Analysis of The Chicago Housing Market

Emily Nam

Fall 2024

Section I: Introduction

I. Economic Issues

Through this project, we will explore the factors currently influencing home values in the urban Chicago area; understanding these determinants is crucial to making informed decisions regarding property investments and urban development. This analysis will identify which structural, locational, and local government spending factors significantly impact home values.

II. Relevance & Interest

The real estate market in urban areas is dynamic and influenced by a multitude of factors, as well as being a major component of the economy. By investigating these determinants, we can gain insights into how different elements are contributing to property values. Having this knowledge is essential for real estate companies, urban planners, policymakers, and even potential homebuyers to know how to optimize their investment strategies, enhance urban infrastructure, and make policy decisions that are more educated and data-driven.

III. Prediction & Explanation

The primary objective of this econometric model will be to predict home values (contract sales price) using a range of explanatory variables; we will quantify the relationship between home values and factors such as the number of rooms, living area, house age, lot size, property tax rate, median income, distance from downtown, school spending, and municipal spending.

IV. Dependent & Explanatory Variables

Dependent Variable:

- SPRICE | Contract sales price of the house | *measured in dollars*

Explanatory Variables

- NROOMS | Number of rooms | measured in count

- LVAREA | Living area | measured in square feet
- HAGEEFF | Age of the house | measured in years
- LSIZE | Lot size | measured in square feet
- PTAXES | Property tax rate | measured in percentage
- MEDINC | Median income of the census tract | measured in dollars
- DFCL | Distance from the Loop area in downtown Chicago | measured in miles
- SSPEND | School district operating expenses per pupil | measured in dollars
- MSPEND | Municipal government expenditure per capita | measured in dollars

V. Original Population Regression Model

$$ln(SPRICE) = \beta 0 + \beta 1 ln(NROOMS) + \beta 2 ln(LVAREA) + \beta 3 ln(HAGEEFF) + \beta 4 ln(LSIZE) + \beta 5 ln(PTAXES) + \beta 6 ln(MEDINC) + \beta 7 ln(DFCL) + \beta 8 ln(SSPEND) + \beta 9 ln(MSPEND) + \epsilon$$

VI. Predicted Explanatory Variable Influence on Dependent

- NROOMS | positive relationship | More rooms generally increase the desirability and value of the house.
- LVAREA | positive relationship | Larger living areas typically increase home value.
- HAGEEFF | negative relationship | Older homes are less appealing due to potential maintenance issues and outdated features.
- LSIZE | positive relationship | Larger lot sizes are associated with higher property values.
- PTAXES | negative relationship | Higher property taxes less attractive to home buyers.
- MEDINC | positive relationship | Higher median income areas have higher property values due to greater purchasing power.
- DFCL | negative relationship | Greater distance from downtown decreases home value due to reduced accessibility to amenities and employment centers.

- SSPEND | positive relationship | Higher school spending indicates better quality education, this being attractive to families.
- MSPEND | positive relationship | Better municipal services enhance the living environment, increasing property values.

Section II: Assumptions & Initial Regression

I. Necessary Assumptions

- Linearity: The relationship between the dependent and independent variables is linear.
  - $y = \beta 0 + \beta 1 x1 + \beta 2 x2 + \cdots + \beta k xk + \epsilon$
- No Endogeneity: The explanatory variables are not correlated with the error term.
  - $Cov(xi, \epsilon) = 0 for\ all\ i$
- Homoscedasticity: The variance of the error term is constant across observations.
  - $Var(\epsilon | x1, x2, \ldots, xk) = \sigma 2$
- No Autocorrelation: The error terms are not correlated with each other.
  - $Cov(\epsilon i, \epsilon j) = 0 for\ i = j$
- Normality: The error terms are normally distributed (for valid hypothesis testing).
  - ϵ~N(0,σ2)
- No Perfect Multicollinearity: There is no perfect linear relationship between variables.

II. Initial Computer Run

Initial Regression Output with Standard Errors in Parenthesis:

$$ln(SPRICE) = 3.237(0.525) + 0.145(0.042)ln(NROOMS) + 0.292(0.038)ln(LVAREA)$$
$$- 0.092(0.010)ln(HAGEEFF) + 0.107(0.011)ln(LSIZE) - 0.586(0.040)ln(PTAXES)$$
$$+ 0.579(0.027)ln(MEDINC) - 0.219(0.016)ln(DFCL) + 0.208(0.045)ln(SSPEND)$$
$$- 0.153(0.019)ln(MSPEND)$$

III. Interpretation of R^2

Our R^2 value of 0.4914 shows that approximately 49.14% of the variability in home values can be explained by the independent variables in the model. This suggests that the model has moderate explanatory power.

