# Chicago Submarket House Price Movements

Dylan Hall
dylan.hall@chicagobooth.edu
University of Chicago Booth School of Business
June 12, 2015

## Project Description

House price movements were one of the causes of the Great Recession. House price models inaccurately forecasted growth and homeowners and investors bore the brunt of these unanticipated fluctuations. This project attempts to better understand house price movements by investigating 14 Chicago submarkets. It employs and contrasts two approaches: vector autoregressive (VAR) models that employ serial cross-submarket correlations, and a lasso regression approach that makes use of attributes of the submarkets such as changes in housing market characteristics, crime, building permits and income. Each approach encounters several limitations but provide different meaningful outcomes that can contribute to understanding of house price fluctuations and the construction of house price models.

## Executive Summary

The investigation of house price movement in Chicago's submarkets is conducted via two forms of linear modeling. VAR models demonstrate the interlocking relationships between local housing markets. Lasso linear regression models produce individual market characteristics that are predictive of house price changes.

***VAR Models***

Employing 16 years of submarket price observations (1997 – 2013), I create VAR models on subsets of submarkets and explore short-term lags. In strong recovery housing markets, these VAR models produce interesting outcomes. The innovations suggest a momentum effect with innovations having a significant positive impact on prices in the short run (0 – 4 years). The innovations have a negative impact in the medium term (5 – 10 years), and finally over the long run settling at slight 2% growth. This has interesting implications for homeowners, as they bear the risk of these house price changes which are typically magnified through the use of leverage.

Further, weak recovery and strong recovery VAR models together demonstrate high levels of serial and cross-submarket dependency that are difficult to remove, and speak to the difficulty in building robust house price models.

***Year-over-Year Lasso Regression***

Unlike the VAR models, the lasso regression employs a naïve approach that does not account for serial and cross submarket dependencies. Doing so allows for greater degrees of freedom and the opportunity to investigate impacts of other characteristic changes on a submarket's change in price. The model identifies certain changes to be predictive, these include:

*Negative Price Impacts*

- Increase in all cash buyers (+)
- Increases in theft (+)

*Positive Price Impacts*

- Increase in Easy Building Permits (+)
- Increase in new construction Permits (+)
- Increase in business establishments
- Increase in REO properties

*(+) attributes that are significant for <u>both</u> lasso and cross-validated lasso regression*

The lasso regression model has shortcomings that extend beyond dynamic dependence. It underpredicts in instances of large absolute growth (either positive or negative) suggesting that there are other important factors left outside of the model that are predictive in price changes, and/or that the linear approach is limiting.

## Empirical Datasets

The Chicago Submarket Prices Indices (see Appendix) are produced by the DePaul Institute for Housing Studies.[1] The data is available on a quarterly basis for 14 Chicago submarkets and is a repeat sales index constructed in a manner similar to the Case Shiller index. I use annual changes on the fourth quarter (Q4) to reduce possible effects from seasonality. Quarterly models were unable to improve the degrees of freedom limitations.

In addition to the price index, the lasso regression also makes use of the following datasets with years ranging from 2007 to 2013. The crimes and permit datasets are at the point level and have been aggregated to submarket levels using spatial joining techniques. The housing market characteristics have been aggregated from the community area to the submarket. Business and household income datasets are available at the zip code level, and have been allocated to the submarkets through the use of census Zip Code Tabulation Areas (ZCTA).

[1] "DePaul Institute for Housing Studies." Cook County House Price Index: Fourth Quarter 2013. Accessed June 9, 2015. Link.

*Crimes*[2]

- Theft (Robbery, Motor Vehicle Theft, Theft)
- Vice (Prostitution, Narcotics, Gambling)
- Behavior (Trespass, Weapons Violation, Interference with Public Officer)
- Property (Arson, Criminal Damage)
- Violence (Assault, Battery, Homicide, Sexual Assault)
- Business (Liquor Law Violation, Deceptive Practice)

*Building Permits*[3]

- Easy Building Permits (simple upgrades of existing structures in homes)
- New Construction
- Wrecking & Demolition
- Renovation and Alteration

*Housing Market Characteristics*[4]

- Foreclosures per parcel
- All cash buyers
- Mortgages per parcel
- Sales per parcel
- Extremely low value sales
- Distressed sales
- Business buyers
- Real-estate Owned (REO)

*Business and Household Income*[5]

- Number of tax filers
- Income (salaries and wages, capital gains and business Income)
- Number of business establishments
- Business establishment earnings before tax, depreciation and amoritization

---

[2] "Crimes - 2001 to Present | City of Chicago | Data Portal." Chicago. Accessed June 9, 2015. Link.
[3] "Building Permits | City of Chicago | Data Portal." Chicago. Accessed June 9, 2015. Link.
[4] "Data Portal." DePaul Institute for Housing Studies. Acessed June 9, 2015. Link.
[5] Proprietary dataset from Powerlytics, Inc that is based on Internal Revenue Service tax datasets.

