable corresponding to a renovated house (i.e., the difference in the log price between pairs of repeat transactions) will be relatively large. In other words, when renovations are unobserved, the residual on a renovated house appears in the upper half of the residual distribution which would not contaminate other points in the distribution. Because the quantile estimator places less weight on outliers than do least squares estimators, targeting quantiles from the middle of the error distribution will reduce the effects of outliers. Using the data for Chicago 1993-2002, they suggest that maintenance account for much of the rapid rise in house prices, and that median-based quantile estimators produce a more accurate view of the price performance of a representative house.

2.8 Possible additional effects

Recent studies argue that additional effects can be incorporated into the interpretation of the age variable. Appealing to the rational expectations framework developed by Campbell and Shiller (1989) and Campbell (1990), Clapp and Giaccotto (1998) develop a new theory of the age coefficient in hedonic models. They suggest that the age coefficient in a hedonic regression measures not only the current level of depreciation but also the present value of present and future expected differences in demand of different properties. Using Fairfax County data, they find that the time
series of age coefficients is non-stationary. The empirical findings corroborate their theoretical analysis that the age coefficient is measuring more than just depreciation, such as shifts in the demand for older housing, over the period of study.

Previous estimates are for the flow of services or stock prices of land and structure combined by literally interpreting the impact of age on the entire value of that real estate. As structures deteriorate but land does not, at least in urban uses, the estimated age-value profile is flatter than the true profile for structures. Therefore, more complete model of housing depreciation requires the separation of structures and land. One of the first studies in this area develops a land-value-adjusted hedonic modeling structure to support the hypothesis that there is a significant difference in the estimated depreciation applied to real property sales observations when the value of the land is extracted from the overall value (Smith, 2004). The results indicate that the land value significantly alters the estimate of depreciation and result in a downward bias towards more valuable properties. The non-stationarity exhibited in the rate of depreciation has also been observed in this study, which implies that empirical work should estimate a time series of age coefficients.

Some works have been motivated by observations on other countries besides the USA. Chau et al. (2005) argue that building age not only accounts for physical deterioration, functional obsolescence, and the lemon effects, but also for the interest rates in dealing with leasehold land. Through analyzing the residential market in Hong Kong and comparing the proposed age-adjusted index to the traditional repeat sales index,
they prove that age bears a close relationship with the term to maturity for leasehold properties. Lee et al. (2005) indicate that due to the shortage of residential land in Korea, city expansion eventually requires redevelopment for further growth which implies that the ageing of the house may not only cause depreciation, but also increase the possibility of redevelopment. In this case, the age variables in hedonic equations will reflect both depreciation effect and effect of redevelopment. In order to test this hypothesis, they incorporate the probability of redevelopment estimated by a probit model as one of the explanatory variables in the apartment price estimation equation. The results based on 3,474 observations on apartments in Seoul in 2001 show that depreciation rates are higher in earlier years and later years of age and are lower in the middle years of age.

The above overview highlights an important issue that although measuring and isolating the age effects, inflation and obsolescence are difficult, the resulting age-index could provide a clearer picture of the performance of a property market. However, a striking feature with the results of all above studies is their variability. This variation seems to be due to differences in definition of depreciation employed, data used in the models and the conditions of the real estate market examined at the time of the study. Imperfect model specification, lacking data for controlling for construction quality and unadjusted sample selection bias may result in incorrect estimation and interpretation of key regression parameters, particularly the housing unit age coefficient, which has been interpreted as an estimation of depreciation.
Further, because the data used in many papers are limited in a particular jurisdiction, generalization of their results to explain national-wide variations of housing prices would be difficult. Additionally, the new findings from Hong Kong and Korea imply that extending our analysis to other countries may help us better understand the role of depreciation and maintenance in calculations of the stock and flow of properties in the economy.

3. Considering index revision volatility

Index stability is subject to either the updated method used to incorporate data or the arrival of new information affecting all of the parameter estimates. The later type, tendency of previously estimated values for prior quarters to change with new release, has been called “revision volatility” by Wang and Zorn (1997). As revision volatility does not only adversely affect practitioners, but also affects the integrity of a market in which transactions are settled with reference to the calculated values of these indexes, house price indexes might then be revised infrequently if desired. Unfortunately, the fact is that revision volatility is a necessary aspect of any price index (Shiller, 1993). Due to the importance of producing reliable and replicable indexes of housing prices, many academic studies seek to identify the incidence, mechanics and magnitude or systematic bias of house price revisions as new information becomes available.

3.1 Are revised price indexes more efficient?

The revision issue has been considered from a positive point of view in the early
studies. As noted by Bailey et al. (1963) and Shiller (1991), the repeat sales model utilizing information about the price index for earlier periods that is contained in sales prices in later period may result in increased efficiency in the estimator. Because new transaction will provide additional information about changes in the price level beyond that obtained from the previous sample, ideal index revisions should be random in nature and center around zero if the index is an unbiased estimate of the real price change.

