The Regional Economics Applications Laboratory (REAL) is a unit in the University of Illinois focusing on the development and use of analytical models for urban and region economic development. The purpose of the Discussion Papers is to circulate intermediate and final results of this research among readers within and outside REAL. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the University of Illinois. All requests and comments should be directed to Geoffrey J. D. Hewings, Director, Regional Economics Applications Laboratory, 607 South Mathews, Urbana, IL, 61801-3671, phone (217) 333-4740, FAX (217) 244-9339.

Web page: www.real.illinois.edu/

Estimating Housing Price Indices for Small Metropolitan Areas (SMSAs)

Esteban Lopez

Minshu Du

REAL 14-T-1 July, 2014

Estimating Housing Price Indices for Small Metropolitan Statistical Areas (SMSAs)

Esteban Lopez[1]
Minshu Du

R | E | A | L
Regional Economics Applications Laboratory
University of Illinois at Urbana-Champaign

Abstract:
In an economy that proved to be highly sensitive to housing prices fluctuations it is at least intriguing why the U.S. has accurate housing price indices for the top 20 Metropolitan Statistical Areas (MSAs) only. This is an important informational gap since in most the states, housing sales in Small Metropolitan Statistical Areas (SMSAs) account for at least 30% of the total state sales. This paper uses matching methods and Fisher Indices to estimate housing price indices (HPIs) using data for Illinois MSAs. The main results suggest that co-movements between Big and Small MSAs are different than the expected.

Key words: Hosing Price Indices, Repeat Sales, Matching Methods.
JEL Classification: R21, R32, O18.

---

[1] Corresponding Author: lopezoch@illinois.edu, 318 Davenport Hall, 607 South Mathews, Urbana, IL, 61801, +1 217 244 7226

Introduction

In the US, the Case-Shiller (CS) Home Price Index is one of the major sources of housing market information. It tracks changes in the value of residential housing both nationally as well as in the top twenty metropolitan regions. The financial crisis that originated in the housing market revealed an urgent need to have more accurate information to track the complexity of how the market operates. In particular, one of the conclusions of the Financial Crisis Inquiry Commission Report (Angelides & Thomas, 2011) is that "widespread failures in financial regulation and supervision proved devastating to the stability of the nation's financial markets," (p. xviii) suggesting that although there were warning signs such as an "unsustainable rise in housing prices," these were ignored. Although much attention has been devoted to the analysis of housing market behavior since then, housing markets other than those in the top twenty metro areas have been highly neglected from analysis to date.

In the case of Illinois for example, the CS price index provides information only for the Chicago Metropolitan Statistical Area (MSA), neglecting Small Metropolitan Statistical Areas (SMSAs), this is those MSAs other than Chicago MSA[2], that account for at least 30% of the total number of housing sales in the state. According to data from the Illinois Realtors Association, in 2013 this 30% accounted for almost 6 billion dollars in housing sales. Additionally, between 2004 and 2010, the growth rate for average housing sale prices in the case of the Chicago MSA was negative (-26%) in contrast to the positive 10% for the case of SMSAs over the same period. Although very basic, these statistics provide enough motivation to justify the need to understand housing markets beyond just the top twenty MSAs.

Reliance on indices only for the largest metropolitan areas may hide important heterogeneity in the rest of the economy. As in many other states, in Illinois there is no information about the behavior of housing prices in SMSAs. In particular, when analyzing housing prices' co-variation over time, there is no evidence of how housing markets within the state move together. This raises the following questions: Do SMSAs follow the Chicago MSA, or is Chicago a particular market separate from the rest? Understanding the co-movements between small and big MSAs in housing prices is important since it provides information about the potential influence of one on the other, providing a better interpretation of housing dynamics over time. Additionally, since the cost of housing is a major component included in the cost of living, housing price indices are essential as an approximate measurement for the cost of living. Therefore, providing housing price indices for lower spatial levels (e.g., smaller MSAs) could be useful indicators to better understand regional disparities in costs of living both within a state and between states.

However, estimating Housing Price Indices (HPIs) for SMSAs is not as straightforward as in the case of big MSAs, where many observations are available. Since fewer sales are available in SMSAs for each period of time, estimating a HPI based on repeated sales –as the CS HPI– is impossible because the imposition of considering only repeated sales will

[2] See Figure 3 for a map showing the classification of MSAs used in this paper.

restrict sample sizes to useless levels. To overcome the data limitation, this paper follows a matching approach that creates comparable samples in terms of housing attributes (Lopez & Aroca, 2012; McMillen, 2012). Instead of requiring the exact same house in two periods of time to be able to control for house attributes and hence create HPIs (repeated sales approach), our approach relaxes this restriction by matching two houses in different time periods as long as they are comparable in terms of their attributes. Matching allows the use of more observations and hence allows the estimation of HPIs for SMSAs.

There are at least two other reasons to advocate for the use of matching estimators in the estimation of HPIs. First, although used in SMSAs because of the lack of better measures, the use of HPIs based on central tendencies (e.g. mean or median prices) is considered inadequate because it does not control for housing attributes. Since housing sales are considered random spatio-temporal occurrences (Dubé & Legros, 2011), each time period will have different houses with different characteristics, making central tendencies dependent on this random pattern and hence not comparable. Second, even if there were enough observations to estimate an HPI based on repeated sales, there is potential selection bias when only selecting a portion of the total sample of housing sales. Although some may argue that houses representing repeated sales are themselves random occurrences, most would argue that it is always better to use all observations and not only those repeated. The use of matching solves these two problems.

Finally, repeat sales' estimates may be subject to bias if the explanatory variables for sales prices (i.e. housing attributes/characteristics) are not constant over time or if their coefficients change. This is a serious problem in places where homes are undergoing extensive renovations, or when some neighborhoods enjoy higher appreciation rates than others (McMillen, 2012). To overcome this problem, a Fisher HPI that considers estimating hedonic regressions in each period of time is estimated following previous methodological contributions by (Lopez & Aroca, 2012; Paredes, 2011; Paredes & Aroca, 2008).

In summary, the construction of housing price indices over comparable samples proposed in this paper involves three steps. First, houses with similar characteristics across time are matched using quasi-experimental methods of control group or matching (Rosenbaum and Rubin, 1983). This step contributes to widening the sample size (relative to the repeated sales methodology) and embracing the heterogeneity of housing market sales samples. Secondly, hedonic regressions are estimated over the treated (t_1) and control (t_0) matched samples to obtain the shadow prices of housing characteristics. Finally, the Fisher index is calculated using the price characteristics estimated in the previous step. This three-step process is repeated for each MSA over time.

The main results suggest that neither SMSAs follow the Chicago MSA, nor Chicago MSA is a particular market separate from the rest; instead, there are two main trend-groups. The first type is formed by the "*price decreasing*" MSAs (Chicago, Rockford and Davenport); and the second type is formed by the "*price steady*" MSAs (Champaign, Springfield, Decatur, Kankakee, Metro-East and Peoria). In addition to these two types of

MSAs, a third one was discovered when the matching approach is used – the "*price increasing*" MSA (Bloomington). The discovery of this price-increasing MSA highlights the importance of controlling for housing characteristics in the construction of the samples to estimate housing price indices.

The remainder of the paper is organized in six sections. The following section describes the methodology and each of the techniques used to calculate HPIs for SMSAs. A description of the data and a discussion of the results are presented in the remaining sections. Due to space limitations, the results section is focused on the analysis of Fisher HPIs in comparison to other methods. However, detailed results from hedonic price regressions are available upon request. Both the results and potential implications of this paper are finally discussed in the concluding section.

Methodology

The main contribution in this paper lies in extending McMillen's (2012) suggestions for using matching as a repeated sales estimator in Chicago to using matching in smaller metropolitan areas, something that has not been considered in the previous US literature. In particular and also acknowledging the contribution in Lonford (2009), matching methods are used in this paper to create comparable data sets over time in terms of housing characteristics. Once matching has been conducted, a data set for each spatial unit r (MSAs) with housing sales for the base period (t_0) and the treated period (t_1) is obtained and used to estimate Hedonic Price Models (HPM). Additionally, the estimated coefficients of each HPM are then used to estimate Fisher HPIs for each period of time. These two methodological features are explained in detail in the following subsections.

