

Description of IHS Hedonic Data Set and Model Developed for PUMA Area Price Index

Introduction

Understanding and measuring house price trends in small geographic areas has been one of the most challenging, but important topics in housing research recently. As national and metropolitan area-level housing markets emerge from the housing downturn, many neighborhoods within those broader geographies have not fully benefitted from this recovery. Being able to measure house price changes in small geographic areas can help housing market stakeholders and policy makers understand which neighborhoods are improving and which continue to struggle so they can make strategic decisions about policy development and implementation.

The following technical paper lays out the methodology used to develop IHS's Cook County Submarket Price Index which is based on a hedonic price model tracking price trends for single family homes.

Background - Measuring House Prices

There are three common methods for measuring house prices and price trends. Each has inherent strengths and limitations, particularly when applied to smaller geographies. These methods include:

- » **Median sales price** – This method looks at all sales taking place in a given geography for a given period of time and tracks the median value of those sales over time. The primary strengths of this method are 1) that data on sales activity and prices are often easily available through local deed transfer recordings or multiple listing services and 2) finding the median is a fairly straightforward and simple calculation. For these reasons, trends in median sales prices are often used by local realtor groups or the media to discuss area house price and trends. The main limitation of this method is that there is no way to control for changes in the underlying composition of properties selling at any two points in time. This has the potential to create “apples to oranges” price comparisons if there are large differences in the mix of the size and quality of properties selling at two points in time. This can be particularly impactful when sample sizes are small such as in small geographic areas.
- » **Repeat sales index** – Repeat sales indices take the sales activity on a property at two points in time and measures the change in value over that period. The change is weighted based on the length of time between the two sales, and the average change in sales prices for all properties in a sample are calculated and indexed to an earlier point in time, often the first quarter of 2000. The repeat sales index is an improvement over median sales price in many ways. By only tracking price changes for properties that sell multiple times, a repeat sales index is better able to ensure that the price change being measured is for properties with the same characteristics. Repeat sales indices also have limitations, however. Most importantly, because the sample uses only properties that sell at least twice, it is often difficult to get a large enough sample of property sales for a given period to measure price trends in a small geographic area. Case Shiller is the best known repeat sales index, and it tracks price trends nationally for a group of large metropolitan areas. Using a similar methodology, the Institute for Housing Studies has a Cook County House Price Index that tracks price changes in Cook County and in very large submarket areas.
- » **Hedonic price index** – Hedonic price indices combine information on property’s sales price with data on the characteristics of that individual property and its location and controls for factors that might affect the sales price of a house. A hedonic model tells you how much influence those individual factors have on sale prices, and, by isolating the effect of those variables on price, allows for the development of an index tracking price changes over a period of time on properties with similar characteristics. Hedonic price models are an improvement over repeat sales models because they include data on a far larger group of sales in a given period of time for a geography, not just those with previous sales. This allows for a larger sample in smaller geographic areas while still controlling for the characteristics and location of the properties being sold in a given period. While hedonic price index models have many advantages, there are also limitations. Hedonic indices require extensive amount of data on property characteristics and location, and developing such a data set is complex and can have extensive upfront costs. Additionally, hedonic models are the most statistically sophisticated of the three methods of tracking housing prices and require significant expertise to develop and extensive testing and monitoring to ensure accuracy.

IHS Hedonic House Price Index

Because of its advantages in tracking small area price trends, IHS developed a hedonic price index to track price changes in smaller geographies. The Institute has adapted this model to track changes in single family house prices in Cook County submarket areas defined by Census Public Use Microdata Areas (PUMAs). PUMA areas contain at least 100,000 people and are built up out of census tracts. There are 17 submarkets in the City of Chicago and 17 that are primarily in suburban Cook County. In the City of Chicago, the submarket surrounding the Loop has been excluded because of insufficient levels of single family home sales. Figure 1 displays all submarkets.

The following sections lay out the data used, variables developed, and the hedonic model used:

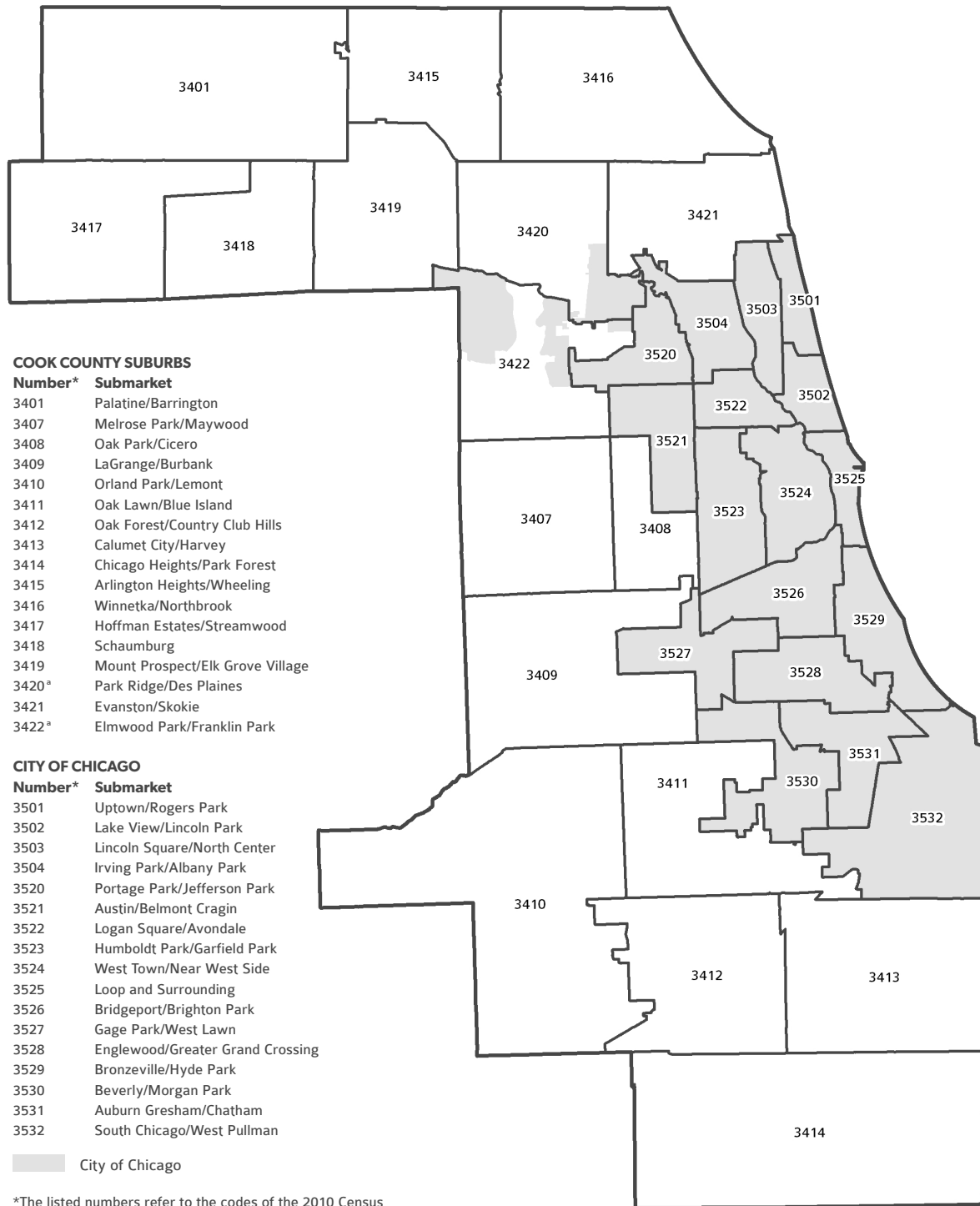
Data and Methodology

A review of existing literature on hedonic models was used to identify a core set of variables related to which property and location characteristics significantly influence house price. Figure 2 highlights variables included in the IHS hedonic model. These data include:

- » **Sales price** - Data on single family sales activity was taken from two sources, 1) property transfer records the Cook County Recorder of Deeds via Property Insight and 2) sales records from Midwest Real Estate Data (MRED), the northwest Illinois Multiple Listing Service (MLS).
- » **Property characteristics** – To identify key physical characteristics of properties such as the building structure, square feet, number of bathrooms and bedroom, age of the building, data from the Cook County Assessor and the northwest Illinois MLS were used.
- » **Location** - Geographic variables were calculated using ArcGIS software. These variables include distance from properties to Chicago Transit Authority (CTA) rail stations, to Lake Michigan, to any type of publicly-accessible open space, to Metra rail stations, and to a lake or river other than Lake Michigan. Spatial data for parcels is obtained annually by IHS from the Cook County Assessor. Distances to CTA and Metra rail stations were calculated by joining the Cook County road network from the Cook County Data Portal and CTA and Metra rail station locations obtained from the City of Chicago Data Portal. Lake Michigan, publicly-accessible open space, and lakes and rivers other than Lake Michigan come from the Chicago Metropolitan Agency for Planning (CMAPs) land use file for 2005.
- » **Distressed sales** - Properties that were likely distressed sales were also flagged. This includes properties identified as short sales, sales at foreclosure auction, and sales occurring after a property entered bank real estate owned (REO) status. Foreclosure distressed status was determined by identifying the date of a foreclosure filing on a property and tracking subsequent transaction activity. These data come from the Cook County Clerk of the Court and Cook County Recorder of Deeds via Property Insight.
- » **Fixed Effects** - All results are controlled by the fixed effect of geographical area (Census Tract) and time of sales (year and quarter).