However, retroactively volatile prices, highly imbalanced transactions and sample selection bias always create practical problems (Wang and Zorn, 1997). First, the linked nature of the paired sales of dwellings might introduce systematic changes, because of the first transaction when it takes place at a date situated before the last period of the index from the addition of new paired-sales to the sample. To examine this hypothesis, Hoesli et al., (1997) propose to compute the number of runs as a test for randomness in the revisions. However, their findings are consistent with the notion that the revised index is a more efficient estimator of the price level. Second, the instability of the index often increases as the total number of observations drop (Abraham and Schauman, 1991). As the first sales always tend to occur disproportionately towards the beginning of the period, and second sales disproportionately towards the end of the period, a tendency for large revisions may occur for the years toward the beginning and end to the sample. Third, in the presence of any sample selection bias such revisions to past indices do also present a theoretical
problem so much as a practical one. Clapp et al. (1991) provide evidence on the significant revisions to repeat sales estimates and sometimes the poor performance in estimating short-term price changes which may stem from the possibility of market price manipulation by speculators towards settlement dates. Steele and Goy (1997) show that where the bias in the change in prices is constant, the bias in the index will vary from one quarter to the next. They hypothesize that the source of latter bias is the presence of opportune traders who tend to hold properties for a short time. Clapp and Giaccotto (1999) further highlight the role of such flips (i.e., properties with very short time between sales) and conclude that the presence of flips is the major cause of revision in most repeat-sales indexes.

3.2 Are systematic revisions significant?

Because of selective use of the available data on sales, variations in the way that new information is incorporated into index estimates and different assumptions about the bundle of attributes and their implicit prices, distinct methods of index construction are not equally exposed to volatility in revision. As additional information on paired sales is used, they will not only augment sales prices in the new period, but also revise the entire history of index values over time, repeat sales indexes seem particularly subject to never ending revision (Deng and Quigley, 2008). The already subsample of all sales, on the other hand, will be further reduced in the most recent time period. This serves as the major reason why revisions are likely more important for the repeat-sales method than for the hedonic method. Compared with repeat sales index,
hedonic-based index is less subject to revision as it uses all sales data as they become available. Clapp and Giaccotto (1999) note that a hedonic method based on a separate regression each time period will not be subject to revision, assuming that the definition of a standard property remains constant.

Bourassa et al. (2006), using the data for regions and cities in New Zealand, provide evidence that the longitudinal hedonic model which includes all periods in a single estimation with time dummy variables should be sensitive to new data being added, but to a lesser degree than repeat sales indexes. Relying upon data of all sales of owner-occupied single family dwellings in Sweden over a 19-year period, Clapham et al. (2006) prove that if the assumption of constant implicit prices in hedonic-based index is maintained, revision will present more accurate estimation as standard errors decrease. In practice, however, this assumption does not typically hold with precision. Hence, if this assumption is violated, index revision will represent an evolving bias resulting from model misspecification.

3.3 Magnitude to Revision Errors

Obviously, this termed revision volatility has induced problems to the interpretability of the index, as additional information changes the past values of the index. Thus, the degree to which the earlier numbers must be revised can have a bearing on the value of index for predictions and the real estate derivatives pricing. For this reason, many studies examine the magnitude to revision errors, with some of warning the revisions are “not inconsequential for the settlement process”. In such studies, index volatility
has always been measured as changes in estimated price levels that results from the remeasurement that occurs when new information is revealed in the form of additional sales. Abraham and Schauman (1991) compare the repeat sales index developed at Freddie Mac with other published indexes at the level of the U.S. Census Region. They find that expansion of the dataset will lead to subsequent downward revision in repeat sales growth rates. That is, index values are found to be lower than previously estimated index values when more transaction data becomes available.

In a more comprehensive study on house index revision, Stephens et al. (1995) examine the sensitivity of annual percentage change in house price index as new data are added to the sample of repeat transaction. They find that almost two-thirds of revisions experience for the nine U.S. census divisions from 1988 to 1992 is negative, and with the negative revisions being of greater average magnitude than the positive changes. The fact that the revision of repeat-sale indexes may be large, asymmetric and more likely with downward revision more prevalent than upward has also been confirmed by Clapp and Giaccotto (1999). The results based on Wilcoxon’s signed-rank test for equality of means and runs test for trends in their analysis both suggests that revisions are not random but systematic. The fact, that revisions based on the Freddie-Fannie data improve efficiency but by relatively small amounts, has been found in this study. Based on Swedish data on all sales of owner-occupied single-family dwellings during 19-year period, Clapham et al. (2006) report that how often revision, either period by period or cumulative, exceeds ideal tolerate levels.
Their finding that the revision in the price level estimates is two to six times greater for the repeat-sales indices relative to the Fisher Ideal Index. They notice that most of the revision occurs in the first ten quarterly estimates and price estimates become more stable thereafter.

