Institute for Housing Studies Cook County House Price Index Technical Paper on Sample and Methodology

The Institute for Housing Studies[Ψ]
DePaul University

March 2013

[Ψ] 1 East Jackson Boulevard Ste 5400, Chicago, IL 60604
Email: housingstudies@depaul.edu, Tel: 312-362-7074.

1 Introduction

The Institute for Housing Studies (IHS) is a research center based at DePaul University. IHS conducts research and provides data and technical assistance to housing practitioners and stakeholders around issues related to the state of the housing market in Cook County. A keystone initiative of the Institute has been the development of the IHS Data Clearinghouse; a database of property-level housing data that allows IHS to conduct in-depth analysis of housing market trends and conditions in Chicago and broader suburban Cook County.

In 2012, data from IHS's Data Clearinghouse was used to develop a repeat sales index to track price changes of residential properties in Cook County. The IHS Cook County House Price Index is released on a quarterly basis and tracks price changes separately for the key property types found in the County: single family homes, condominium units, smaller multifamily rental buildings with between two and four units, and larger multifamily buildings with more than five units. Unlike many national indices which are broader in scope and focus on tracking price change in the single family market, the IHS Cook County House Price Index is more granular in focus and allows users the opportunity to compare price trends for different types of properties.

2 About the Data and the Sample

The IHS Cook County House Price Index uses real estate transactions data from the Cook County Recorder of Deeds (via Property Insight) as a primary data source as well as data from the Midwest Real Estate Data (MRED) to identify the matched sales pairs used to calculate the Index. Data from the Recorder of Deeds includes information on the date that the transaction occurred, the date that the transaction was recorded by the County, the sales price, and information on the buyer and seller involved in the transaction. When possible, data from the MRED is used as a secondary source to confirm the transactions observed in the Cook County Recorder data and to supplement the primary data source when fields are missing.

IHS also uses data from the Cook County Assessors Office to identify the geographic location of a property and the property type. This data set has information on every parcel in Cook County and is also used to identify the universe of parcels on which the Index weighting process is based. Data from the Cook County Assessor is updated yearly and has information on the characteristics of a parcel, including property type, as well as the assessed value of a parcel in a given year. This data set is geocoded based on the parcel address and each parcel is assigned a location based on the PUMA (Public Use Microdata Area) location of each parcel. [1]

The IHS Cook County House Price Index is a weighted repeat sales index similar to those developed and used by major national indices including Case-Shiller, Corelogic and the Federal Housing Finance Agency. IHS uses sales transactions that occurred on the same property within the Index period (a matched pair) and assumes that the quality and size of the property is unchanged between transactions. IHS cleans the data for obvious outliers that could skew the index including non-arm's-length transactions, conversions from one property

[1] Appendix I contains a map of all 33 PUMAs within Cook County.

type to another property type, obvious renovations, likely fraud, and data entry errors of sales prices. To reduce the number of non-arm's-length transactions, transactions without a recorded value or a subsequent recording of taxes paid were not included in the data set used to calculate the Index. A non-arm's -length transaction is a transaction where the price observed is likely not reflective of market price due to a relationship between the parties that influences those individuals from acting in their self-interest. Annual price changes occurring in the upper and lower five percent of the data set were removed to correct for both potential data entry errors and for unpredicted depreciation, total renovation, or fraud. IHS does not create a matched pair of transactions on a property where the successive transaction occurs within a six month window. Prices associated with these types of transactions might not reflect market price, such as in a flipping scenario. IHS also does not include transactions on properties that were identified as new construction as these transactions lack a previous sale on which to base price change.

IHS waits six weeks after the end of a quarter to update its Index data with all transactions that fit the qualifying criteria. This window accounts for delays in the recording of new transactions and missing transactions are added to the subsequent quarter's data set. To account for the mismatch between the date that the transaction was executed and the date that the transaction was recorded by the County, IHS employs a rolling window of three months. For example, a given quarter's data will include matched pairs of transactions that occurred in the previous quarter. This process ensures that all transactions are accounted for and that there are sufficient transactions within a quarter for analysis. Table 1 shows the resultant sample sizes for repeat sales by property type in Cook County from 1997 to 2012.