## Analysis

### *VAR Models*

Two separate VAR models were constructed for different groups of submarkets.

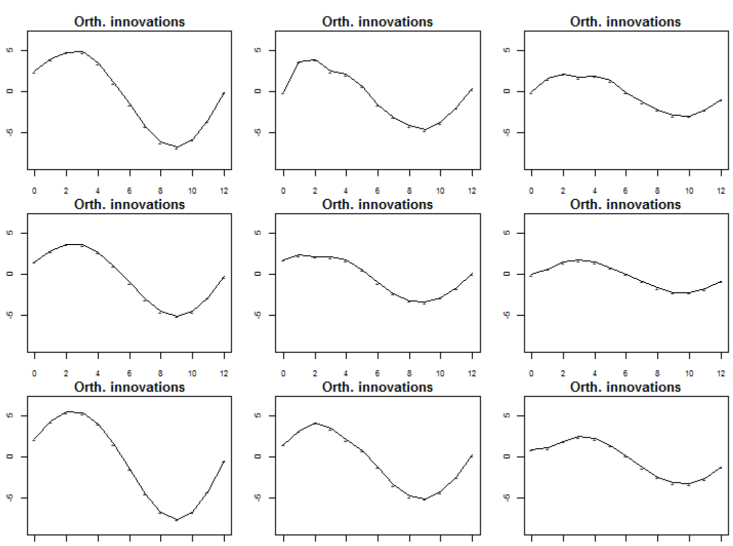
The first group consists of 3 submarkets that experienced a strong rebound from the Great Recession and include downtown and the near north areas of Chicago, areas which are more affluent in general. The refined VAR(2) model produces the best residuals according to MQ statistics.

*VAR(2) Model for Strong Recovery Submarkets*

$$\begin{bmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{bmatrix} = \begin{bmatrix} 3.3 \\ 3.2 \\ 3.2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1.7 \\ 0 & 0.6 & 1.8 \\ 0 & 0 & 1.8 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \\ z_{3,t-1} \end{bmatrix} + \begin{bmatrix} -0.6 & -0.8 & 0 \\ 0 & -1.1 & 0 \\ -0.5 & -0.9 & 0 \end{bmatrix} \begin{bmatrix} z_{1,t-2} \\ z_{2,t-2} \\ z_{3,t-2} \end{bmatrix}$$

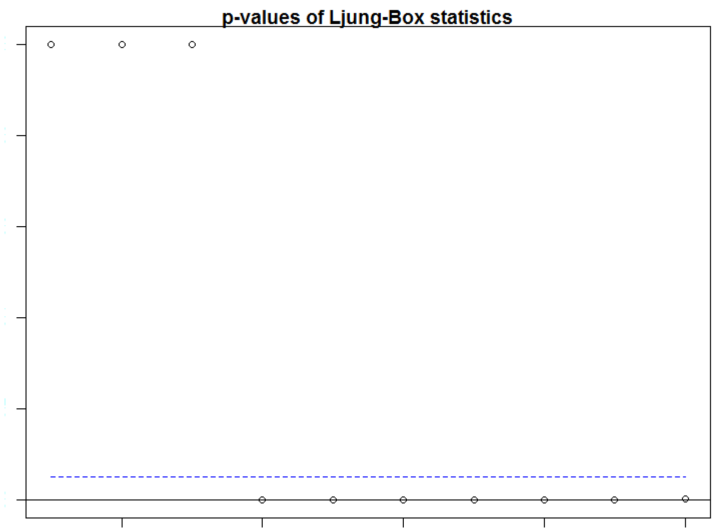
Submarket 3 (West Town / Logan Square) has a positive impact on all submarkets in period t – 1, and is the only submarket with an impact in this time period. Submarket 1 (Loop / Hyde Park) and Submarket 2 (Lakeview / Lincoln Park) had negative impacts on the other submarkets in period t – 2. The resulting impulse response functions result in a series of fluctuations with the positive peak occurring in year 2,-4 negative peak occurring in years 7 – 9, and reaching a net gain by years 10 – 12.