Reference Map of Cook County Housing Submarkets

FIGURE 1



*The listed numbers refer to the codes of the 2010 Census PUMAs upon which the listed housing submarkets are based.

*PUMAs 3420 and 3422 include parts of the City of Chicago.

Figure 2) Descriptions of Variables

Variable Name	Description of Variable
Sale Price and Distressed Sale	
house price	House Price Sold (\$)
log_price	Log of House Price
dsale	=1 if sold as a distress sale (Short Sale, Foreclosure, REO)
Property Characteristics	
sqft	Square Feet of Building Area
lotsize	Square Feet of Lot Size
log_sqft	Log of Square Feet of Building Area
log_lot	Log of Square feet of Lot Size
bedroom	Number of Bedrooms
bathroom	Number of Bathrooms (Full.Half)
totalroom	Total Number of Rooms in the Property
garage	Number of Cars in Garage
brick	=1 if full or partial Brick Building)
age	Building Age or Age after Improvement
age_sq	Squre of building age
centralair	=1 if Central Air conditioning
fireplace	Number of Fireplace
Location and Distance Variables	
waterfront	=1 if located at waterfront
cta_stop	=1 if within 660 feet near CTA Station
cta_nearstop	=1 if within 661 to 1320 feet near CTA Station
cc_cal_distance	Distance from the Central Business District (CBD)
metra_stop	=1 if located within a quarter mile
pubopen	=1 if having a public open space within 660 feet
michlake	=1 if located within 1 mile from Lake Michigan
lake_river	=1 if located within 660 feet from river and lake

Building a final data set for the base hedonic model required creating a large master data set. To start, there were 833,821 detached single family property transactions recorded in Cook County from 1997 to the fourth quarter of 2014. Hedonic variables were constructed for each property using methodologies described above. Properties where transactions repeated within 90 days were excluded to avoid any potential recording errors and to reduce potential bias in the index due to frequently traded properties. Additionally, transactions were dropped if there was found to be substantial missing information on essential property characteristics such as the number of bedrooms, existence of an air conditioning system, or because of errors such as missing property identification numbers, or conflicting sales price information. The overall sample rate is 75.7 percent for the entire sample periods of 1997 to 2014, and Figure 3 shows the annual total number of valid observations included in the IHS hedonic model data set is 631,589. The valid sample rate is substantially higher starting in 2009 where over 86 percent of transactions match hedonic variables in each year.

Figure 3) Cook County Single Family Sample Data with Hedonic Variables

Year	Total Transactions	Variables	Valid Sample Rate
1997	42,562	31,057	73.0%
1998	36,909	27,149	73.6%
1999	49,978	35,562	71.2%
2000	52,659	37,153	70.6%
2001	56,306	39,517	70.2%
2002	56,627	41,714	73.7%
2003	68,339	45,993	67.3%
2004	79,253	51,019	64.4%
2005	72,586	54,095	74.5%
2006	58,281	44,457	76.3%
2007	39,382	31,616	80.3%
2008	26,823	22,316	83.2%
2009	28,105	24,201	86.1%
2010	28,142	24,341	86.5%
2011	26,416	23,285	88.1%
2012	33,297	29,586	88.9%
2013	41,382	36,457	88.1%
2014	36,774	32,071	87.2%
1997-2014	833,821	631,589	75.7%

Calibrating the Small Area Index

Even in a hedonic model, a sufficient sample size is required to consistently and accurately track price trends. While a sample size of 631,589 records is sufficient to produce a quarterly hedonic house price index for the entire Chicago area, large variation in levels of transaction activity made it challenging to produce quarterly updates for small geographies. To compensate for declining transaction volume and the lower number of transaction in small geographies, a rolling sample method with a 365 day window was adopted. This means that in addition to data from the current quarter, sales data from the previous three quarters were also included. Additional data from previous quarters helps smooth out the volatile nature of transaction activity in small areas from quarter to quarter. Due to the lack of single family houses in Chicago downtown area, PUMA 3525 is excluded. Valid sample sizes for all other submarket areas from 1997 to 2014 can be found in Appendix A. Sample sizes are much smaller for submarkets in the City of Chicago compared to those in Suburban Cook County. This is due to the more diverse housing stock found in many Chicago neighborhoods which include both small and large multi-unit rental properties and condominium units which are not included in this hedonic model. All PUMAs included had large enough valid samples of single family sales to produce stable trends.