IV. Test of Overall Significance of Regression Equation

- Null Hypothesis (H0): All regression coefficients are zero ($\beta 1 = \beta 2 = \ldots = \beta 9 = 0$); none of the independent variables have a significant relationship with ln(SPRICE).
- Alternative Hypothesis (H1): At least one regression coefficient is not zero ($\beta j \neq 0$).
- F-statistic: The reported value is $F(9, 1990) = 215.57$
- Significance Level: Use $\alpha = 0.05$
- P-value: $Prob > F = 0.0000$, which is less than 0.05.
- Decision: Reject the null hypothesis.
- Conclusion: The model is statistically significant overall, meaning that at least one independent variable has a significant impact on ln(SPRICE).

V. Test HAGEEFF Significance

- Null Hypothesis (H0): $\beta ln(HAGEEFF) = 0$, meaning the variable has no effect.
- Alternative Hypothesis (H1): $\beta ln(HAGEEFF) \neq 0$, meaning the variable has an effect.
- t-Statistic: The reported value is $t =- 9.11$.
- P-value: The reported value is $P > |t| = 0.000$, which is less than 0.05.
- Decision: Reject the null hypothesis.
- Conclusion: ln(HAGEEFF) is statistically significant at the 5% level and hurts ln(SPRICE), as indicated by the negative coefficient of (-0.09175).

VI. Dropping Insignificant Variables

All of the variables have p-values of less than 0.05, indicating that they are all statistically significant at the 5% level. There are no insignificant variables to drop.

VII. Performing Subset Test

As all the variables in the original regression are significant at the 0.05 level, the subset test is unnecessary due to no variables being dropped due to insignificance.

VIII. Final Regression Equation

No variables were dropped; the final regression equation remains the same as the initial equation.

Initial Regression Output With Standard Errors in Parenthesis:

$$ln(SPRICE) = 3.237(0.525) + 0.145(0.042)ln(NROOMS) + 0.292(0.038)ln(LVAREA) - 0.092(0.010)ln(HAGEEFF) + 0.107(0.011)ln(LSIZE) - 0.586(0.040)ln(PTAXES) + 0.579(0.027)ln(MEDINC) - 0.219(0.016)ln(DFCL) + 0.208(0.045)ln(SSPEND) - 0.153(0.019)ln(MSPEND)$$

Key Statistics:

- R^2 = 0.4914 | adjusted R^2 = 0.4937 | F-statistic = 215.57 | Root MSE = 0.24728

Section III: Interpretation of Key Results & Confidence Intervals

I. Interpreting Most Highly Significant Estimated Regression Coefficients

ln_MEDINC (Median Income)

- coefficient = 0.5794 | t-value = 21.58 | p-value = 0.000
- A 1% increase in the median income is associated with an increase of approximately 0.5794% in the sale price of the property (SPRICE). This suggests a positive relationship between median income and property values; higher median income tends to raise the value of the properties. This aligns with the expectation that more affluent areas often have higher property values.

ln_LVAREA (Living Area)

- coefficient = 0.2921 | t-value = 7.74 | p-value = 0.000
- A 1% increase in the living area of a property is associated with an increase of approximately 0.2921% in the sale price. This positive relationship shows that larger homes, in terms of square footage and living area, are more valuable. This is logical, as buyers typically pay more for more space in real estate markets.

ln_PTAXES (Property Taxes)

- coefficient = -0.5860 | t-value = -14.64 | p-value = 0.000
- A 1% increase in property taxes is associated with an approximate 0.5860% decrease in the sale price of a property; this negative relationship shows that higher property taxes tend to lower property values. Buyers deterred by high taxes lead to a decrease in demand, meaning a lower sale price.

II. Constructing and Interpreting Confidence Intervals

$$Confidence\ Interval = Coefficient \pm (tcritical \times Standard\ Error)$$

- t-critical value = critical value for a 95% confidence level (1.96 due to large sample size)
- the standard error for each coefficient provided in the regression output

Interval of ln_MEDINC (Median Income)

- coefficient = 0.5794 | standard error = 0.0268
- $Confidence\ Interval\ =\ 0.5794 \pm (1.96 \times 0.0268)\ =\ 0.5794 \pm 0.0525$
    - $Confidence\ Interval\ =\ [0.5269, 0.6320]$
- We are 95% confident that the true coefficient for median income is between 0.5269 and 0.6320. For each 1% increase in median income, the sale price of the property increases by an amount between 0.5269% and 0.6320%. This positive relationship remains strong when in this range.

Interval of ln_LVAREA (Living Area)

- coefficient = 0.2921 | standard error = 0.0377
- $Confidence\ Interval\ =\ 0.2921 \pm (1.96 \times 0.0377)\ =\ 0.2921 \pm 0.0739$
    - $Confidence\ Interval\ =\ [0.2182, 0.3660]$
- We are 95% confident that the true coefficient for living area lies between 0.2182 and 0.3660. For each 1% increase in living area, the sale price of the property will increase by an amount between 0.2182% and 0.3660%. The relationship between living area and property price is both positive and significant.