3 Methodology and Models

Using data from the Cook County Assessor, the IHS Cook County Index uses data on location, property type and assessed value to assign a weight to each property and to correct for sampling errors inherent to a repeat sales index.[2] Within a PUMA and separately for each property type, IHS has created a tiered "assessor value class" based on the average of the 2009 and 2010 Assessed value of a property. These value tiers weight each property in the data set. The first tier is defined as properties with values in the bottom 30 percent of observed values for that property type and geographic location, the second tier as properties with values in the 30 to 70 percent range, and the third tier as properties with values in the top 30 percent.

The IHS Cook County House Price Index uses a Weighted Least Square (WLS) approach to correct for the significant variation, or heteroskedasticity, observed in prices between properties that occur across time periods and based on property location. Indices that use the repeat sales approach, such as Case-Shiller (Case and Shiller, (1987,1989)) also use a WLS approach to correct for the heteroskedastic nature of the housing market. However, while

[2]Some selection bias is indigenous to the repeat sales approach and some variation cannot be corrected. For example, some selection bias occurs because the Index measures only price changes on properties that were sold within the Index period, not all properties. Additionally, bias can occur because certain types of properties, particularly lower-value properties, tend to trade more frequently than other properties, particularly higher-price properties. Numerous factors contribute to the value of a particular property that results in variation in price between two otherwise similar properties from a data perspective. For more on the sample selection bias issue as it relates to repeat sales price indices, please see Hwang and Quigley (2004) and Steele and Goy (1997).

Table 1 Paired Transaction in Cook County by Property Type

	SINGLE FAMILY	CONDO	2-4 UNIT Multifamily	5+ UNIT Multifamily	TOTAL
1997	5,179	3,844	829	166	10,018
1998	6,927	5,156	1,001	218	13,302
1999	8,372	6,220	1,403	290	16,285
2000	9,259	7,656	1,489	174	18,578
2001	10,734	8,946	1,819	212	21,711
2002	12,650	10,133	2,174	307	25,264
2003	15,078	11,851	2,485	259	29,673
2004	17,260	13,366	3,421	340	34,387
2005	18,509	14,643	4,121	346	37,619
2006	16,138	12,871	3,490	279	32,778
2007	11,463	10,304	1,803	223	23,793
2008	6,829	6,331	888	139	14,187
2009	5,925	5,101	779	82	11,887
2010	6,345	4,549	996	82	11,972
2011	6,176	4,417	1,070	108	11,771
2012*	2,447	1,853	400	48	4,748
Total	159,291	127,241	28,168	3,273	317,973

*: Transactions that were executed as of June 31, 2012 were included.

the Case-Shiller Index uses a least squares approach to weighting based on the errors of the regression model to correct the heteroskedasticity, IHS creates a weight using a full population approach given the Institute's unique access to data on all parcels and properties in Cook County. Information on the entire universe of parcels, not just those with transactions, allows IHS to assign a precise weight to each transaction based on the universe of properties and based on geography and price tiers at the time of each transaction.

The Repeat Sale Index:

The underlying hedonic type of model shows:

$$ln(p_i) = \beta_0 + \beta_1 \mathbf{D}^T + \beta_2 \mathbf{H}_t + e_t, \tag{1}$$

$$\begin{aligned} where\ i &= 1, ..., n, \\ \beta_1 \mathbf{D}_i^T &= \beta_{11} D_{1i} + \beta_{12} D_{2i} + ... + \beta_{1t} D_{ti}, \\ \beta_2 \mathbf{H}_i &= \beta_{21} H_{1i} + \beta_{22} H_{2i} + ... + \beta_{2k} H_{ki}. \end{aligned}$$

The equation uses n number of transactions t periods, and k number of hedonic variables.