3.4 Extent of index revision

Considering the big influence of revision volatility in the light of subsequent information, some recent studies try to systematically analyze the extent of these revisions. Baroni et al. (2008), using a Paris housing dataset, measure the impact of revision on prices, returns and volatility of two indices derived from the Case and Shiller classical methodology (WRS) and a factorial method using a Principal Component Analysis (PCA). Most of their conclusions are consistent with the literature for the US or the Swedish markets. An exception is that cumulative revisions for the PCA index are systematically upwards. Meanwhile, the impact of revision on the volatility, as showed in their analysis, is often lower for the sub-markets than for Paris as a whole. Another study using data on the sales of over 500,000 owner-occupied homes in Netherland also finds the different revision volatility between the overall Netherland market and sub-markets (Jansen et al., 2008). Such facts imply that risk measured by the volatility may be sensitive to regional disaggregation.

Deng and Quigley (2008), using a complete set of the OFHEO housing price index published in 2001 and 2007, analyze the distribution and predictability of revision
volatility over time and across geographic regions. They find that even over a short interval the geographical definitions of metropolitan areas are subject to substantial revision. This suggests the necessity for the agency to preserve geographical identifiers for each sale in its data base and to produce current price index estimates for all geographical configurations of MSAs which have been used in the past. In addition, there is little evidence, as shown in this study, that the index revisions are strongly predictable, either on the basis of lags and serial correlation or on the basis of simple macroeconomic factors.

3.5 Attempts to solve the index revision volatility problem

Seeking to shed some light on how to solve the index revision volatility problem is a common area of exploration in the literature. Although many authors propose some practical solutions to the revision process, they ultimately conclude that home price futures markets might better be served by hedonic price indices. Because a mathematical fix to the revision problem is neither easy nor likely to be correct, to get a more accurate and stable measurement of price levels based on repeat sales model, some studies suggest that very short holds should be eliminated from the data set or observations from housing types that are sampled “too frequently” should be downweighted (Gatzlaff and Ling, 1994; Clapp and Giaccotto, 1999). However, the question is if there is something that the current model cannot provide consistent explanations. Because blindly removing outliers will incur hazards (see MacDonald and Robinson 1985), it seems not enough to identify and delete such flips. We need to
examine the flips to determine if there is any systematic pattern to them. Others provide candidates for use by government agencies in developing house price indexes, in particular, if the cost and difficulty of index construction is highly concerned. For example, Bourassa et al. (2006) show that sale price appraisal ratio (SPAR) index not requiring detailed databases of property attributes is comparable to multiple equation hedonic indexes in terms of index revision.

A particularly attractive feature of the SPAR approach is that it does not require any historical index revisions when new data are added. Gao and Wang (2007) suggest using consecutive transactions to do the pairing of house prices when multiple transactions occur for individual houses. The multiple transactions model proposed in their paper with a panel data is shown to keep the same features of the repeat sales model and own nearly zero index revision volatility. Because the relative strengths and weakness of the various types of indexes should be assessed according to more than one criterion, increasing the attractiveness of an index according to revision stability may limit its appeal with respect to another. In other words, the importance of changes in historical values will depend on the use to which the index is put. For example, while revision issue is not so important when dealing with the time series of price returns, it really important on estimating the index price level and dealing with the forecasting ability of the model (Baroni et al., 2007). In addition, if new data do not materially differ from what was obtained before, and enters dataset used for constructing the house price indexes subsequently in a random fashion, the revision is
causes should be pure noise.

4. Controlling for the impact of illiquidity

Real estate market always goes through “hot” and “cold” periods, and the time lag between real estate is placed on the market and when it is sold has frequently been substantial. Thus, it displays characteristics of illiquidity which varies dramatically over time and different states of nature. This intertemporal variation in the ease of selling an asset suggests that changes or shocks to the fundamental value of housing are not transmitted solely through market prices, but through market liquidity as well (Krainer, 2001). Unfortunately, the traditional housing price index construction methods borrowing in a naïve fashion from finance theory to estimate indices of property values over time ignore real estate illiquidity (Fisher et al., 2003).

In presence of seller reserves, price indexes based on such traditional methods may provide inadequate or even misleading information regarding market demand and supply (Goetzmann and Liang, 2006). Moreover, as Lin and Vandell (2007) emphasize, traditional indices of using the \textit{ex post} variance as a proxy for the \textit{ex ante} variance can seriously underestimate real estate risk, especially when the expected marketing period is high and the holding period is relatively short. Unlike the presumption in the market value concept, liquidity is not held constant across time in a transaction price series. Yet researchers, policymakers, regulators and investors care not only about the average price at which transactions are consummated, but also about how long it takes and how easy it is to sell at those prices. Therefore, it seems
imperative for constructing a constant-liquidity index which reflects the price changes that hold constant the degree of difficulty in selling assets in the market.