Results of the Model

The results of the hedonic models for Cook County, Chicago, and suburban Cook are shown in Figure 5. The results for most of the individual independent variables are statistically significant and the magnitude and direction of their effect on house prices are consistent with expectations. The results are also largely consistent across geographic regions of Cook County. The r-square for all three models is roughly .77, which indicates the included control variables explain the house price variation strongly. R-squared results for individual submarket areas are not shown, but they are at acceptable levels ranging from 0.53 to 0.80.

Figure 5) Hedonic Regression by Geographic Area, 1997Q1-2014Q3

	Cook County	Chicago	Suburb Cook
log sqft	0.317***	0.289***	0.320***
	-298.05	-145.28	-256.67
log lot	0.113***	0.157***	0.105***
	-200.01	-92.94	-182.32
bedroom	0.015***	0.018***	0.015***
	-36.87	-27.33	-29.83
bathroom	0.006***	0.018***	0.002**
	-12.57	-17.85	-2.87
totalroom	0.028***	0.019***	0.032***
	-138.41	-52.26	-133.04
garage	0.036***	0.037***	0.034***
	-102.21	-59.34	-79.88
brick	0.024***	0.007***	0.029***
	-43.48	-6.41	-47.96
age	-0.004***	-0.001***	-0.005***
	(-117.34)	(-13.15)	(-122.35)
age sq	0.000***	-0.000***	0.000***
	-69.26	(-10.23)	-76.37
waterfront	0.064***	0.019***	0.073***
	-31.36	-3.87	-33.45
centralair	0.070***	0.061***	0.079***
	-105.48	-53.75	-97.07
fireplace	0.036***	0.033***	0.036***
	-79.48	-35.69	-72.88
cta stop	-0.070***	-0.039***	-0.138***
	(-12.19)	(-5.09)	(-14.45)
cta nearstop	-0.035***	-0.014***	-0.081***
	(-12.06)	(-3.66)	(-15.26)
cc calc dist	0.017***	0.016***	0.016***
	-28.35	-6.94	-26.3

	Cook County	Chicago	Suburb Cook
metra stop	-0.004	0.007*	-0.004
	(-1.82)	-2.09	(-1.44)
pubopen	0.003***	0.009***	0
	-5.75	-7.65	-0.69
michlake	0.035***	-0.024***	0.074***
	-9.4	(-3.86)	-16.09
lake river	0.015***	0.024***	0.012***
	-16	-7.77	-13.41
dsale	-0.121***	-0.158***	-0.098***
	(-102.06)	(-82.24)	(-64.36)
ds2007	-0.055***	-0.064***	-0.063***
	(-18.36)	(-12.86)	(-16.78)
ds2008	-0.329***	-0.462***	-0.241***
	(-118.97)	(-98.04)	(-70.40)
ds2009	-0.487***	-0.640***	-0.389***
	(-193.44)	(-145.76)	(-126.89)
ds2010	-0.432***	-0.545***	-0.359***
	(-168.62)	(-121.82)	(-115.40)
ds2011	-0.399***	-0.472***	-0.361***
	(-150.34)	(-98.53)	(-114.15)
ds2012	-0.385***	-0.440***	-0.360***
	(-157.95)	(-98.13)	(-125.34)
ds2013	-0.369***	-0.436***	-0.340***
	(-157.55)	(-101.25)	(-122.47)
ds2014	-0.364***	-0.452***	-0.325***
	(-94.78)	(-62.54)	(-72.99)
N	2,460,333	798,395	1,661,938
R2	0.774	0.772	0.775