Now we have properties sold twice over the sample period time. A house sold in period t and $t+\tau$, a model of logarithm of its price change is found by subtracting equation (1) for period t from the period $t+\tau$, it follows that

$$ln(p_{i,t+\tau}) - ln(p_{i,t}) = (\beta_0 - \beta_0) + \beta_1 \mathbf{D}^{T+\tau} - \beta_1 \mathbf{D}^{T} + \beta_2 \mathbf{H}_i + \epsilon_i^{T+\tau} - \epsilon_i^{T}, \tag{2}$$

$$\begin{aligned} where\ T &= 0, ..., t, \\ T+\tau &= 0, ..., t, t+1, ..., t+\tau. \end{aligned}$$

From model equation (2), a traditional repeat sale can be defined:

$$ln\left(\frac{p_{i,t+\tau}}{p_{i,t}}\right) = \beta_1 \mathbf{D}^{\tau *} + \epsilon_i^{\tau *}, \tag{3}$$

$$where\ \tau * = t, t+1, ..., t+\tau.$$

where $\mathbf{D}_i^{\tau *}$ is the series of dummy variables with the value 1 within the periods that the resale occurred, -1 in the period that the previous sale occurs, and 0 otherwise. The repeat sale price index going from 1 to period t $(t = 1, ..., t)$ is obtained from the exponential of the sum of corresponding estimated regression coefficients divided by the base year exponential of the sum of corresponding numbers until the base year $(t = 1, ..., b)$.

$$RPI_t = \frac{exp\left(\sum_{t=1}^{t} \hat{\beta}_t\right)}{exp\left(\sum_{t=1}^{b} \hat{\beta}_b\right)}, \tag{4}$$

where β_b = base year coefficient of regression.

The Sample Weighting Method:

The initial geographical and three tier based weights can be created using the universe of parcel data.[3] The weight has been generated the following way:

$$w_i jk = \frac{n_{ijk}}{N_k}, \tag{5}$$

[3]The appendix includes a detailed table of the weighting by geography, property type and price tier.

$$
\begin{aligned}
where\ N_k &= \textit{Total Number of Properties of k property type},\\
i &= 1, ..., 33\ \textit{(PUMAs)},\\
j &= \textit{tier1, tier2, tier3},\\
k &= \textit{property types (single family, condo,}\\
&\quad \textit{2-4 unit multifamily, 5+ unit multifamily).}
\end{aligned}
$$

The final weights can be obtained by the number of transactions that occurred by geography and price tier at t. Equation (6) shows the weights based on initial weight (5) for each transaction at the time occurred. Using the weights (6), an equation is estimated (3) with the weight as weighted least square.

$$
h_{ijkt} = \frac{\frac{w_{ijk}}{m_{ijkt}}}{\sum_j^3 \sum_{i=1}^{33} w_{ijk}}, \tag{6}
$$

$$
\begin{aligned}
where\ w_{ijk} &= \textit{Weight of i PUMA j tier property type},\\
m_{ijkt} &= \textit{Number of transactions of i PUMA j tier of k property at time t},\\
i &= 1, ..., 33\ \textit{(PUMAs)},\\
j &= \textit{tier1, tier2, tier3},\\
k &= \textit{property types (single family, condo,}\\
&\quad \textit{2-4 unit multifamily, 5+ unit multifamily).}
\end{aligned}
$$

The Hodrick-Prescott (HP) Filtering for Submarket Area (PUMA):

The IHS house price index uses the Hodrick-Prescott [4] filtering approach in order to identify the long-term trends of price changes for submarket areas. At the submarket level, cyclical components such as distressed sales can heavily influence the price in the short-term, which might distort the long-term price trend. Assuming an aggregated submarket, a house price(p_t) can be expressed as a growth component(τ_t) and a cyclical component (c_t),

$$
p_t = \tau_t + c_t.
$$

The HP filter choose the growth component by solving

$$
Min_\tau \left(\sum_{t=1}^{T} (p_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right).
$$

[4] Hodrick and Prescott (1980) developed a filter to understand the growth component in a aggregated time series in business cycle by isolating the cyclical component in raw data. Since then, the filter has been widely used to remove the temporal or cyclical components in a aggregate time series while maintaining the growth trend.