Matching

Following Rosenbaum and Rubin (1985), the matching procedure can be broken down into several stages depending on the matching method. The present paper compares matching quality at the annual level as a pre-filter to choose the matching method that best fits the data in terms of the percentage of bias reduction.[3] Following Paredes (2011), it is expected that the percentage of bias reduction will be higher in the cases when matching methods are based on minimizing the difference in housing characteristics (Mahalanobis distance matching) instead of minimizing the distance based on a single representation of the distribution (propensity score/kernel matching). Also, it is expected that a matching method that has a poor level of bias reduction in the annual case (more data makes it easier to find a pair) will not have a better performance at the monthly level (where the number of observations decrease per time period).

As a brief explanation of the matching methods used at the annual level, there are the following options: (1) *One2One* that looks for a clone house in the control sample based on the closest propensity score on the treated; (2) *k-Nearest neighbors* that follows the same procedure as the One2One matching, but within k nearest neighbors in terms of

[3] We follow Rosenbaum and Rubin (1985) definition of percentage of bias reduction as the number of variables that have less than 10% on standardized difference between treated and control samples, respect to the total number of variables considered for matching.

characteristics (when k is usually set to 5); (3) *Kernel* that uses a weighted average of the individuals in the control group to construct the counterfactual outcome.; (4) *Mahalanobis*, based on Mahalanobis distances over the matching attributes, this could be done alone or after a propensity score matching (such as One2One or k-nearest matching) was performed in the first stage.

General propensity-score-based matching follows a two-stage estimation procedure. In the first stage, the propensity score is estimated, using a binary discrete choice model such as:

$$q(\mathbf{x}) \equiv \log[1-e(\mathbf{x})/e(\mathbf{x})] = \alpha + \mathbf{b}'\mathbf{f}(\mathbf{x}) \tag{1}$$

In the second stage, houses are matched on the basis of their predicted probabilities $\hat{q}(\mathbf{x})$ of participation, where $e(\mathbf{x}) = \Pr(\mathbf{x} \mid z = t)$. In this paper, the following algorithm is used to construct the matched samples:

1. For each treated year t_1 (or month), a logit model was estimated using all sales taking place in the base year (month) t_0 and a future and treated year (month) t_1. The dependent variable in the logit model then equals one if the sale took place in the treated year (or month) t_1, and zero if the sale is from the base period t_0. The explanatory variables of the logit regressions are house attributes/characteristics, which are the same as those used for the hedonic price model estimations. The fitted values of these regressions correspond to the propensity score, which provides a continuous metric of a house's probability of belonging to the treated sample.
2. The estimated propensity score from each logit regression was used to match N_1 observations from a treated year (or month) t_1 to sales N_0 from the base period t_0. Note that in this paper, for the yearly case, various matching methods were explored and tested at the annual level to find the two best matching methods with the highest bias reduction. Later, these two matching methods are applied to estimate the housing price indices for the monthly sales data.

As can be seen from figure 4, the annual case method that delivered the highest percentage of bias reduction was a type of Mahalanobis matching. Hence, following Rosenbaum and Rubin (1985), the subsequent methods that expand on the Mahalanobis metric were estimated for higher frequency time frameworks:

- *Nearest available Matching on the estimated propensity score (One2One):* As explained before, this matching follows the previous two-stage estimation algorithm choosing the clone house as the one with the closest propensity score estimation between treated and control samples. This was estimated as a baseline.
- *Mahalanobis matching with propensity scores as covariates (MPSCov):* This method follows the first stage of estimating the propensity score, but it chooses the clone houses based on the Mahalanobis distance $d_M(\mathbf{x}, \hat{q}(\mathbf{x}))$ over the matching attributes including the propensity score as a covariate $\hat{V}_{\mathbf{x},q(\mathbf{x})}$.

$$d_M(\mathbf{x}, \hat{q}(\mathbf{x})) = (\mathbf{x} - \hat{q}(\mathbf{x}))'(\hat{V}_{\mathbf{x},q(\mathbf{x})})^{-1}(\mathbf{x} - \hat{q}(\mathbf{x})) \tag{2}$$

- *Mahalanobis matching with propensity score as calipers (MahalPSCal):* this method provides a more sophisticated pre-filtering matching that could be considered a combination of the previous two methods. It first estimates the *One2One* and drops those observations without a clone. Then, it performs the *Mahalanobis* matching over the remaining data but this time using the estimated propensity scores as calipers. A caliper is defined as a “window” to conduct the search for clones based on Mahalabobis distances. As Rosenbaum and Rubin (1985) suggest, the caliper is defined as: $c = 0.2\sigma$, where $\sigma = [(\sigma_1^2 + \sigma_0^2)/2]^{0.5}$, and the subscripts 1 and 0 denote treatment and control propensity score variances respectively.

Hedonic Price Model (HPM)

The hedonic approach considers the price of a good as the sum of the shadow or implicit prices of its characteristics or attributes (Rosen, 1974). In the housing case, since only the total house sale price is observed, hedonic regressions explain housing prices as a function of house attributes, hence obtaining the shadow or implicit price of each housing characteristic.

There are two features that vary in the literature when estimating HPM. First, studies have varied in the functional form chosen; being the most used the log-log and log-linear specifications. This has depended on the preference for interpreting marginal effects, where estimated coefficients are simply elasticities in the log-log context. However, many other applications also use log-linear regressions since some of the independent variables include the zero value (such as in number of bedrooms where zero means a ‘studio-type’ apartment). This paper uses a log-linear hedonic regression as in Eq. (3).

$$\begin{aligned} \ln y_{it_0} &= \delta_{t_0} + \sum_{k=1}^{K} \beta_{t_0}^{k} X_{it_0}^{k} + \sum_{l=1}^{L} \lambda_{t_0}^{l} Z_{it_0}^{l} + \varepsilon_{it_0} \\ \ln y_{it_1} &= \delta_{t_1} + \sum_{k=1}^{K} \beta_{t_1}^{k} X_{it_1}^{k} + \sum_{l=1}^{L} \lambda_{t_1}^{l} Z_{it_1}^{l} + \varepsilon_{it_1} \end{aligned} \tag{3}$$

Where y_i represents sale price for house *i*, δ is regression intercept, β^k is the estimated hedonic prices for a housing attribute *k={1, 2, ..., K}* and X^k is a vector variable a for housing attribute *k*. Finally, λ^l and Z^l are coefficients and vector variables for each of the *l={1, 2, ..., L}* non-observed attributes, which are assumed to be controlled for due to the benefits of matching.

The second feature of the HPM used in this paper involves estimation of a HPM for each period of time. Subscripts t_0 and t_1 in (3) were used to specify this feature, which is here considered appropriate because, by estimating HPM for each time period, this paper does not assume that the coefficients of housing attributes are constant over time. In this way, the estimated coefficients will vary over time hence capturing changes on the valuation that consumers do of housing attributes in time. This is an important feature that needs to be included since time varying factors such as housing crisis, migration, urban renewal, among others may change consumer preferences. It is important to note that the data used in this paper contains cross-sectional transactions pooled over time, where housing attribute and prices are available at every period.

Fisher Housing Price Indices (Fisher-HPIs)

An additional feature of this paper is the use of Fisher HPIs. Although constructed from a Locally Weighted Regression (LWR), the use of Fisher indices is inspired in the work of (Meese & Wallace, 1991), who estimated HPIs for the each Municipality in the San Francisco/Bay Area. More recently, the use of Fisher indices has also been applied to the case of Spatial HPIs (Paredes, 2011; Paredes & Aroca, 2008) in a single period of time, and for spatial and temporal calculations of this index (Lopez & Aroca, 2012).