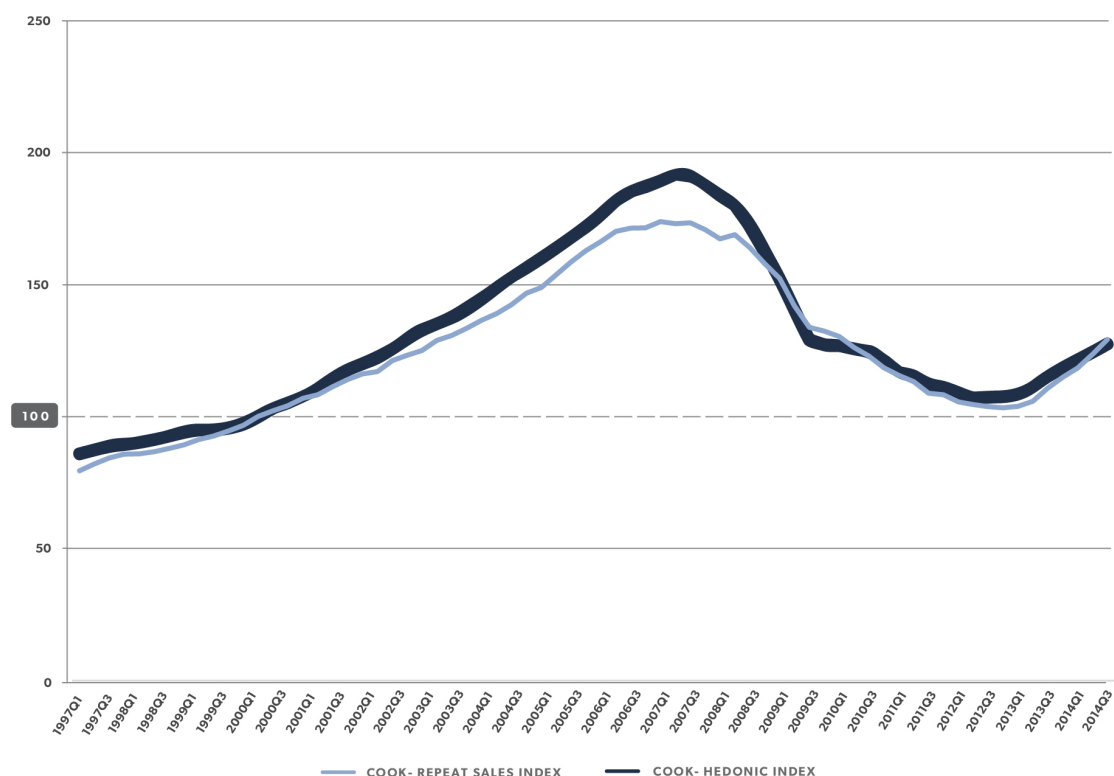
ALL RESULTS ARE CONTROLLED BY THE FIXED EFFECT OF GEOGRAPHICAL AREA (CENSUS TRACT) AND TIME OF SALES (YEAR AND QUARTER).
ALL T-STATISTICS ARE CALCULATED USING HETEROSKEDASTICITY CORRECTED ROBUST STANDARD ERRORS

* : 10 %, **: 5%, ***: 1% SIGNIFICANT

Interestingly, the distressed sale dummy variable returns highly significant results. The coefficients are very stable for the three geographic regions ranging from -0.533 in suburban Cook to -0.627 in the City of Chicago. Another way to state this would be assuming a median sale price of \$187,500, the impact of a distressed sale would drop the price to \$100,161 in the City of Chicago or to \$110,032 in suburban Cook County. This means the value of a distressed property will be depreciated by 41 to 47 percent compared with a non-distressed property, respectively. The interactive dummy variable of distress sale by year after 2007 show the significant from log of -0.3 to -0.4. By controlling these annual distressed sales, we are able to calculate the general price changes after the financial crisis. Without the distressed dummy variables, there might be downward bias on the general house price trend due to relatively high concentrated distressed sales after 2007, particularly in certain areas with high levels of distressed sales. For example, if a community has higher level of distresses sales while housing turnover rate relative low, the transactions from the distresses sales will be over-represented in the price index and that will create downward bias in overall price trends.