4.1 Frequency of transaction and house price modeling

Understanding transaction frequencies affecting real estate price levels and their movements has been a common area in economic literature. A few of papers provide micro-analytic insights into what constitutes the fundamental essence of the time-variant liquidity characterizing the transaction process in real estate. Haurin (1988) applies the standard search model to sellers and argues that the higher the variance of the distribution of offers for a house, the longer the expected wait for an acceptable offer. Stein (1995) develops a model where sales volume is cyclical due to the down payment effects and the market interaction among credit-constrained households with different ages. Glower et al. (1998) highlight the importance of seller heterogeneity in the search process for housing. Their finding based on a telephone survey indicates that home sellers who are motivated to sell quickly will set a lower list price, have a lower reservation price and accept earlier. While such papers do not formally model liquidity, their results have direct implications for the house price changes, as well as for the correlation between prices and trading volume.

In order to show how house price, liquidity and sales volume depend on the value of the housing service flow, Krainer (2001) develops a search-theoretical, rational agent model in which house prices and liquidity are derived from the maximizing behavior of both buyers and sellers. In his model, both prices and liquidity are jointly
determined and adjust to reflect changes in the value of real estate good. Absence of smoothly functioning rental market has also been shown in this study to increase the frictions giving rise to strong and falling markets. Fisher et al. (2004), using data on sold and unsold properties from the underlying database of the National Council of Real Estate Investment Fiduciaries (NCREIF) index, examine the primary factors that explain intertemporal variation in transaction frequency of institution-grade commercial real estate. The results show that market condition, ownership characteristics and property features appear to play significant, independent and approximately equivalent roles in determining transaction frequency.

Two of the recent studies provide new insights into the underlying causes of time variation of liquidity in the private real estate market. Lin and Vandell (2007) theoretically delve into the state-varying liquidity analysis. They argue that the most important aspects defining real estate illiquidity are the time required for sale and the uncertainty of the marketing period. Based on three different classes of models, Clayton et al. (2008) derive alternative testable hypothesis for why seller valuations lag changes in the market. Since the annual frequency may be too low to detect the dynamics of price and liquidity, they explore this consideration in more detail with an analysis of quarterly frequency data. Their empirical findings based on the NCREIF data sample are generally consistent with models of optimal valuation with rational updating and provide support for the opportunity cost approach. Little evidence suggests that high liquidity results from the presence of overconfident traders.
4.2 Short holds and bias in the housing price index

Empirically, several studies identify the sample selection bias associated with differences in the ease of selling a property over time (e.g., Haurin and Hendershott, 1991; Clapp and Giaccotto, 1992; Steel and Goy, 1997; Munneke and Slade, 2001). The consensus among such studies is that the sample sold frequently in the sample period will tend to have a shorter holding period than the sample as a whole. Unfortunately, few studies have further analyzed intertemporal variations in the ease of a house sale. Studies by Gatzlaff and Haurin (1997, 1998) are among the first to point out that if variations in economics conditions affect offer and reservation prices, then price indices based on a sample composed only of the sold houses do not hold constant liquidity in the market. In this sense, the transaction prices may relate to the average of buyers’ or sellers’ valuations, but do not reflect the spread between them or fail to capture changes in market demand and supply. These transactions would be consummated only at appraised values of the targeted property and a perfectly observed and measured transaction price index would be the same as an appraisal-based index (Fisher et al., 1994).

Using a simple econometric model of trading in illiquid markets, Goetzmann and Liang (2006) show that both hedonic and repeat sale indices are biased measures of true market values unless the effects of seller reserves are controlled, and without controlling for seller reserves the raw index will smooth out drops and volatility of returns may be underestimated. In particular, if market valuation leads private
valuation in a dynamic setting, both repeat sale and hedonic methods has been proven to underestimate the volatility of actual market indices. This Lin and Vandell (2007) have termed liquidation bias as the value of the potentially constrained time lag between the event of sale and event of placing the property on the market. They use simulations to validate the notion that there is a positive relationship between the expected marketing period and liquidation bias. Given the often longer average marketing periods than those in residential markets, their finding also suggests that liquidation bias in the commercial market could be an even more serious problem than in the residential market when using the return and volatility estimated from transaction prices.