As pointed in (Meese & Wallace, 1991), the use of Fisher indices (also called Fisher Ideal Indices) have several advantages over other indices. From (Diewert, 1976) contribution, Fisher Ideal indices have been shown to be both superlative and exact, which are attractive properties when doing index construction since they allow direct comparison between indices and are derived from an underlying utility or production function[4]. Additionally, Fisher indices reduced potential bias since they were calculated as the geometric mean (Eq.4) between Laspeyres and Paasche Indices (Griliches, 1971). Specifically, the bias of using Laspeyres and Passche alone would produce underestimate price indices and the latter one would overestimate the price indices. Equation (4) and figure 1 illustrates this point:

$$P = \frac{p_1 q_1}{p_0 q_1}, \quad L = \frac{p_1 q_0}{p_0 q_0}, \quad F = \sqrt{P * L} \tag{4}$$

«Insert figure 1 here»

Hence, from (4) and using some algebraic manipulation the Fisher index can be derived as:

$$\begin{aligned} \ln F &= 0.5\left[\ln P\right] + 0.5\left[\ln L\right] \\ &= 0.5\left[\ln\left(p_1 q_1\right) - \ln\left(p_0 q_1\right)\right] + 0.5\left[\ln\left(p_1 q_0\right) - \ln\left(p_0 q_0\right)\right] \end{aligned} \tag{5}$$

Where subscripts 1 and 0 denote treatment (t_1) and control/base (t_0) observations respectively. In (5), each component can be interpreted as:

$\ln\left(p_1 q_1\right)$: House sale price based on attributes observed in t_1 valued at t_1 prices.

$\ln\left(p_0 q_1\right)$: House sale price based on attributes observed in t_1 valued at t_0 prices.

---

[4] As pointed in (Meese & Wallace, 1991), (Rosen, 1974) strongly argues that hedonic demand equations can be assumed to represent compensated demand functions if demander/buyers are assumed to be similar.

$\ln\left(p_1 q_0\right)$: House sale price based on attributes observed in t_0 valued at t_1 prices.

$\ln\left(p_0 q_0\right)$: House sale price based on attributes observed in t_0 valued at t_0 prices.

The components $\ln\left(p_1 q_1\right)$ and $\ln\left(p_0 q_0\right)$ correspond to $\ln \bar{y}_{it_1}$ and $\ln \bar{y}_{it_0}$ respectively, which are the mean of the observed house sale prices in each period. However, the components $\ln\left(p_0 q_1\right)$ and $\ln\left(p_1 q_0\right)$ need to be calculated from a previous estimations of the hedonic regressions in (3) as:

$$\begin{aligned} \ln\left(p_0 q_1\right) &= \hat{\delta}_{t_0} + \sum_{k=1}^{K} \hat{\beta}_{t_0}^{k} \bar{X}_{it_1}^{k} \\ \ln\left(p_1 q_0\right) &= \hat{\delta}_{t_1} + \sum_{k=1}^{K} \hat{\beta}_{t_1}^{k} \bar{X}_{it_0}^{k} \end{aligned} \tag{6}$$

Then, replacing (5) in (6) we have:

$$\ln F = 0.5\left[\ln \bar{y}_{it_1} - \hat{\delta}_{t_0} + \sum_{k=1}^{K} \hat{\beta}_{t_0}^{k} \bar{X}_{it_1}^{k}\right] + 0.5\left[\hat{\delta}_{t_1} + \sum_{k=1}^{K} \hat{\beta}_{t_1}^{k} \bar{X}_{it_0}^{k} - \ln \bar{y}_{it_0}\right] \tag{7}$$

For finally applying exponential function in both sides we end up with:

$$F = \left[\exp\left(\ln \bar{y}_{it_1} - \hat{\delta}_{t_0} + \sum_{k=1}^{K} \hat{\beta}_{t_0}^{k} \bar{X}_{it_1}^{k}\right) * \exp\left(\hat{\delta}_{t_1} + \sum_{k=1}^{K} \hat{\beta}_{t_1}^{k} \bar{X}_{it_0}^{k} - \ln \bar{y}_{it_0}\right)\right]^{0.5} \tag{8}$$

which is the expression used in this paper to estimate the Fisher price index for each treated time period $t_1=\{1, 2, \ldots, T\}$ relative to the base/control time period t_0. These estimations were applied for each MSA.

As mentioned before, Fisher HPIs were calculated not only monthly but also by using a three-month moving average algorithm. The construction of both monthly and moving average samples through matching is illustrated in figure 2. These matched samples are later used to estimate hedonic regressions.

«Insert figure 2 here»

Data

The analysis was conducted using monthly housing sales data from January 2005 to June 2012 provided by the Illinois Association of Realtors (IAR). The data contains cross-sectional housing sales' transactions pooled over time, where housing attribute and prices are available at every period. Information about house sales is available for the 10 Metropolitan Statistical Areas (MSAs) in Illinois. The main variables are listing price, closing price, as well as housing characteristics such as square footage, number of bedrooms and number of bathrooms. Although it is acknowledged that the spatial housing location is an important attribute explaining housing price, this paper has not included this variable because it was not available at the time of the estimations[5]. However, the methodology presented in the previous section is capable of including spatial-related variables either in the matching or hedonic regression stages, or both.

Using the most recent classification for the MSAs in Illinois, the housing price data has been assembled for the 10 MSAs representing the most important urbanized areas in Illinois. Figure 3 shows the MSA classification by counties used in this paper. In addition, table 1 presents basic descriptive statistics for the first month (control) and last month of available data. Most variables have the expected number range with exception of some extreme values that were dropped after conducting the matching technique.

«Insert figure 3 here»

«Insert Table 1 here»

Results

Using annual data, matching quality results show that the Mahalanobis Matching with Propensity Scores as Caliper (*MahalPSCal*) is the best method in terms of making the samples comparable. This is verified in figure 4, which shows that the differences in bias reduction[6] on covariates (housing characteristics) are almost complete when using this method. A second measure that supports this result is the propensity score difference between both samples (treated and control), where the *MahalPSCal* presents the lowest differences[7] when compared to other matching methods.

«Insert figure 4 here»

At the annual level, the main finding suggests that results could be overestimated when using the median approach to compare price evolution. Figures 5 and 6 show the evolution of price indices by each MSA, where the Median and the *MahalPSCal* matching approach were used respectively. In the first case, the median approach shows that there are two types of MSA based on their housing price evolution. The first type is formed by the "*price decreasing*" MSAs (Chicago, Rockford and Davenport). The second type is formed by the "*price steady*" MSAs (Champaign, Springfield, Decatur, Kankakee, Metro-East and Peoria).

In addition to these two types of MSAs, a third one was discovered when the matching approach is used – the "*price increasing*" MSA (Bloomington). The discovery of this

[5] Since the Regional Economics Applications Laboratory (REAL) has recently renewed a contract with the Illinois Association of Realtors (IAR) that provides houses' addresses, future versions of this paper will include spatial variables in the analysis.
[6] Percentage of cases with a Standardized Difference lower than 10% as explained in Footnote 3.
[7] These results are not shown here for space reasons. However, they are available upon request.

price-increasing MSA highlights the importance of controlling for housing characteristics in the construction of the samples to estimate housing price indices. Similarly, for MSAs that were classified as "*price steady*" based on the median, a slightly decreasing pattern was discovered when the matching approach was used. Then, the grouping arising from HPIs based on the median approach is somewhat misleading because it suggests little or no movement in prices. Furthermore, finding three different groups (decreasing, steady and increasing) reveals unexpected housing prices' co-movements between MSAs. As mentioned in the introduction, because of the size of Chicago, one could have expected that either other MSAs would follow Chicago's housing trends, or that Chicago would have unique housing trends. However, results suggest that only Rockford and Davenport have similar decreasing behavior as with the Chicago MSA, and the rest of the Illinois MSAs (SMSAs) evidence different housing market behavior in other groups (i.e. steady and increasing).

«Insert figures 5 and 6 here»

How can these results be explained? One possible explanation could be a change in consumer preferences over time. If this were the case, HPIs based on the median would not control for changes in preferences –i.e. consumers that previously preferred/bought bigger houses, might have changed to preferring/buying smaller ones, and hence comparing house prices with different house characteristics distributions will lead to less comparable samples. Figure 7a and 7b support the previous hypothesis, where it can be seen that the Bloomington MSA had an increase in sales for more expensive houses relative to previous years in contrast to observations from the other MSAs. Since the matching method considers a fixed basket of reference (2005), this change in consumer behavior is taken into account in the construction of the price index in contrast to the median price index (figure 5). Similarly, figure 7b shows the change of consumer preferences for the case of the Chicago MSA, and it can be seen that as time passes, there is a shift toward the consumption of less expensive houses and a corresponding decrease in the consumption of more expensive houses. In summary, there was a change in consumer preferences for all MSAs after 2007; however, each MSA changed differently and the median price has no way to account for these changes.