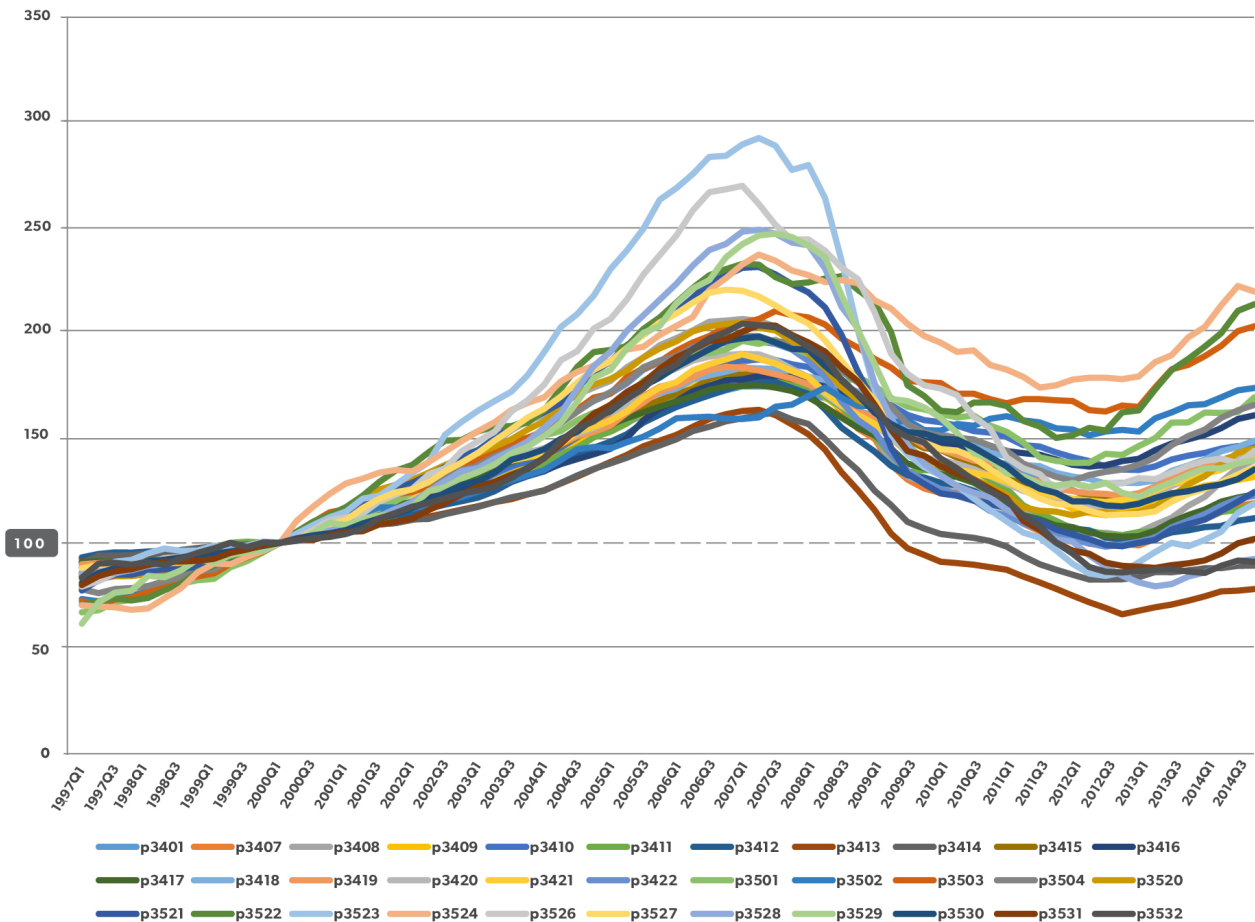
The results of the Cook County-wide hedonic model are generally consistent with those found in other price index models for Cook County. Figure 6 compares quarterly price trends in Cook County calculated by this hedonic price index and IHS's Cook County House Price Index which is calculated using a repeat sales methodology. As the figure shows, the direction of quarterly price changes are generally consistent for the hedonic and repeat sales indices. The hedonic index shows a more substantial price build up leading to the market peak in 2007, but after the market collapse in 2008, price trends for the hedonic and repeat sales models track fairly closely. One possible explanation for why the hedonic model saw greater price increases leading to the market peak is that it includes all property sales, including new properties which may be of higher quality. Because a newly constructed property only counts as one transaction (and doesn't yet have a repeat sale) it is not captured in the repeat sales index which only tracks sales of existing, older properties.