*Plot 1: Impulse Response Functions for Strong Recovery Submarkets*



The second VAR model consists of 3 submarkets that have not rebounded from the Great Recession. These are south and west of downtown Chicago. A refined VAR(3) model produces the highest AIC, though it still exhibits strong dynamic dependence amongst the residuals. This dynamic dependence is evident in lower-order VAR models as well. It is not possible to expand to VAR(4) due to the limitation of the degrees of freedom.

*Plot 2: Persistent Dynamic Dependence in Weak Recovery Submarket Model Residuals*




*VAR(3) Model for Weak Recovery Submarkets*

$$\begin{bmatrix} z_{4,t} \\ z_{5,t} \\ z_{9,t} \end{bmatrix} = \begin{bmatrix} -6.0 \\ -8.7 \\ 0 \end{bmatrix} + \begin{bmatrix} 2.1 & -1.7 & 0.3 \\ 1.9 & -0.8 & 1.8 \\ 2.7 & -1.2 & -0.5 \end{bmatrix} \begin{bmatrix} z_{4,t-1} \\ z_{5,t-1} \\ z_{9,t-1} \end{bmatrix} + \begin{bmatrix} 0.9 & 0.6 & 0.6 \\ 0.9 & 0 & -0.7 \\ 0.5 & 0.9 & -0.4 \end{bmatrix} \begin{bmatrix} z_{4,t-2} \\ z_{5,t-2} \\ z_{9,t-2} \end{bmatrix} + \begin{bmatrix} -0.3 & 0 & -0.8 \\ 0 & -0.6 & 0 \\ -1.4 & 0.9 & -0.9 \end{bmatrix} \begin{bmatrix} z_{4,t-3} \\ z_{5,t-3} \\ z_{9,t-3} \end{bmatrix}$$

Submarket 4 (Little Village/Pilsen) has a cumulative positive impact on the submarkets, while Submarket 5 (Englewood) has a cumulative negative impact on the submarkets. Submarket 9 (Garfield Park) has a slight positive impact on the submarkets.

***Year-over-Year Lasso Regression***

$$y_{yoygrowth} = b + x_{yoygrowth}$$

The lasso regression consists of the percent change in a submarket's index regressed onto the percent change of its attributes. The attribute data spans years 2007 to 2013. With 14 submarkets and 6 years' worth of year-over-year differences, there are 84 observations total. The models below are selected using minimum AICC. The intercept term suggests an annual 4% drop in house prices perhaps reflecting the downward nature of house prices during the time period of investigation, and the poor performance of most but not all submarkets.

| **Lasso Regression** | | **Cross-validated Lasso Regression** | |
|---|---|---|---|
| **Variable** | **Coefficient** | **Variable** | **Coefficient** |
| Intercept | -4.05 | Intercept | -4.43 |
| Δ Foreclosures | . | Δ Foreclosures | . |
| Δ Sales | . | Δ Sales | . |
| Δ Mortgages | . | Δ Mortgages | . |
| Δ Low Value Sales | . | Δ Low Value Sales | . |
| Δ Distressed Sales | . | Δ Distressed Sales | . |
| Δ Business Buyers | . | Δ Business Buyers | . |
| Δ Cash Buyers | -0.52 | Δ Cash Buyers | -0.31 |
| Δ REO | 1.87 | Δ REO | . |
| Δ Theft (Crime) | -0.37 | Δ Theft (Crime) | -0.16 |
| Δ Violence (Crime) | . | Δ Violence (Crime) | . |
| Δ Behavior (Crime) | . | Δ Behavior (Crime) | . |
| Δ Vice (Crime) | . | Δ Vice (Crime) | . |
| Δ Property (Crime) | . | Δ Property (Crime) | . |
| Δ Easy Permits | 0.13 | Δ Easy Permits | 0.02 |
| Δ Construction | 0.07 | Δ Construction | 0.04 |
| Δ Demolition | . | Δ Demolition | . |
| Δ Renovation | . | Δ Renovation | . |
| Δ Firms | 0.05 | Δ Firms | . |
| Δ EBITDA | . | Δ EBITDA | . |
| Δ Filers | . | Δ Filers | . |
| Δ Income | . | Δ Income | . |

## Discussion

***VAR Models***

*Results*

The VAR models provide compelling observations:

- Local submarkets influence each other's prices in a bi-directional manner
- Price changes have cumulative effects that vary depending on the time period of examination. That is, positive price changes can have positive impacts in the short-term but negative price changes in the intermediate term.
- Submarket effects may vary depending on the nature of the submarkets involved. In this project, the refined strong recovery models looked incredibly different from the weak recovery models.
- It is difficult to remove serial and cross-submarket dependence from models.