4.3 Attempts to control the impact of liquidity

It is safe to say that an accurate index should not only represent typical transaction prices prevailing among consummated deals in the market, but also reflect varying ease or ability to sell properties across time. While widely understood as important, it is only recently that researchers have attempted to quantitatively control for this problem in the construction of house price indexes. In the context of a private asset market model characterized by pro-cyclical variable volume of trading, Fisher et al. (2003), hereafter FGGH (2003), define a “constant-liquidity value” of a property, as the value assuming no change in the expected time on the market (TOM), and derive a constant liquidity version of the NCREIF property value index. The key insight from their theoretical model describing the functioning of a real estate market is that the
market is characterized by two jointly determined statistics: transaction price and trading volume. The transaction price reflects a type of “average” of buyer and seller valuations, while the transaction volume varies over time only in response to the differential movement between the buyer and seller reservation price distribution. Their empirical results indicate that the mean constant-liquidity price is higher than the mean variable-liquidity transaction price during up markets and lower than the observed transaction average in down markets.

Based on these findings, Fisher et al. (2003, 2007) argue that the constant-liquidity values present a more complete of metric of the changes in market conditions in the private asset market, and enables estimation of empirically based constant liquidity value indices of market capital returns. This argument has been corroborated by Clayton et al. (2008) using their finding that constant liquidity and observed transaction prices are tightly connected over long-run periods, but can be divergent in the short-run. As noted above, the FGGH (2003) is of interest for purpose of tracing out movements on the demand side of the market.

Goetzmann and Liang (2006), hereafter GL (2006), on the other hand, define a constant-liquidity index relative to the seller’s distribution of reservation prices, and construct a reserve-conditional unbiased index for the Los Angeles housing market, which substantially differs from a traditional repeat sale index. In their view, the marketing period variations over time represent proxies for variations in the degree of liquidity present in the real estate market at any period of time. They argue that a
A sensible definition of housing price index should track not only the market value, which reflects the market demand, but also the seller reserves, which reflect the spread between demand and supply as well as market liquidity. Their paper documents the negative relationship between sellers’ reserve ratio and trading volume and the positive relationship between sellers’ reserve ratio and the upward bias of the index level from the market value. Furthermore, based on a three-step approach to mitigate the liquidation bias, the reserve conditional unbiased index by GL (2006) has been presented to be less data-demanding than the Heckman approach, easier to calculate than the maximum likelihood and economically equivalent to the constant-liquidity index by FGGH (2003). In words, the accuracy of estimation has been achieved by eliminating the bias due to the time-varying seller reserve, whereas their reserve-conditional unbiased index reflects drops more immediately.

In summation, the findings in the above literature suggest how variations in a seller’s cost of holding a property affect the optimal type of contract, the transaction price and the TOM. Because the information between buyers and sellers is asymmetric, housing market illiquidity is really caused by a selection problem. The micro-analytic foundations of illiquidity proposed in the above papers help us to understand the risk-premium puzzle in real estate and aid in the development of constant-liquidity value by a broad definition. Because many papers are primarily intended to derive proper econometric bias corrections, the theoretical models in such papers are solely for the purpose of illustration. Thus, certain simplifying assumptions should bear
further scrutiny throughout their analysis for the sake of analytical tractability. While the liquidity adjusted price index more completely tracks the changes in the condition of the private market over time, which facilitates comparing the risk across the asset classes, the merit comes at a cost. For example, while the model developed by Fisher et al. (2003) is an excellent technique, its onerous data requirements limit the datasets that can be used, and it is necessarily constrained by a number of measurable items that can be statistically found to be important.

Compared with FGHH technique to construct constant liquidity index, the GL (2006) method is less data-demanding and more easily manipulated. However, this new econometric procedure to mitigate the bias used a few strong assumptions which may not be appropriate under all market conditions. Hence, challenges remain regarding how to eliminate the bias under weaker assumptions. As noted in Clayton et al. (2008) a higher temporal partition tends to increase serial correlation and makes the index lag behind the true price movement. Therefore, a narrower frequency may be more relevant in terms of capturing market dynamics. However, a narrow temporal partition may be heavily influenced by noise, hence overstating variance and underestimating serial correlation (Geltner, 1997). In other words, narrower temporal disaggregation is done at the expense of taking on more “noise”. Since the time intervals are of arbitrary length and are not subject to choice by the observer in many cases, it is important to study how temporal aggregation works. Finally, under a broader definition of constant liquidity or given real estate as an asset class in a mixed-asset
portfolio, the optional definition of liquidity as TOM and trading volume may not be sufficient to adjust real estate price movements to represent full liquidity (see Lin and Vandell, 2007).

5. Criteria for a good price index

Having set out some concerns in the above section, it is now necessary to put forward some criteria for choosing among indices in general.