«Insert figures 7a and 7b here»

A good way to make sense of these changes in consumer preferences after the housing crises is by considering changes in the economy that could provide some insights into the changes in housing prices revealed in this paper. Figure 8 shows the Total Non-Farm Employment Growth Rate by MSA between 1995 and 2011. From simple economic reasoning, it could be expected that an increase in employment in an economy would affect (positively) housing prices since the rise in demand is too fast for housing supply to adjust in time (housing supply is usually characterized as inelastic in the short run). Note that Bloomington MSA's employment as an indicator of their economic performance reveals the same trend as the housing price indices here calculated. This fact suggests that the housing indices obtained by the matching approach in this paper might be a better indicator to trace economic performance at the MSA level. The rest of MSAs seem to have a similar trend in employment as for housing prices, especially Rockford that shows the lowest rate of employment growth that is in accordance with the housing price trend resulting from our analysis.

«Insert Figure 8 here»

Turning to the monthly and moving average level estimations obtained by the same *MahalPSCal* matching approach as in the annual case, figure 9 shows that the moving average results also produce three groups of MSAs in terms of different trends in price evolution. This confirms the previous results shown in the annual case. The first group is formed by the downward-trend MSAs, led by Chicago, which shows a clear structural change in the early months of 2008. The trend of the second group is more stable in time, showing a less sensitive market to the housing crisis in 2008 when compared to the previous group. Finally, Bloomington MSA, representing the third group, shows an upward trend that was not previously discovered when the analysis was based on the median price. In contrast to the median estimations, the revealed upward pattern is now in accordance with other economic indicators such as employment (see figure 8). Although the jump of Bloomington employment growth starts earlier than the Bloomington Housing Price Index jump, this might be associated with housing being a non-tradable[8] good that reacts more slowly to economic shocks. Although it is acknowledged that other variables should be controlled for before stating a relationship between housing prices and employment, the revealed trend for Bloomington should at least be considered as a sign to stimulate more research about this relationship and the importance of estimating accurate housing price indices. Housing price indices only based on median prices should not be used since they show a biased representation of the housing market and fail to consider how other market forces might influence it. Figure 10 presents different price indices for Bloomington showing clear evidence of the bias incurred when using only the median versus the matching-based Fisher price indices.

«Insert figures 9 and 10 here»

Figure 11 shows the comparisons among different housing price indices for Chicago MSA, using the moving-average Fisher HPIs with the *MahalPSCal* matching method, the CS HPIs based on the repeat sales approach, and the traditional housing indices using median prices. Compared with the Fisher HPIs, the indices estimated by the traditional median approach and the repeat sales approaches were significantly over-estimating the price indices for Chicago MSA. As mentioned in the methodology section, this over-estimation is expected since Median and CS indices are based on data sets that are affected by outliers and are only a sample of the total houses sold in a period of time.

«Insert figure 11 here»

As in the annual case, these monthly-level differences could be a result of changes in consumer preferences, and they can be better understood when decomposing the changes

[8] Fixed in space and therefore not subject to arbitrage.

in prices by characterizing them in terms of price stratifications. Figure 12 provides a detailed evaluation of the last 12 months of available data showing the differences between the indices (Fisher based on *MahalPSCal* Matching vs. Median, right axis) along with the price stratification (left axis) for the case of Chicago MSA. Although both indices seem to have a similar behavior, there are three important differences between them: First, the median price overestimates housing prices by 0.2 points on average. This means that while the Fisher HPI reveals that housing prices in the last time period were around 55% of the levels in January 2005, the median index suggests that they were around 73%. Secondly, the Fisher HPI shows a more stable behavior than the median price index. This suggests that the Fisher HPI is not affected by short-term economic shocks but that it captures the underlying trend. Thirdly, the divergence of the median price index from the Fisher HPI starting on February 2012 is mainly due to both the increase of the sales of more expensive houses (300K~500K level up), and the decline of the less expensive houses in the 0~100K level. Although the Fisher HPI also shows an upward trend in these last months, it is much more conservative than the median price that could be affected by very expensive sales that only account for a very small fraction of the total sales. In fact, in June 2012, the more expensive sales categorized in the 300K~500K level to the 700K~UP level, account for only 30% of the total sales; the rest of the price strata (from 0~100K to 200K~300K) account for the remaining 70%.

«Insert figure 12 here»

Conclusion

Policy-makers commonly use HPIs as an important indicator to measure and track the behavior of housing markets. As mentioned before, this paper contributes by bringing together different methodologies for estimating HPIs that were not available for SMSAs until now. Since it has been argued that the nature of housing sales requires having HPIs that control for housing attributes and its potential variation over time, the HPIs estimated in this paper will potentially provide policy-makers of improved better measures to tract the behavior of housing markets and take decisions accordingly including all MSAs.

There are several additional benefits of having accurate information not only for big MSAs but also for all MSAs in a state. First, understanding the co-movements between small and big MSAs in housing prices is important since it provides information about the potential influence of one on the other, providing a better interpretation of housing dynamics over time. Additionally, since the cost of housing is a major component included in the cost of living, housing price indices are essential as an approximate measurement for the cost of living. Therefore, providing housing price indices for lower spatial levels (e.g., smaller MSAs) could be useful indicators to better understand regional disparities in costs of living both within a state and between states.

Despite the aforementioned contributions, several improvements and remain pending. First, spatial variables and models should be introduced in future versions of this index. This extension will enhance the accuracy of the estimations since it will capture the influence of spatial spillover effects of nearby houses (neighborhood effects) and amenities on house prices. Second, the presented Fisher HPIs can also be estimated at more disaggregated spatial levels such as county and even at neighborhood levels. Assuming that housing markets differ depending on the spatial location, this will provide a finer picture of how housing markets behave over time. Third, the price indices provided here could be improved both by testing the dependency from the base period chosen and by estimating confidence intervals. These extensions will contribute to further support the use of this technique.

References

Angelides, P., and Thomas, B. 2011. The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States (Revised Corrected Copy). *United States Government Printing Office*.

Diewert, W. E. 1976. Exact and superlative index numbers. *Journal of Econometrics*, 4(2): 115–145. doi:10.1016/0304-4076(76)90009-9

Dubé, J., and Legros, D. 2011. A spatio-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science*, 92(1): 19–31. doi:10.1111/j.1435-5957.2011.00402.x

Lopez, E., and Aroca, P. 2012. Estimación de la inflación regional de los precios de la vivienda en Chile. *Trimestre Economico*, 79(315): 1–31.

McMillen, D. P. 2012. Repeat sales as a matching estimator. *Real Estate Economics*, 40(4): 745–773. doi:10.1111/j.1540-6229.2012.00343.x

Meese, R., and Wallace, N. E. 1991. Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Real Estate Economics.* 19(3): 308-331.

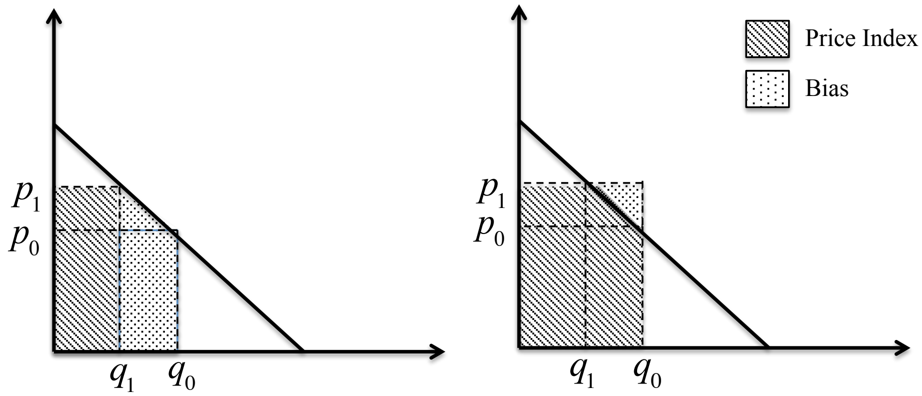
Paredes, D. (2011). A methodology to compute regional housing price index using matching estimator methods. *The Annals of Regional Science*, 46(1): 139–157. doi:10.1007/s00168-009-0346-z

Paredes, D., and Aroca, P. 2008. Metodología para estimar un índice regional de costo de vivienda en Chile. *Cuadernos De Economía*, 45: 129–143.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy*, 82(1): 34–55.