4 References

Case, K., and Shiller, R. (1987). Prices of single-family homes since 1970: new indexes for four cities. New England Economic Review, 87(1987), pp. 4556.

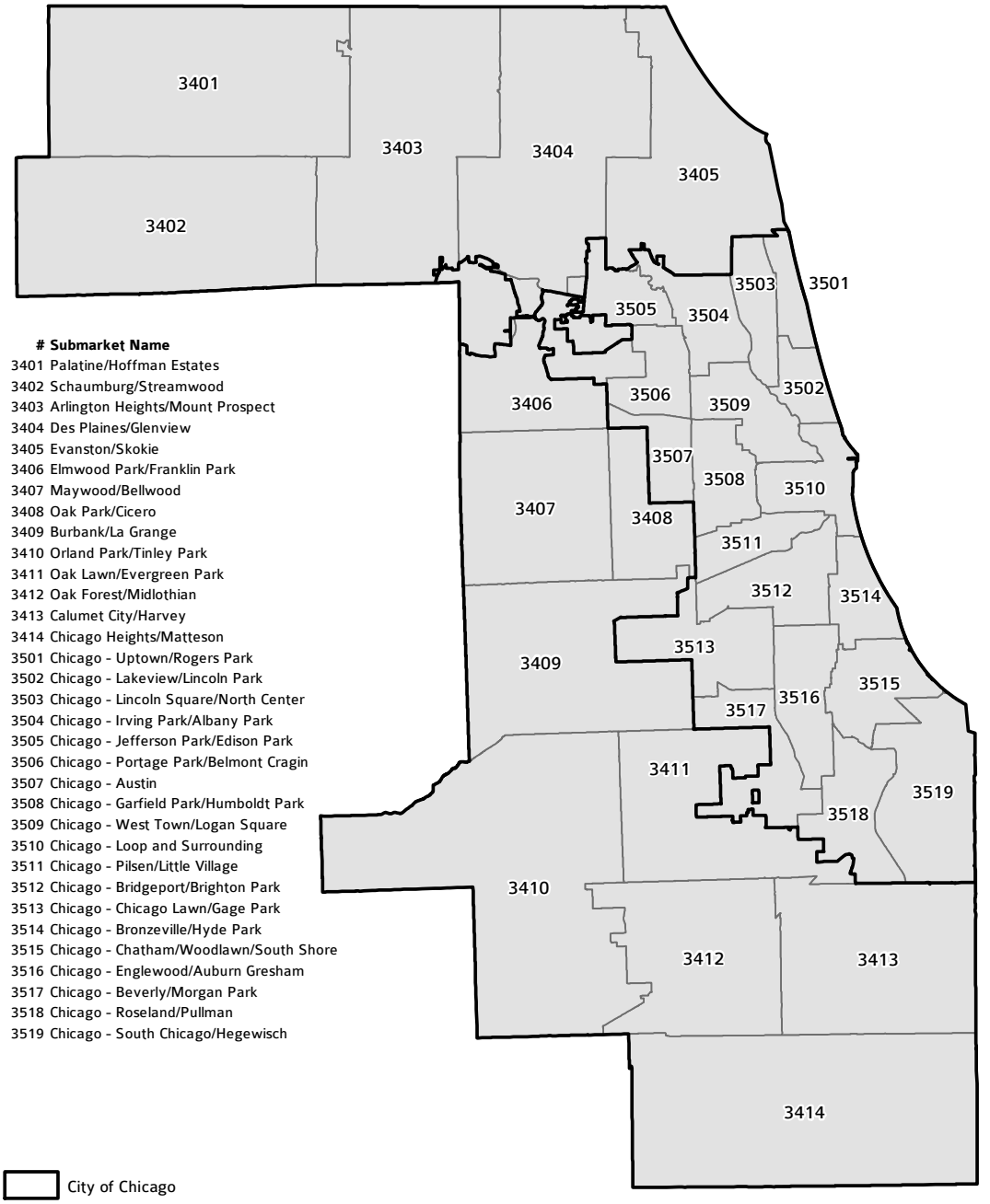
Case, K. E. and Shiller, R. J. (1989). The e ciency of the market for single-family homes, American Economic Association 79, pp. 125-137.