Section IV: Conclusion

Through this econometric model, we identified several key determinants of home values in the Chicago metropolitan area. The model incorporated variables related to the physical characteristics, as well as economic and locational factors of the homes. Our results highlight the most significant variables impacting home prices which are median income, living area, and property tax rates. The findings showed that higher median income and larger living areas positively correlate with higher property values, while increased property taxes negatively impact home prices.

The original and final models are identical, as all variables in the initial regression were statistically significant at the 5% level. Although this model shows insightful relationships, there are limitations, such as the potential for omitted variables, margins of measurement error, as well as the assumption of linearity. However, the explanatory power of R^2 being 0.4914 shows a strong fit for the data, explaining nearly half of the variability in home prices.

The strengths of this model lie in its inclusion of a variety of factors that align with theoretical expectations, such as the influence of living area size and property taxes on home values. The use of logarithmic transformations allows for intuitive interpretations of elasticity, as well as our R^2 value indicating a good fit. However, the model could have omitted variable bias if other significant factors were not included, as well as the assumption of linearity and homoscedasticity not fully capturing the dynamics of the real estate market. This model also suffers from the potential of multicollinearity among the explanatory variables, which could affect the reliability of coefficient estimates.

Overall, our findings provide valuable insights for real estate investment and urban planning. To enhance property values, strategies should focus on increasing the square footage of homes, improving local income levels, as well as reevaluating property tax policies. However, this report should be interpreted with caution, as additional factors could also influence home prices not captured in this model. These findings can guide strategic investment decisions and urban development policies alike, ultimately contributing to more informed and effective real estate management.

References

Chattopadhyay, Sudip. "Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market." *Land Economics*, vol. 75, no. 1, 1999, pp. 22–38. *JSTOR*, https://doi.org/10.2307/3146991. Accessed 2024.

Appendix

```
. *(17 variables, 2000 observations pasted into data edit
> or)
. gen ln_SPRICE = ln(SPRICE)

.
. gen ln_NROOMS = ln(NROOMS)

.
. gen ln_LVAREA = ln(LVAREA)

.
. gen ln_HAGEEFF = ln(HAGEEFF)

.
. gen ln_LSIZE = ln(LSIZE)

.
. gen ln_PTAXES = ln(PTAXES)

.
. gen ln_MEDINC = ln(MEDINC)

.
. gen ln_DFCL = ln(DFCL)

.
. gen ln_SSPEND = ln(SSPEND)

.
. gen ln_MSPEND = ln(MSPEND)
. regress ln_SPRICE ln_NROOMS ln_LVAREA ln_HAGEEFF ln_LSI
> ZE ln_PTAXES ln_MEDINC ln_DFCL ln_SSPEND ln_MSPEND

      Source |       SS           df       MS      Number
>  of obs   =     2,000
-------------+----------------------------------   F(9, 1
> 990)      =    215.57
       Model |  118.634916         9  13.1816573   Prob >
>  F        =    0.0000
    Residual |  121.684092     1,990  .061147785   R-squa
> red       =    0.4937
-------------+----------------------------------   Adj R-
> squared   =    0.4914
       Total |  240.319008     1,999  .120219614   Root M
> SE        =    .24728

-------------+--------------------------------------------
> -------------------
   ln_SPRICE | Coefficient  Std. err.      t    P>|t|
>  [95% con
>         f. interval]
-------------+--------------------------------------------
> -------------------
   ln_NROOMS |   .1445533   .0418017     3.46   0.001
>  .0625736
>              .226533
```

```
   ln_LVAREA |   .2920687   .0377187     7.74   0.000
>  .2180964
>               .3660411
  ln_HAGEEFF |  -.0917527    .010076    -9.11   0.000
> -.1115133
>              -.071992
    ln_LSIZE |   .1067204   .0113157     9.43   0.000
>  .0845285
>               .1289124
   ln_PTAXES |  -.5859227   .0400284   -14.64   0.000
> -.6644247
>              -.5074208
   ln_MEDINC |   .5793913   .0268488    21.58   0.000
>  .5267366
>               .6320461
     ln_DFCL |  -.2186553   .0163153   -13.40   0.000
> -.2506521
>              -.1866585
   ln_SSPEND |   .2078268   .0446858     4.65   0.000
>  .1201908
>               .2954627
   ln_MSPEND |  -.1531695   .0191963    -7.98   0.000
> -.1908165
>              -.1155226
       _cons |   3.236998   .5249566     6.17   0.000
>  2.207476
>               4.266521
-------------+----------------------------------------
> ---------------------
```