List of Figures

Figure 1. Laspeyres and Paasche Price Indices Bias



Paasche Price Index Underestimation *Laspeyres Price Index Underestimation*

Figure 2. Monthly and Moving Average matching sample construction

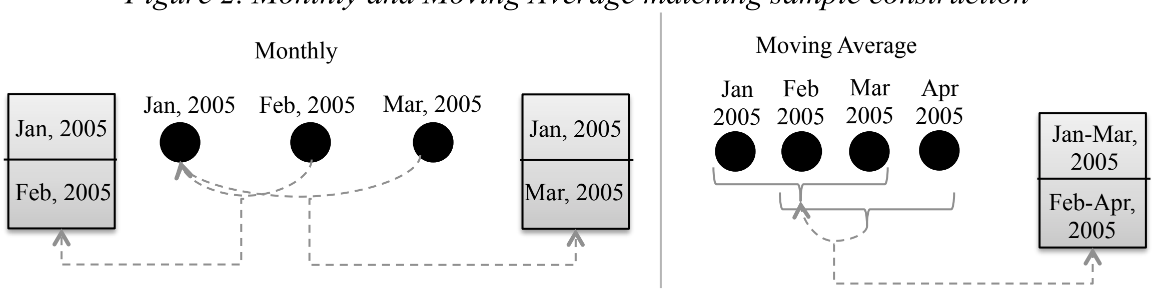
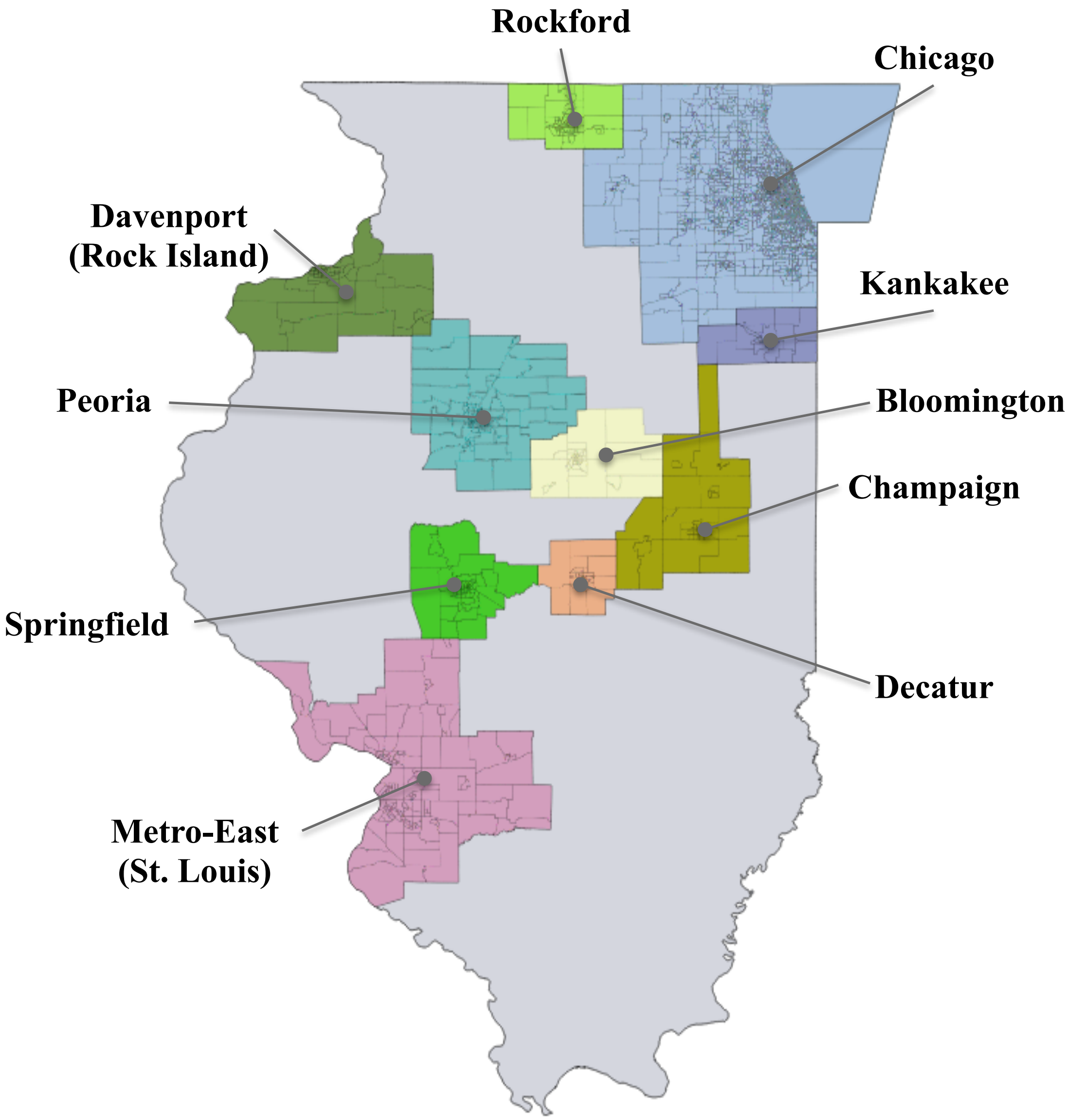


Figure 3. MSA classification by county in Illinois



Source: Statistical Policy Office. United States Office of Management and Budget.

Figure 4. Bias Reduction Evolution by Matching Method
Based on the standardized differences of housing characteristics' variables