These observations demonstrate the dangers in building house price models. There appear to be high levels of dynamic dependence, and even within the city of Chicago, the effects and their transmission can significantly vary depending on the submarket(s) of interest. This challenges the notion that a single model will effectively capture the nature of house prices, and that more specialized approaches are likely needed to effectively assess risk and returns in various submarkets. Further, the complicated nature of housing markets underscore the price risk homeowners take on in purchasing a home with a mortgage and the difficulty in diversifying away the price risk. Additionally, the alternating positive and negative impact of innovations over time in strong recovery submarkets demonstrate that homeowners encounter may encounter house price risk over different periods. A dangerous belief of the Great Recession was the persistent growth in house prices. Even the Strong Recovery model shows how price movements are not unidirectional and constant.

*Limitations*

The VAR models were clearly restricted by the limitations of the dataset. In particular, the limitations on degrees of freedom prevented efforts to better remove of dynamic dependence from either of the fitted models with VAR(3) models representing the greater order for analysis. It was additionally difficult to add attribute data, examine a larger group of submarkets or to entertain other models such as a moving average model that could assist in producing a more descriptive model that better captures the relationship of Chicago submarket prices changes.

*Improvement*

To overcome these limitations, further work should look at aggregate indices than span longer time periods such as those available from the Federal Housing Finance Agency (FHFA). Their availability at different micro- and metropolitan statistical areas will allow for adjacency comparisons similar to those employed here. National and metropolitan trends could also impact submarket trends as well, so it could prove informative to investigate national and Chicago MSA trends and their impact on individual Chicago submarkets, and not simply compare submarkets amongst each other. The intervals could be expanded as well (e.g. 5 year changes) to strip some of the more extreme events and better match homeowners and investor time horizons.

### *Year-over-Year Lasso Regression*

#### *Results*

The cross-validated lasso regression model produced individual factors that are predictive of submarket house price changes:

- Changes in all-cash buyers (+)
- Changes in theft (+)
- Changes in Easy Building Permits (+)
- Changes in new construction permits (+)
- Changes in REO properties
- Changes in number of establishments

These variables demonstrate that local housing markets are dynamic. New construction, new business establishments, property upgrades and theft may provide price signals to potential buyers. All cash buyers, who tend to be investors, may be more detrimental to neighborhood investment than mortgage buyers. It is important to distinguish though that while these attribute changes are predictive, they are not necessarily causal. For example, year-over-year changes in Easy Building Permits could be the result of increasing prices, as homeowners better see the possibility of capturing the returns of their home investment and undertake rehabilitation projects as a result which in further improves prices.

These dynamic effects also expose homeowners to risk. Homeowners do not really control the attributes found to be predictive of neighborhood price changes, and their investment is subject to price movements related to theft, construction, investment and business formation.

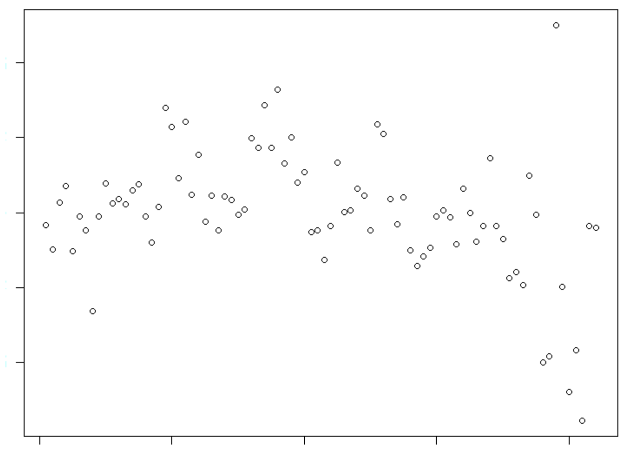
#### *Limitations and Improvements*

The predictive relationships could be biased from the house price index methodology. For example, consider the significance of REO properties. Their importance in the model may signify that REO properties have not been resold and have not affected the submarket prices yet since they may sell at a discount. An alternative index could be used like an appraisal-based index as opposed to repeat sales in the future.

Additionally, the model itself is limited. The submarket attributes could be expanded to include additional variables, or additional lags. It is conceivable that changes in household income or business sales could have an initial delay in effecting house prices that show up in future years.

From a modeling standpoint, the lasso regression does tend to under predict growth in extreme growth years. Plot 3 shows the downward bias in examining the difference between fitted and actual values. Improvements could be made in including serial lags and possibly weighted average lags of nearby submarkets since there are clearly larger effects at play.

*Plot 3: Biased Lasso Regression Residuals (Predicted % increase – Actual % increase)*



## Appendix

*Plot 4: Chicago Submarket Price indices over time*