Figure 6) Comparison of Cook County Repeat and Hedonic Price Index Results

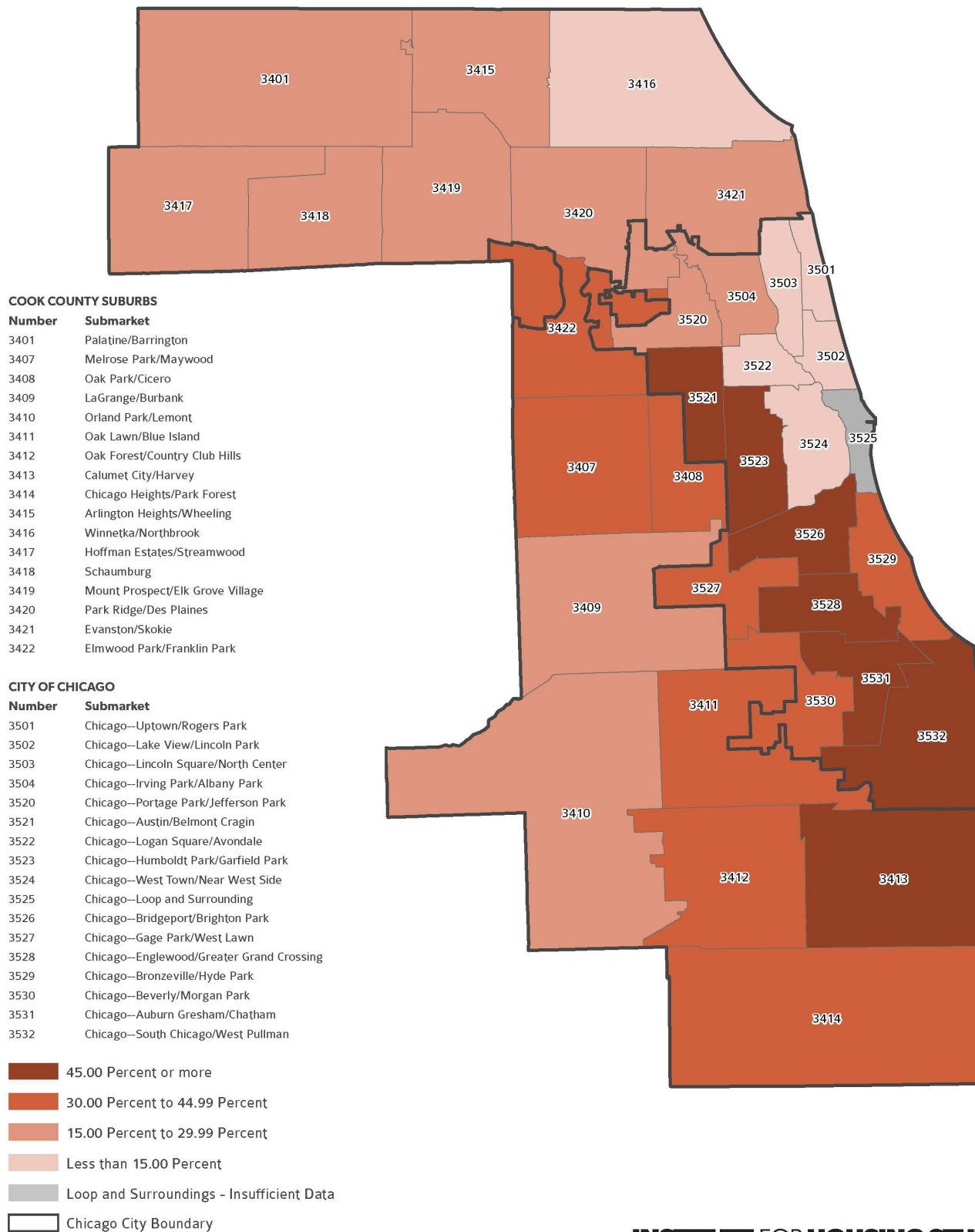


Similarly to how results from the hedonic model can be converted to track changes in house prices countywide, they can also be converted to track price changes at the submarket level. The estimated average price level at time t on the condition of all control variables was used to build hedonic price index. To create a relative measurement compare to other time. The trend lines for all submarkets can be seen in Figure 7.

Figure 7) Hedonic House Price Index by Cook County Submarket, 1997-2014



IHS Price Index - Decline from Peak to Q4 2014 Cook County Submarkets



SOURCES: IHS DATA CLEARINGHOUSE

**INSTITUTE FOR HOUSING STUDIES
AT DEPAUL UNIVERSITY**

INSTITUTE FOR HOUSING STUDIES AT DEPAUL UNIVERSITY

HOUSINGSTUDIES.ORG

APPENDIX A

	03401	03407	03408	03409	03410	03411	03412	03413
	Palatine/ Barrington	Melrose Park/ Maywood	Oak Park/ Cicero	LaGrange/ Burbank	Oriand Park/ Lemont	Oak Lawn/Blue Island	Oak Forest/ Country Club Hills	Calumet City/ Harvey
1997	1,291	1,525	1,131	1,316	1,695	1,192	962	1,616
1998	1,288	1,369	1,003	1,145	1,447	1,040	763	1,297
1999	1,474	1,778	1,327	1,510	1,678	1,367	1,117	1,814
2000	1,664	1,770	1,441	1,628	1,836	1,471	1,228	1,939
2001	1,623	1,962	1,546	1,743	1,879	1,675	1,384	1,924
2002	1,592	1,961	1,633	1,867	1,991	1,740	1,390	2,168
2003	1,708	2,230	1,771	2,055	2,123	1,789	1,611	2,401
2004	1,613	2,526	1,926	2,179	2,142	1,960	1,772	3,098
2005	1,641	2,746	2,038	2,212	2,187	2,148	1,898	3,413
2006	1,360	2,159	1,638	1,776	1,731	1,848	1,632	3,255
2007	995	1,534	1,077	1,184	1,244	1,287	1,113	2,256
2008	720	999	829	834	980	897	774	1,534
2009	750	1,195	1,014	963	962	981	809	1,426
2010	718	1,208	985	1,041	952	930	728	1,337
2011	726	1,240	962	994	908	979	706	1,089
2012	1,068	1,698	1,237	1,323	1,197	1,255	901	1,331
2013	1,404	2,169	1,572	1,588	1,481	1,595	1,229	1,728
2014	1,230	1,753	1,348	1,412	1,458	1,460	1,105	1,455
Total	22,865	31,822	24,478	26,770	27,891	25,614	21,122	35,081