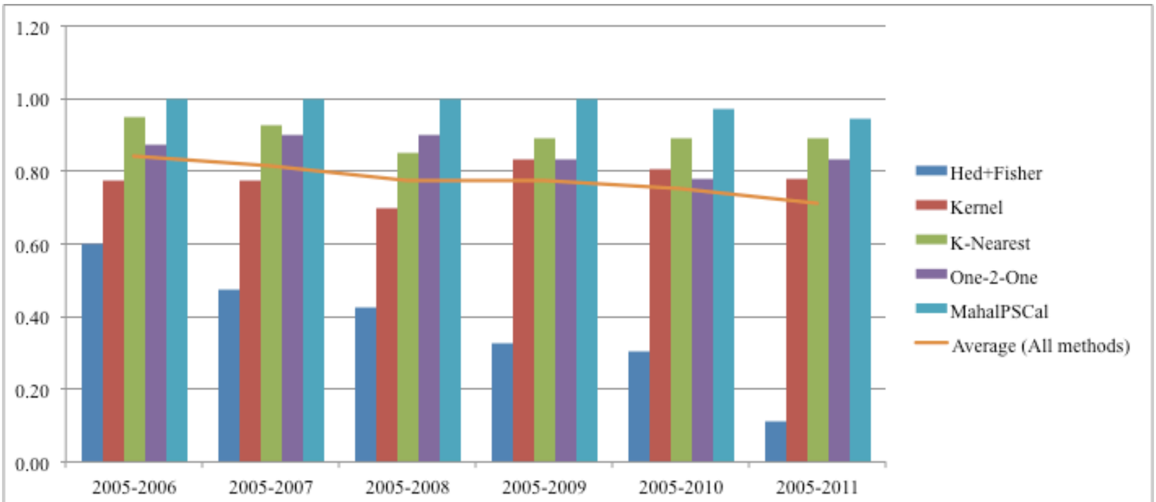


Figure 5. Housing Price Annual Variation
Based on the Median

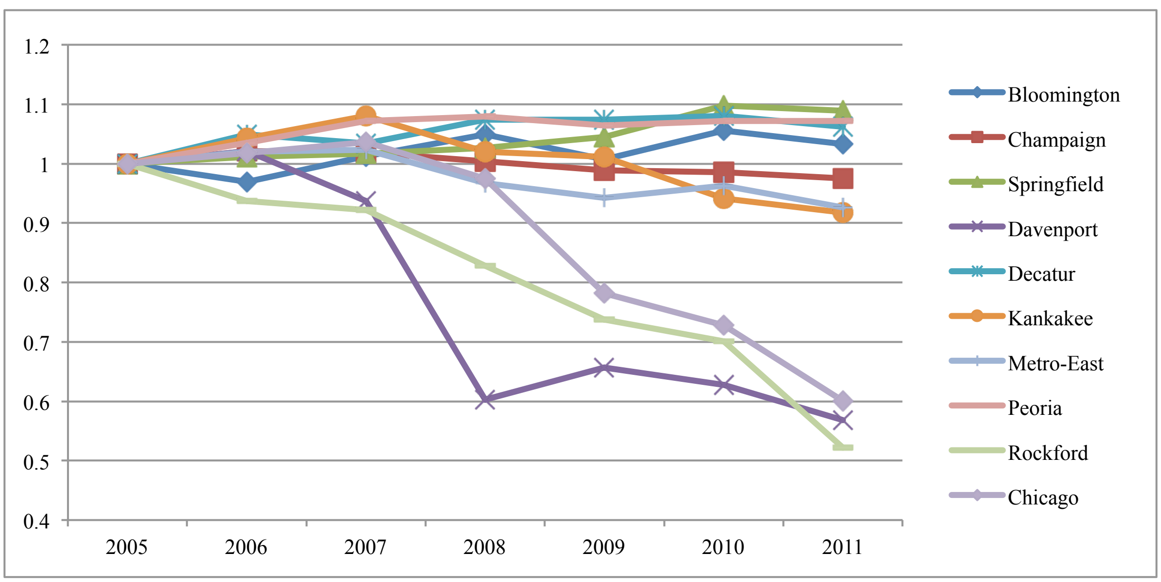


Figure 6. Housing Price Annual Variation
Based on MahalPSCal Fisher Price Index

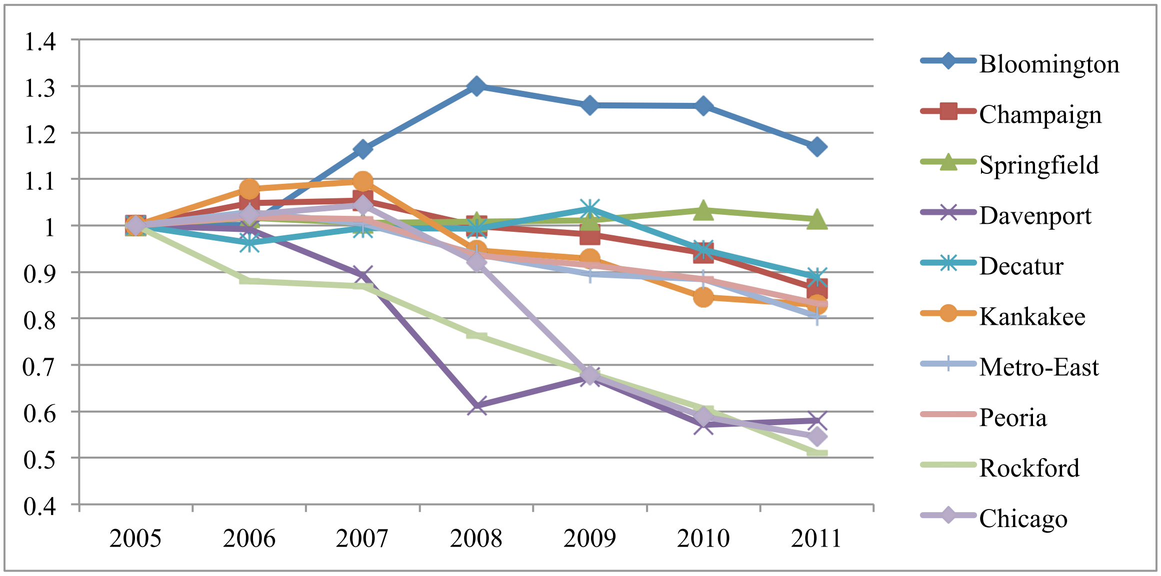


Figure 7a. Price Stratification of Housing Sales, Bloomington MSA

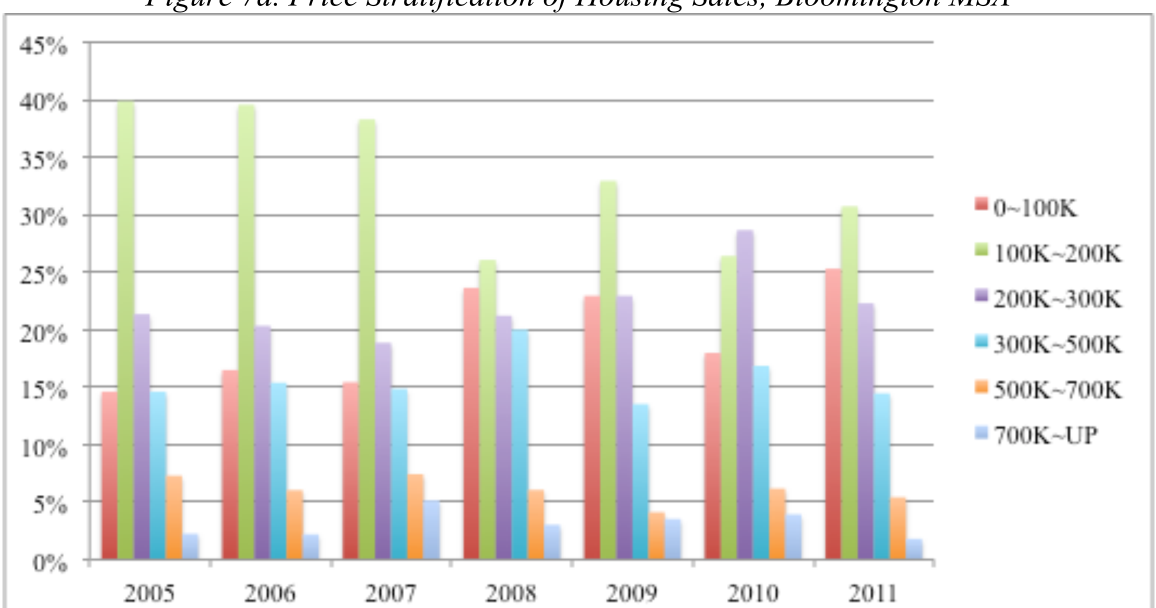


Figure 7b. Price Stratification of Housing Sales, Chicago MSA