Hwang, M. and Quigley, J. M. (2004), Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values, Journal of Real Estate Finance and Economics, 28(2-3), pp. 161-78.

Hodrick, Robert J., and Edward C. Prescott (1980), Postwar U.S. Business Cycles: An Empirical Investigation, Carnegie Mellon University discussion paper no. 451.

Steele, Marion and Goy, Richard (1997). Short Holds, the Distributions of First and Second Sales, and Bias in the Repeat-Sales Price Index, The Journal of Real Estate Finance and Economics, 14(1-2), pp. 133-54.

APPENDIX I: PUMA (PUBLIC USE MICRODATA AREA)



#	Submarket Name
3401	Palatine/Hoffman Estates
3402	Schaumburg/Streamwood
3403	Arlington Heights/Mount Prospect
3404	Des Plaines/Glenview
3405	Evanston/Skokie
3406	Elmwood Park/Franklin Park
3407	Maywood/Bellwood
3408	Oak Park/Cicero
3409	Burbank/La Grange
3410	Orland Park/Tinley Park
3411	Oak Lawn/Evergreen Park
3412	Oak Forest/Midlothian
3413	Calumet City/Harvey
3414	Chicago Heights/Matteson
3501	Chicago - Uptown/Rogers Park
3502	Chicago - Lakeview/Lincoln Park
3503	Chicago - Lincoln Square/North Center
3504	Chicago - Irving Park/Albany Park
3505	Chicago - Jefferson Park/Edison Park
3506	Chicago - Portage Park/Belmont Cragin
3507	Chicago - Austin
3508	Chicago - Garfield Park/Humboldt Park
3509	Chicago - West Town/Logan Square
3510	Chicago - Loop and Surrounding
3511	Chicago - Pilsen/Little Village
3512	Chicago - Bridgeport/Brighton Park
3513	Chicago - Chicago Lawn/Gage Park
3514	Chicago - Bronzeville/Hyde Park
3515	Chicago - Chatham/Woodlawn/South Shore
3516	Chicago - Englewood/Auburn Gresham
3517	Chicago - Beverly/Morgan Park
3518	Chicago - Roseland/Pullman
3519	Chicago - South Chicago/Hegewisch

City of Chicago

Source: United States Census Bureau
Prepared by Institute for Housing Studies at DePaul University

INSTITUTE FOR HOUSING STUDIES

APPENDIX II: IHS Housing Unit Weights by Property Type and Three Price Tiers in PUMA