5.1 Criteria based on in-sample variation

Based on “reliable” criteria, a number of articles have appeared in the real estate literature comparing the various methodologies. In the earlier studies, the efficiency of models has been often compared by how well they explain the in-sample variation in house prices (e.g., Palmquist, 1980; Mark and Goldberg, 1984; Can, 1990; Case et al., 1991; Goetzmann, 1992; Pace and Gilley, 1997). The mean square error (MSE), estimated standard deviation of the residuals (SD) and width of a 95% confidence interval drawn around the predicted transaction price have always been used to measure the validity of estimation procedures. According to the theory behind the housing price indices, the first and second should be lower for a more efficient index. Because the third offers the best compromise between a high level of confidence on the one hand and a high level of accuracy on the other, it becomes the most commonly used value. Related to the topic of confidence intervals is the number of observations needed to obtain an accurate estimate. Jansen et al. (2008) shows that the number of pairs of repeat sales needed to calculate an accurate index, which is quite different for
the various segmentations.

Besides the criteria mentioned above, Case and Shiller (1989) suggest that signal-to-noise ratio is useful to assess the reliability of the comprehensive indices. This ratio is calculated by dividing the standard deviation of the price changes in the index by the mean of standard error of the time dummy parameters. Hence, larger ratios of this relative measure reveal a higher degree of index precision. Later, Gatzlaff and Ling (1994), Gatzlaff and Haurin (1997) and Munneke and Slade (2001) also use this precision measures to examine the reliability of the respective indices. Besides, Case et al. (1991, 1997) compute the square of the correlation coefficient between actual transaction prices and the transaction prices predicted by their model to assist evaluation of the model performance. Clapp and Giaccotto (1999) borrow an idea from time-series analysis by examining the sum and difference of the true and estimated indexes to test if the reduction in variance statistically significant. In addition, Goetzmann (1992) identifies which method performs better according to the amount of variance in the actual New York Stock Exchange index explained by seven different repeat sales estimators. Pace et al. (1998) and Sun et al. (2005) assess the model performance by comparing their ability in reducing spatial and temporal autocorrelations across models.

5.2 Criteria based on out-of-sample excise

Another way to judge the accuracy of each method is to subject them to an out-of-sample excise, This test has always been based on aggregate market indices
and model parameters. Since this empirical method is conducted for a short time horizon, it focuses on the accuracy of forecasts at the individual house level rather than on time series properties of the forecasts. In general, the out-of-sample prediction analysis can be designed through two different procedures. Some researchers use the house transaction data in earlier periods as the estimation sample and use the house transaction data in the last six or five periods as the holdout sample (e.g., Basu and Thibodeau, 1998; Sun et al., 2005). Alternatively, Pace and Gilley (1998) and Gao and Wang (2007) employ the rotation of estimation and holdout samples. Statistically, the quality of the out-of-sample performance has always been measured by the mean square prediction error (MSPE), standard deviation of the prediction errors (SD) and mean absolute prediction error (MAPE).

In forecasting literature, criteria for accurate forecasts also include a battery of preferred distributional properties for prediction errors; that is, predication errors should have low means, skewness and kurtosis. Some recent researches add to earlier studies of price indices comparison by using various transformations of the mean percentage error to test model performance. Applying a test statistic developed by Diebold and Mariano (1991), Crone and Voith (1992) determine in a pairwise fashion whether the predictive accuracy of the estimated derived from each method significantly differs. Dubin (1998) and Clapp and Giaccotto (2002) employ the Theil (1985) inequality coefficient and a statistic first proposed by Granger and Newbold (1986) to rank forecast performance of their targeted approaches. Unfortunately,
some models can not be fitted using likelihood methods because neither estimates of effects (e.g., spatial effects, temporal effects and interaction effects) nor their standard error needs exist. Gelfand et al. (1998), adopting the criterion presented by Laud and Ibrahim (1995) and further developed by Gelfand and Ghosh (1998), judges the predictive performance of their models in terms of their ability to predict a replicate of the data while still being faithful to the observed sample. Such findings also imply that the model comparison approach adopted for penalizing both underfitting and overfitting might be objectively driven by the data.

5.3 Ideal benchmark for comparison among indices
Since the “true” constant quality house price index is unobservable, many studies used to employ a conventional index as a benchmark against which to compare (e.g., Gatzlaff and Ling, 1994; Anglin and Genay, 1996; Meese and Wallace, 1997; Schwann, 1998). In these studies, the benchmark index serves as a reference point used to qualify the relative performance of different methods. In particular, (1) how well changes in each of the alternative indices explain changes in the benchmark indices and (2) whether changes in the alternatives provide unbiased estimates of changes in the benchmarks are the focus of these studies. To examine the time path of each index for accuracy relative to the benchmark index, some performance criteria have been often employed: (1) the root mean squared error (RMSE) of the log index from the true log index, (2) the number of correctly identifies turning points in the series, (3) the number of periods the log index lies outside the certain confidence
bounds of the true log index. The RMSE measures the overall accuracy of the estimation procedure in capturing the true log index.