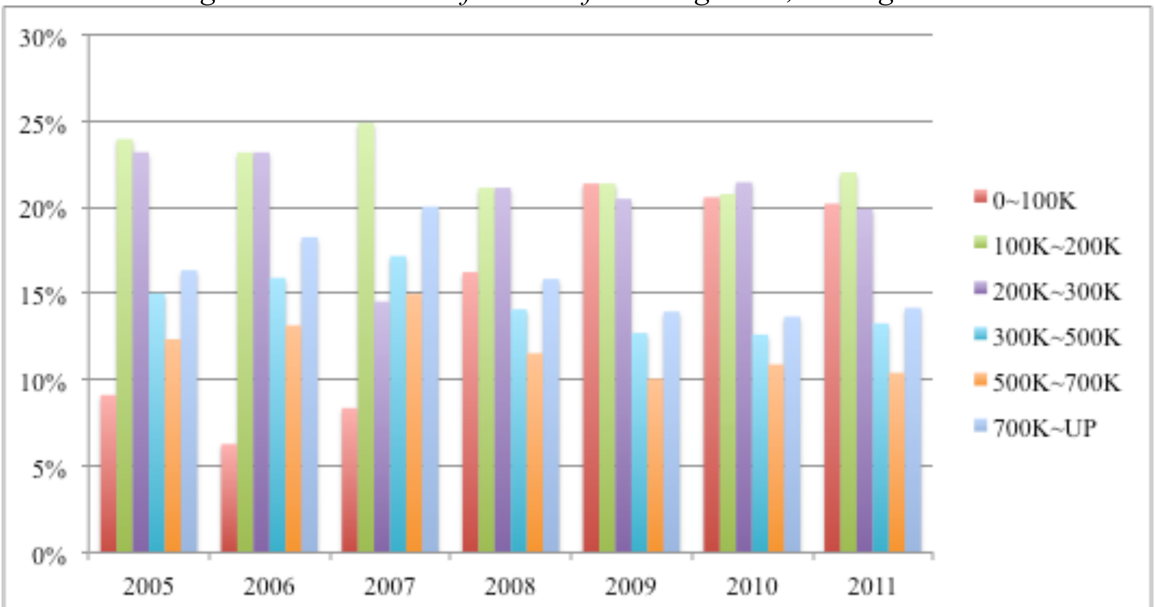
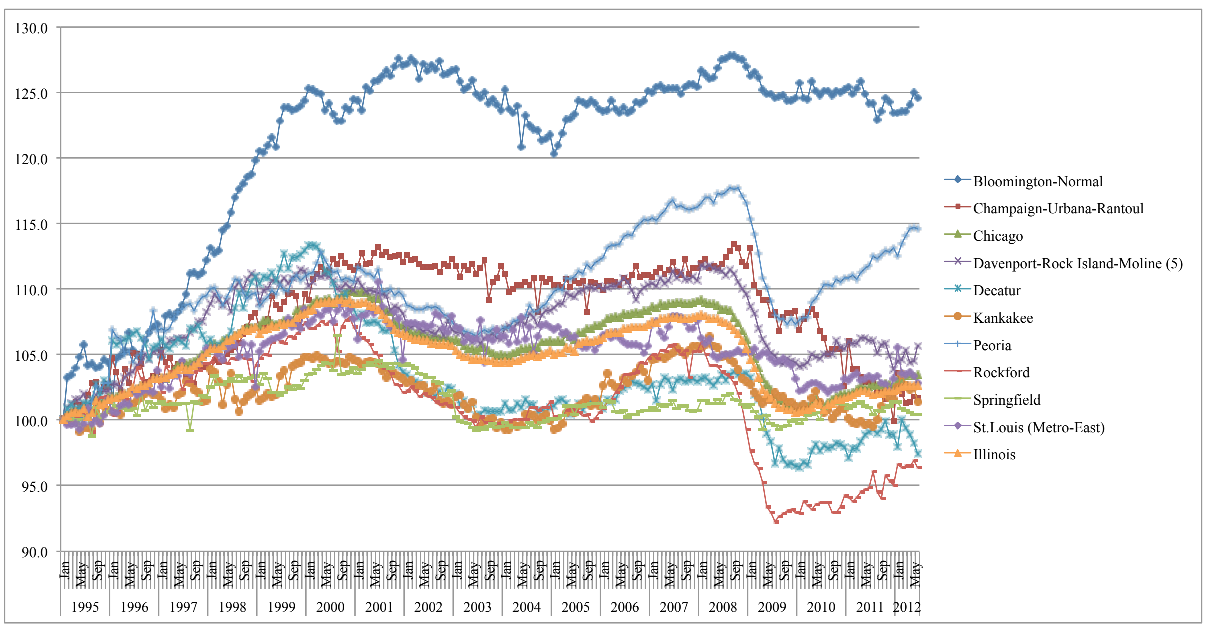
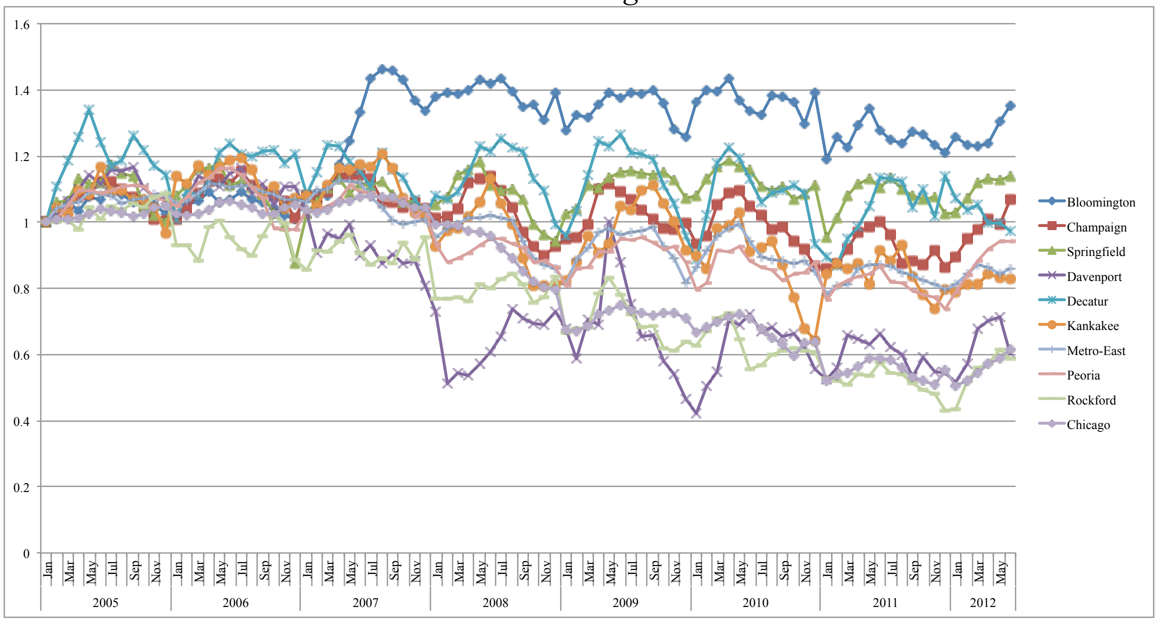


Figure 8. Total non-farm Employment growth rate Jan 1995 – Jun 2012



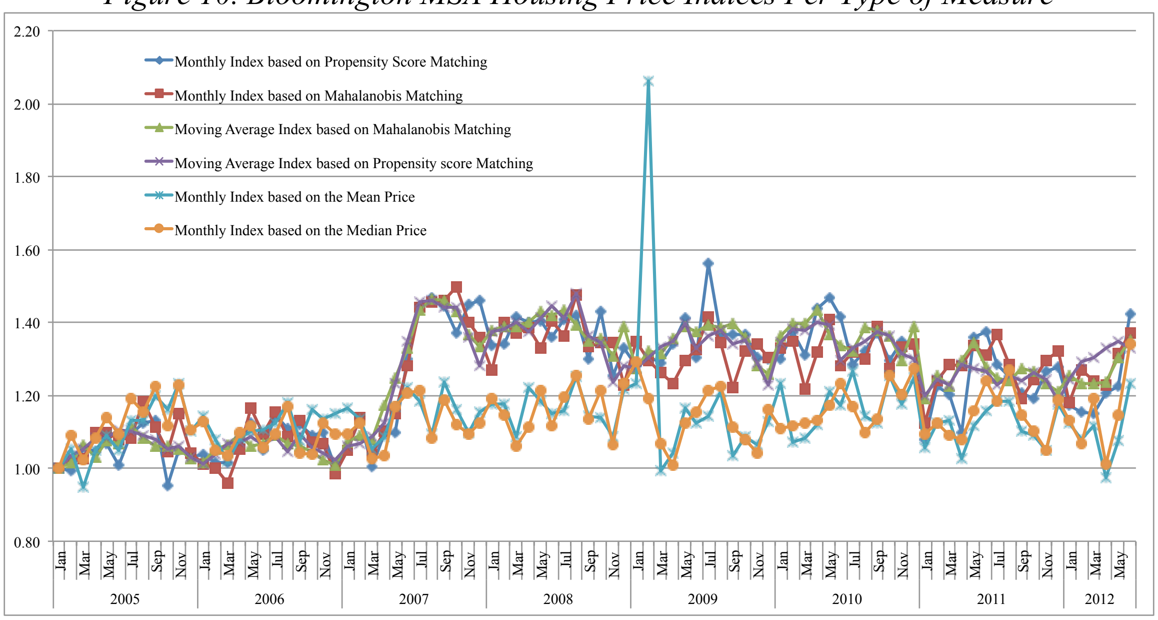
Source: Illinois Job Index, MSA report, released 10/03/2011

Figure 9. Housing Price Indices Per MSA based on Mahalanobis Moving Average Matching