PUMA*	SINGLE FAMILY			CONDO		
	Tier1) (0- 30%)	Tier2 (30-70%)	Tier3 (70-100%)	Tier1 (0- 30%)	Tier2 (30-70%)	Tier3 (70-100%)
3401	0.057	1.281	2.033	0.822	1.586	0.391
3402	0.785	3.552	1.565	0.777	3.372	0.264
3403	0.102	2.434	3.236	2.160	3.709	0.998
3404	0.174	1.684	4.382	1.046	2.345	1.321
3405	0.473	1.859	4.141	0.859	1.817	1.675
3406	0.230	2.318	0.554	0.585	0.700	0.107
3407	0.986	2.896	1.061	0.483	0.615	0.244
3408	0.655	1.524	0.928	0.391	0.732	0.287
3409	0.507	2.217	1.571	0.440	0.826	0.277
3410	0.158	2.459	2.577	0.659	2.060	0.403
3411	1.664	2.952	0.343	1.186	1.522	0.126
3412	1.848	1.466	0.159	0.566	0.915	0.012
3413	5.101	0.358	0.024	0.614	0.106	0.000
3414	3.646	1.304	0.277	0.739	0.468	0.017
3501	0.002	0.086	0.328	1.837	3.432	2.375
3502	0.001	0.005	0.717	1.860	2.531	5.320
3503	0.011	0.241	0.951	0.738	1.504	1.252
3504	0.015	0.509	1.335	0.231	0.933	0.405
3505	0.054	1.953	1.389	0.224	1.275	0.219
3506	0.019	1.734	0.445	0.091	0.383	0.069
3507	0.573	0.388	0.071	0.071	0.133	0.017
3508	0.441	0.244	0.013	0.090	0.204	0.073
3509	0.015	0.517	0.659	1.104	0.934	2.548
3510	0.018	0.034	0.381	10.186	5.294	10.578
3511	0.271	0.141	0.005	0.034	0.035	0.047
3512	0.485	0.802	0.205	0.285	0.278	0.094
3513	1.522	2.667	0.054	0.250	0.242	0.014
3514	0.073	0.184	0.232	0.822	1.372	0.685
3515	1.549	0.235	0.039	0.564	0.623	0.113
3516	2.712	0.058	0.000	0.047	0.009	0.004
3517	1.439	1.711	0.304	0.125	0.148	0.014
3518	2.306	0.040	0.001	0.001	0.004	0.005
3519	2.115	0.162	0.004	0.044	0.009	0.006

*See Appendix I for PUMA names and their locations.

APPENDIX II: IHS Housing Unit Weights by Property Type and Three Price Tiers in PUMA (Continued)

PUMA*	2-4 UNIT Multifamily			5+ UNIT Multifamily		
	Tier1) (0- 30%)	Tier2 (30-70%)	Tier3 (70-100%)	Tier1 (0- 30%)	Tier2 (30-70%)	Tier3 (70-100%)
3401	0.002	0.034	0.096	0.111	1.028	0.138
3402	0.005	0.054	0.006	0.138	0.464	0.272
3403	0.015	0.027	0.168	0.467	2.036	0.581
3404	0.027	0.102	0.323	0.262	1.912	0.393
3405	0.073	0.591	1.361	0.491	0.763	1.300
3406	0.062	0.498	0.486	0.726	2.137	0.343
3407	0.524	1.505	0.460	2.124	1.216	0.464
3408	1.738	3.395	0.558	1.260	1.452	0.827
3409	0.224	0.741	0.404	0.544	1.341	0.339
3410	0.032	0.117	0.142	0.222	1.210	0.198
3411	0.658	0.370	0.154	1.455	2.013	0.454
3412	0.153	0.079	0.021	0.628	0.595	0.171
3413	1.282	0.414	0.016	2.688	0.447	0.215
3414	0.721	0.085	0.014	0.571	0.185	0.151
3501	0.009	0.123	1.928	0.114	1.243	4.641
3502	0.003	0.029	3.792	0.067	0.544	4.735
3503	0.031	0.176	4.985	0.383	1.482	2.258
3504	0.041	1.881	2.397	0.255	1.529	1.243
3505	0.042	0.608	1.636	0.343	1.593	0.232
3506	0.067	3.107	1.917	0.259	1.667	0.578
3507	2.468	1.924	0.046	0.948	0.860	0.807
3508	2.175	5.329	0.163	2.527	0.951	0.554
3509	0.101	4.424	6.604	0.554	4.933	1.855
3510	0.020	0.164	0.832	0.474	0.649	1.579
3511	2.388	2.788	0.261	1.774	0.497	0.252
3512	3.164	4.998	0.792	0.985	1.048	0.185
3513	1.539	2.225	0.142	1.028	1.169	0.350
3514	0.516	0.716	0.174	1.737	0.911	1.502
3515	2.785	1.981	0.032	3.451	2.221	2.241
3516	4.791	0.550	0.011	1.680	1.075	0.598
3517	0.156	0.276	0.080	0.326	0.356	0.178
3518	1.701	0.178	0.001	0.591	0.178	0.175
3519	2.489	0.507	0.004	0.820	0.306	0.178

*See Appendix I for PUMA names and their locations.