	03414	03415	03416	03417	03418	03419	03420	03421	03422
	Chicago Heights/Park Forest	Arlington Heights/Wheeling	Winnetka/Northbrook	Hoffman Estates/Streamwood	Schaumburg	Mount Prospect/Elk Grove Village	Park Ridge/Des Plaines	Evanston/Skokie	Elmwood Park/Franklin Park
1997	1,770	1,019	1,508	1,421	1,076	835	1,539	1,292	887
1998	1,304	959	1,633	1,351	907	732	1,466	1,333	768
1999	2,036	1,018	1,906	1,657	1,010	886	1,824	1,514	1,020
2000	2,074	1,062	1,873	2,054	1,075	925	1,829	1,584	1,042
2001	2,244	1,081	1,988	1,977	1,072	992	1,970	1,634	1,201
2002	2,333	1,115	1,998	2,064	1,183	1,077	1,961	1,745	1,149
2003	2,821	1,247	2,125	2,089	1,353	1,101	2,183	1,922	1,262
2004	3,330	1,386	2,007	2,298	1,359	1,170	2,324	2,083	1,411
2005	3,425	1,299	2,111	2,507	1,419	1,117	2,268	1,969	1,485
2006	3,290	1,014	1,504	1,810	1,118	890	1,809	1,603	1,161
2007	2,316	826	1,279	1,262	774	768	1,338	1,230	767
2008	1,497	561	1,007	835	579	533	986	966	528
2009	1,497	633	1,031	822	627	538	1,199	1,038	778
2010	1,434	649	1,287	802	509	605	1,220	1,142	815
2011	1,222	639	1,311	850	554	559	1,242	1,099	855
2012	1,619	932	1,585	1,123	827	793	1,646	1,402	1,021
2013	2,004	1,131	1,892	1,403	999	1,095	1,927	1,621	1,217
2014	1,668	949	1,671	1,201	897	920	1,780	1,504	1,026
Total	37,884	17,520	29,716	27,526	17,338	15,536	30,511	26,681	18,393

	03501	03502	03503	03504	03520	03521	03522	03523
	Chicago-- -Uptown/ Rogers Park	Chicago--Lake View/Lincoln Park	Chicago-- Lincoln Square/ North Center	Chicago--Irving Park/Albany Park	Chicago-- Portage Park/ Jefferson Park	Chicago-- Austin/Beimont Cragin	Chicago--Logan Square/Avondale	Chicago-- Humboldt Park/ Garfield Park
1997	201	505	456	553	922	777	353	227
1998	147	423	407	516	820	610	372	159
1999	176	537	557	691	1,122	931	517	270
2000	172	487	580	662	1,109	909	556	306
2001	160	603	562	696	1,169	986	528	362
2002	206	519	648	710	1,219	1,090	601	385
2003	222	510	675	818	1,292	1,229	582	457
2004	272	526	649	862	1,318	1,445	611	602
2005	260	547	737	921	1,483	1,476	675	682
2006	189	428	559	679	1,108	1,216	528	584
2007	152	359	465	587	728	648	349	323
2008	107	232	367	425	537	447	244	216
2009	122	232	376	446	752	661	313	235
2010	104	284	374	491	775	715	315	236
2011	127	275	386	478	734	626	347	195
2012	123	349	441	603	916	713	397	246
2013	188	416	589	765	1,090	817	460	285
2014	178	353	469	650	961	749	403	247
Total	3,106	7,585	9,297	11,553	18,055	16,045	8,151	6,017

	03524	03526	03527	03528	03529	03530	03531	03532
	Chicago--West Town/Near West Side	Chicago-- Bridgeport/ Brighton Park	Chicago--Gage Park/West Lawn	Chicago-- Englewood/ Greater Grand Crossing	Chicago-- Bronzeville/ Hyde Park	Chicago- -Beverly/ Morgan Park	Chicago--Auburn Gresham/ Chatham	Chicago--South Chicago/West Pullman
1997	322	407	1,037	586	256	1,181	538	661
1998	248	269	893	450	255	901	421	453
1999	411	417	1,217	793	324	1,275	687	701
2000	332	406	1,163	819	296	1,245	823	793
2001	423	467	1,507	869	306	1,348	801	835
2002	411	526	1,641	1,084	371	1,426	954	956
2003	399	618	1,765	1,385	386	1,654	1,120	1,090
2004	388	673	1,993	1,705	510	1,842	1,554	1,485
2005	464	791	2,170	1,994	594	2,050	1,710	1,658
2006	418	678	1,515	1,565	550	1,768	1,592	1,482
2007	326	450	1,001	1,091	354	1,292	1,170	1,071
2008	230	288	674	706	308	934	834	708
2009	216	301	889	634	264	1,032	796	669
2010	231	314	905	610	257	1,018	709	641
2011	227	304	879	495	234	1,004	576	463
2012	297	326	1,129	467	269	1,060	677	615
2013	349	389	1,298	564	316	1,357	818	701
2014	298	361	1,072	537	296	1,229	791	640
Total	5,990	7,985	22,748	16,354	6,146	23,616	16,571	15,622