The second and third criteria gauge the potential making “wrong” statistical inferences. Gatzlaff and Ling (1994) employ regression analysis to further investigate the relationships between their benchmark indices and the alternatives. In order to avoid some measurement errors resulting in inconsistent estimation, they use the corresponding change in the alternative index as a left-hand-side variable to measure whether it provides an unbiased estimate of the benchmark index. Prasad and Richards (2008) use the “trend” change in prices as a proxy of underlying housing price movements. This proxy is obtained by constructing a measure of trend growth for each measure and then averaging these trends. They assume that the larger the deviation from trend, the less information the series is about the underlying state of the housing market.

5.4 A much more noise index vs a more smoothing one

For the above-mentioned analyses, house price indexes should possess some specific qualities. Perhaps most notably, both academicians and practitioners have wondered about the extent to which the indices presents a “smoothed” or “lagged” representation of asset price changes in the property market that the index is often assume to represent. Previous studies on volatility have related mainly to the smoothing debate (e.g., Geltner, 1991; Lai and Wang, 1998; Clayton et al., 2001; Cho et al., 2003). In general, a bad price index has always been assumed to exhibit much
less volatility than others and at times has failed to register movements in real estate values that were widely perceived by market participants. In contrast, an ideal index can not only tells us what the duration and amplitudes of the historical cycles in various segments of the markets, but provides a leading indicator of market values will soon be doing. First-order autocorrelation is another source of information on the second moment of property returns. Positive first-order autocorrelation in annual returns implies “inertia”, which suggests a partial response of market prices to the arrival of information, or lack of information efficiency. However, each measure has a potential source of noise associated with it. Noise from compositional effects may contribute volatility to housing price movements rather than being indicative of true price trends in the housing market (Gatzlaff and Geltner, 1998).

Whether the first-order autocorrelation is a desirable property for a real estate price index is also an open question as the literature on the serial correlation of house price appreciation contains contradictory evidence (e.g., Gunterman and Smith, 1987; Gatzlaff, 1994). Schwann (1998) finds that imprecision may be manifest in a much more volatile index; that is, higher volatility may partially reflect stronger “noise” (spurious volatility), which biases upward the estimated volatility of the index returns and injects a negative bias into the first-order autocorrelation of the returns. On the other hand, “smoothing” injects positive bias into the first-order autocorrelation of the return series, and results in the index value changes lagging in time behind the true underlying housing market value changes. Therefore, the interaction of the two
sources of error which tend to offset and mask one another complicates the interpretation of index return second moments (Gentler, 1997). Additionally, much of the debate about smoothing has occurred in the context of comparing the real estate returns with those of publicly traded stocks and bonds, to arrive at an apple to apples metric of risk (Goetzmann, 1992; Case et al., 1997). However, as Gatzlaff and Geltner (1998) point out, due to the difference between the property market and stock or bond market, existing returns data for this asset class are inadequate for comparison pricing with market-determined stock and bond returns. In other words, one is still comparing apples to oranges.

5.5 Scientific sampling as the basis of a quality research index

It is important to note that the actual fit to the true underlying housing price index does not only depend on the comparison results based on the criteria mentioned above, but also on the sample size and what the aggregation level is. So it is necessary to know how its accuracy varies with parameters such as number of properties in the sample and size of interval estimated. Schwann (1998) investigates whether real estate price indexes retain the greater fidelity in thinly traded markets. His findings confirm the expectation that the precision of the index will drop in thinly traded markets. In fact, non-randomness of the data being unrepresentative of the universe of property values in the locality is a major concern in the construction of housing indices (e.g., Haurin and Hendershott, 1991; Case et al., 1991, 1997; Gatzlaff and Haurin, 1997, 1998, and many others). Since samples are incomplete (i.e., nonmovers are excluded),
by ignoring the problem of sample selection, price indexes based on sales transactions might misrepresent price trends in the overall market. Thus, whenever possible, the sensitivity analysis on the base of selectivity of subsample has always been made (Clapp and Giaccotto, 1992; DiPasquale and Somerville, 1995; Steele and Goy, 1997).

Varying temporal aggregation has also been used as a test of robustness of the estimates. Englund et al. (1999) and Sommervoll (2006) shows the instability of the index estimates with respect to small changes of temporal aggregation, and this instability is higher than expected from looking at R squared or estimated index covariance alone. Goetzmann (1992) find from his sensitivity analysis that more data do not compensate well for increasing the number of intervals to be estimated. As noted above, housing price indexes on geographically aggregated data implicitly accept a representative market approach to modeling. It means that even if the in-sample fit is good with geographically aggregated data, there is no guarantee that postsample observations will continue the distributional pattern of the past and continue to fit the aggregate data. Therefore, Dombrow et al. (1997) and Goodman (1998) argue that the robustness of the results should be tested based on different spatial aggregation levels. In words, the accuracy of each method should also be judged by its sensitivity to changes in sample size and selection method.