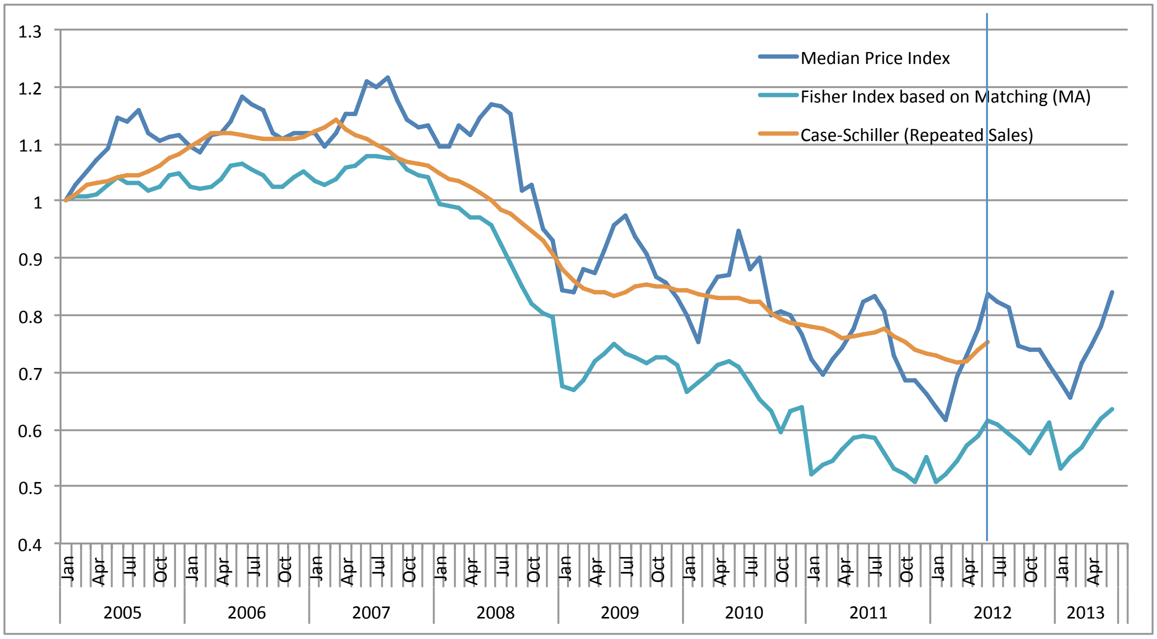
Note: This graph can be found in its interactive version at: http://goo.gl/wrTj5

Figure 10. Bloomington MSA Housing Price Indices Per Type of Measure



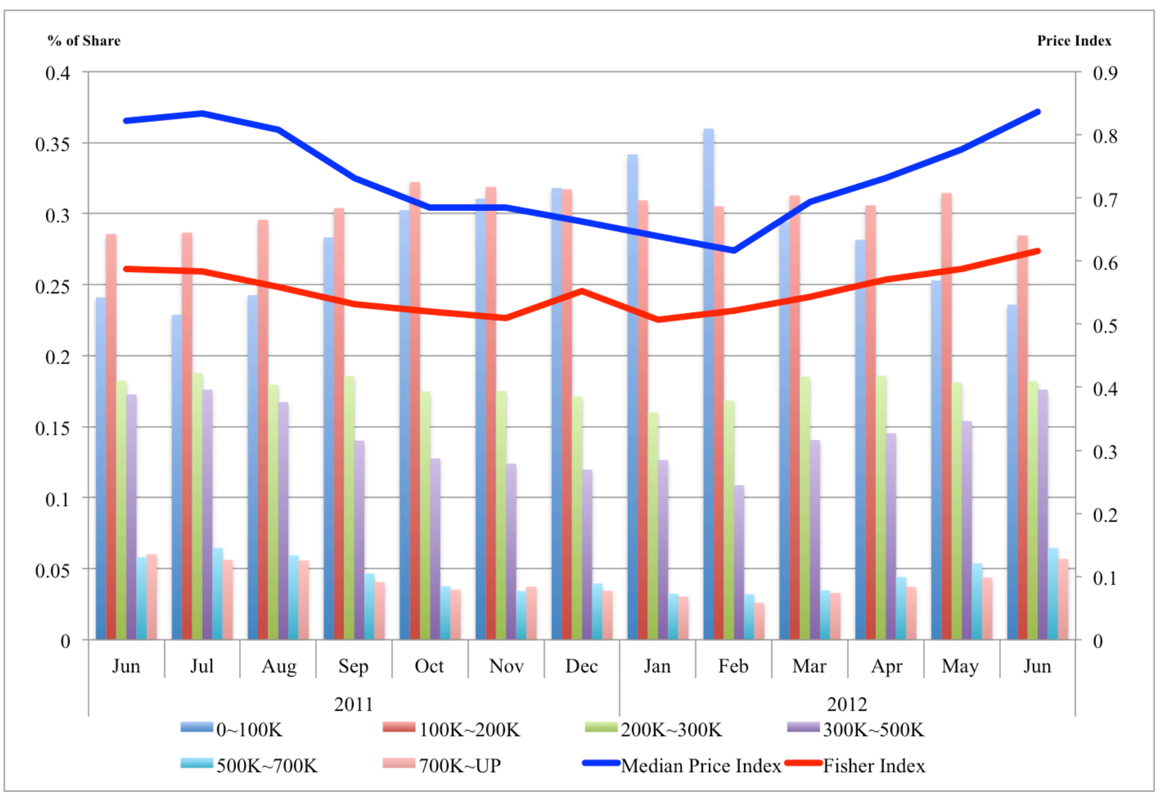
Note: Comparisons of other MSAs bias can be explored in the interactive version of this graph at: http://goo.gl/wrTj5

Figure 11. Comparisons among Fisher Indices (Mahalanobis Matching Method), Case-Schiller Indices (Repeat Sales) and Indices of Median Price for Chicago MSA



Note: The horizontal line sets the beginning of to the forecast for each series from July 2012 to June 2013.

Figure 12. Price Stratification and Price Index Comparison Chicago MSA



List of Tables

Table 1. Basic Summary Statistics of the used Data - January 2005

MSA	Name	Bedroom			Bathroom			Closing Price			Square Footage			Sample size
		Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	
MSA 1	Bloomington-Normal	1	3	6	1	2	5	18079	134500	505000	541	1680	7319	99
MSA 2	Champaign	2	3	5	1	2	3	15000	125750	372500	535	1600	4452	108
MSA 3	Chicago	0	3	7	1	2	6	27000	262661	5002368	460	1533.5	15022	2344
MSA 4	Springfield	1	3	5	1	2	3	8500	106500	458000	650	1668.5	5380	152
MSA 5	Davenport-Moline-Rock Island	1	3	4	1	1	3	4000	82000	505000	480	1302	7650	117
MSA 6	Decatur	1	3	6	1	1.5	5	4000	57750	320000	672	1170	5540	102
MSA 7	Kankakee	1	3	6	1	2	5	9000	115572	438000	560	1334	4118	56
MSA 8	Metro-East	1	3	6	1	2	5	0	102900	620000	580	1523	5670	373
MSA 9	Peoria	1	3	5	1	1	3	8500	91000	465000	600	1420.5	5501	258
MSA 10	Rockford	2	3	5	1	2	3	65000	140500	265434	1000	1476	3800	38

June 2012

MSA	Name	Bedroom			Bathroom			Closing Price			Square Footage			Sample size
		Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max	
MSA 1	Bloomington-Normal	2	4	6	1	2	6	20000	180500	805000	720	1832	6246	252
MSA 2	Champaign	1	3	6	1	2	5	8500	154000	875000	600	1700	5788	311
MSA 3	Chicago	0	3	7	1	2	9	1	180000	6000000	440	1600	43117	7124
MSA 4	Springfield	1	3	6	1	2	4	12500	136000	795000	616	1800	20725	278
MSA 5	Davenport-Moline-Rock Island	2	3	4	1	1	2	12900	57000	322000	978	1624	2781	17
MSA 6	Decatur	1	3	5	1	2	5	2200	86000	510000	624	1636	4522	97
MSA 7	Kankakee	1	3	5	1	2	5	13900	129110	325500	720	1592	4000	146
MSA 8	Metro-East	1	3	5	1	2	6	3000	119250	545000	560	1709.5	6244	558
MSA 9	Peoria	1	3	5	1	2	8	5500	130000	1100000	503	1561	7650	441
MSA 10	Rockford	2	3	6	1	2	4	7550	85000	500000	779	1446	9148	100