5.6 Easiness of Administration
Lastly, one criterion pertains to whether the index method is administratively simple and does not rely on expensive and awkward sampling. In fact, the choice among
different methods for estimating changes in housing prices will always involve a trade-off among various competing concerns including the ease of construction, the extent to which the methodology controls for quality, and particularly the extensiveness of data requirement (Bourassa et al., 2006; Prasad and Richards, 2008). Thus, an ideal index should be substantially easier than others to construct and not require unduly complex estimation technique or huge amounts of data.\textsuperscript{10} Ironically, the somewhat surprising finding is that there appears to be relatively little difference between indexes in terms of overall performance. In some studies, the indices appear broadly similar, and are contemporaneously correlated as over 60\% in the annual price changes (\textit{e.g.}, Abraham and Schauman, 1991; Clapp and Giaccotto, 1992; Gatzlaff and Ling, 1994; Can, 1997).

Given the potential real-time difficulties with the more sophisticated methodologies, it becomes worthwhile to consider if some simple measures can be improved to be more useful before reliable estimates can be produced using the more sophisticated regression techniques. A study by Prasad and Richards (2008) is among the first to point out the need for simpler approaches for calculating housing price indexes. Their findings suggest that the growth rates produced by simply methods might also line up closely with estimated based on the sophisticated ones.

In providing a clear metric to evaluate statistical measures of accuracy, this should aid researchers in deciding if a method is appropriate. However, in practice, conflicts may arise when using different statistics (Goetzmann, 1992; Crone and Voith, 1992; Clapp
and Giaccotto, 2002). For example, as noted above, many researches assess model prediction ability using the MSE. However, predictors having the same properties must be applied to these models (Kato, 2008). The reason is that MSE is affected by the efficiency of the predictor. In other words, the difference in MSE may not be accounted for by the difference in the prediction ability itself. Moreover, there seem no indications in the literature on how narrow a confidence interval had to be in order to be described as “accurate”. Nor was there any consensus on the minimum required accuracy of a sample. In summation, it seems that we have no gold standard of which level of accuracy is still acceptable.

Still, there remains a healthy research agenda before any price index will be ready to be adopted for widespread use, along with the ongoing comparisons among such techniques. Finally, a caveat should be noted in conclusion. Due to the costliness of data collection, all indexes are necessarily compromises relative to an “ideal” measure and can be judged relative to one another rather than on an absolute scale. As shown by Wang and Zorn (1997), much of the disagreement over the choice of method is due to unspoken and perhaps previously unrecognized disagreement over targets or aims. In this regard, an index based on transactions in the property market does not get us all the way there. Rather, researchers must choose these in accordance with the needs at hand. Nonetheless, a comparison is instructive and highlights the issues discussed previously about using indices carefully and being aware of the uses to which the index are being put.
Notes
1. To identify and correct sample selection bias, this two-step procedure has been more recently applied to real estate markets in several studies (e.g., Jud and Seaks, 1994; Gatzlaff and Haurin, 1997, 1998; Munneke and Slade, 2001; Fisher et al., 2003, 2007).
2. Depreciation has been defined by Randolph (1988b) to be the change in quality due to generally unmeasured characteristics changing systematically with aging of a housing unit.
3. An outlier is an observation that does not fit the pattern of most data.
4. See Baroni et al. (2007) for a description of this method.
5. “Hot” and “cold” markets have always been defined as degree to which property owners can take advantage of being in the high state of economy (e.g., Krainer, 2001; Clayton et al., 2008).
6. Information criteria such as AIC or the BIC have also been frequently used to panelize for model complexity.
7. It has been regarded as the only available measure of accuracy for nonparametric methods (Crone and Voith, 1992).
8. More specifically, the entire set of observations of house transactions is randomly divided into certain groups with roughly the same number of observations in each group. One group of data is the holdout sample, and the remaining groups are the estimation sample. The holdout sample is rotated throughout all groups. For each rotation of the holdout sample, the market indices and prediction errors are computed.
9. Theil’s U Statistic based on the MSPE is in a similar fashion to the coefficient of determination, $R^2$, from linear regression. See Dubin (1998) for more discussion on the Theil’s U Statistic.
10. Ironically, median price measures have still been widely cited in the press and frequently used in a range of countries by housing lenders, industry bodies, and sometimes government agencies. The reason is presumably that the more advanced techniques require detailed data.
11. This statistic can be used to measure “goodness of fit” between the actual and predicted house price.
12. For example, kring as best linear unbiased prediction in geosciences is a minimum mean squared error statistical procedure for spatial prediction that assigns a differential weight to observations that are spatially closer to the dependent variable’s location (Goldberger, 1962).
13. For example, if the ultimate goal is prediction, a logical means of processing should be to pick what predicts best.

References


McMillen, D. P. 2008. Changes in the Distribution of House Prices over Time: Structural Characteristics,


