David C. Lincoln Fellowships: Land Valuation Methods
A Matching Method for Land Valuation

Jeffrey Zabel
Economics Department
Tufts University
January 28, 2021

Introduction

- One approach to land valuation, particularly used by accessors, is to base estimates on comparable properties that recently sold.

Introduction

- One approach to land valuation, particularly used by accessors, is to base estimates on comparable properties that recently sold.
- The idea is to find recently transacted properties that are close matches to the target unit and use their transaction prices to predict the market price of the target unit.

Introduction

- One approach to land valuation, particularly used by accessors, is to base estimates on comparable properties that recently sold.
- The idea is to find recently transacted properties that are close matches to the target unit and use their transaction prices to predict the market price of the target unit.
- In this study, an analytical procedure for choosing comparables is developed that is based on the Mahalanobis distance measure.

Introduction

- One approach to land valuation, particularly used by accessors, is to base estimates on comparable properties that recently sold.
- The idea is to find recently transacted properties that are close matches to the target unit and use their transaction prices to predict the market price of the target unit.
- In this study, an analytical procedure for choosing comparables is developed that is based on the Mahalanobis distance measure.
- A hedonic regression using these comparables is then run to predict land value as well as market price.

Introduction

- This method is applied using transaction data for single family residential properties from Maricopa County from 2007-2018.

Introduction

- This method is applied using transaction data for single family residential properties from Maricopa County from 2007-2018.
- This was a period including the great recession and its recovery that was particularly impactful in Maricopa County which suffered from significant overbuilding.

Introduction

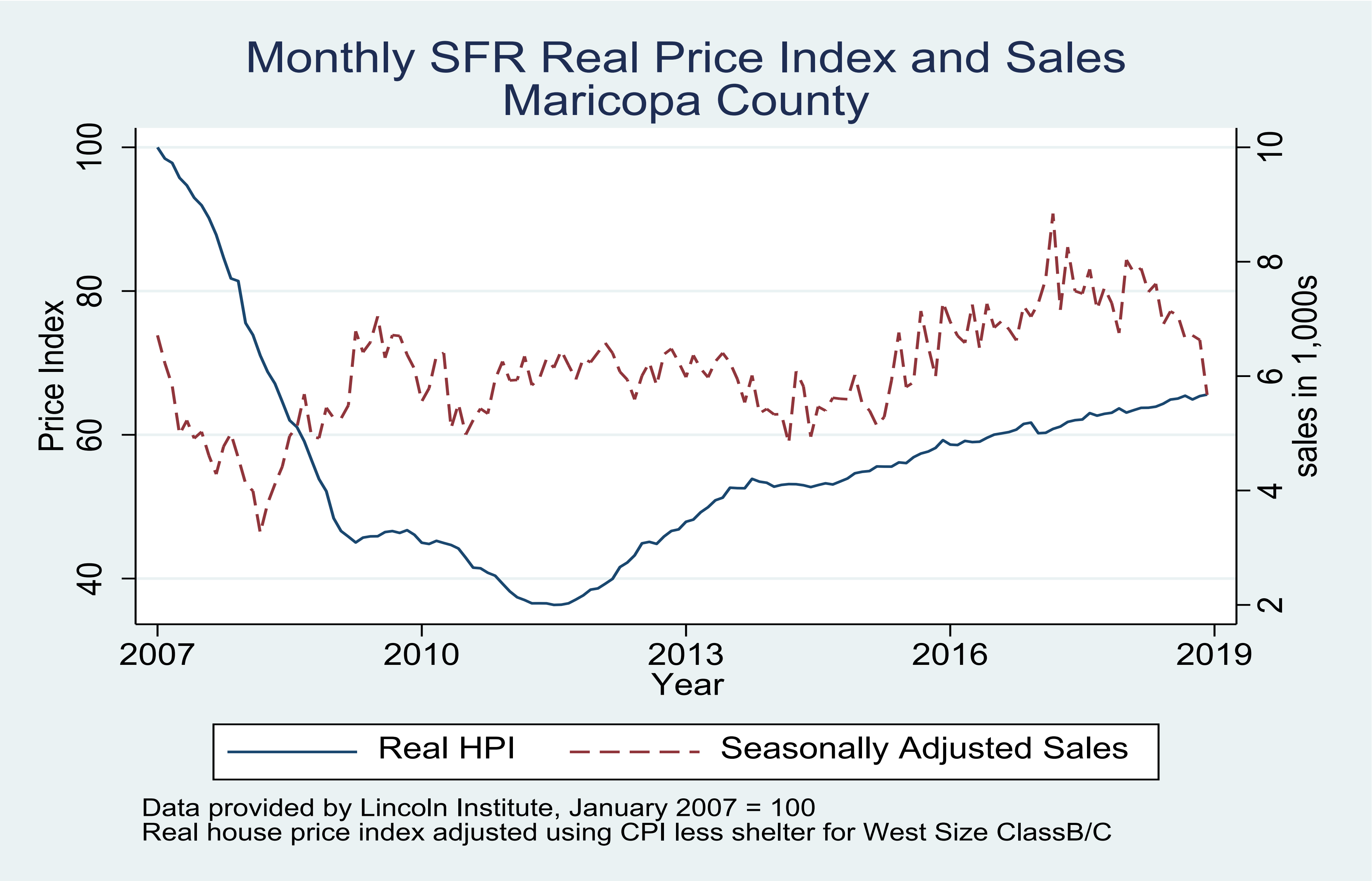
- This method is applied using transaction data for single family residential properties from Maricopa County from 2007-2018.
- This was a period including the great recession and its recovery that was particularly impactful in Maricopa County which suffered from significant overbuilding.
- Prices bottomed out in July 2011 at 36% of its January 2007 value.

Introduction

- This method is applied using transaction data for single family residential properties from Maricopa County from 2007-2018.
- This was a period including the great recession and its recovery that was particularly impactful in Maricopa County which suffered from significant overbuilding.
- Prices bottomed out in July 2011 at 36% of its January 2007 value.
- It then experienced a steady increase and stood at 66% of its original value in December 2018.

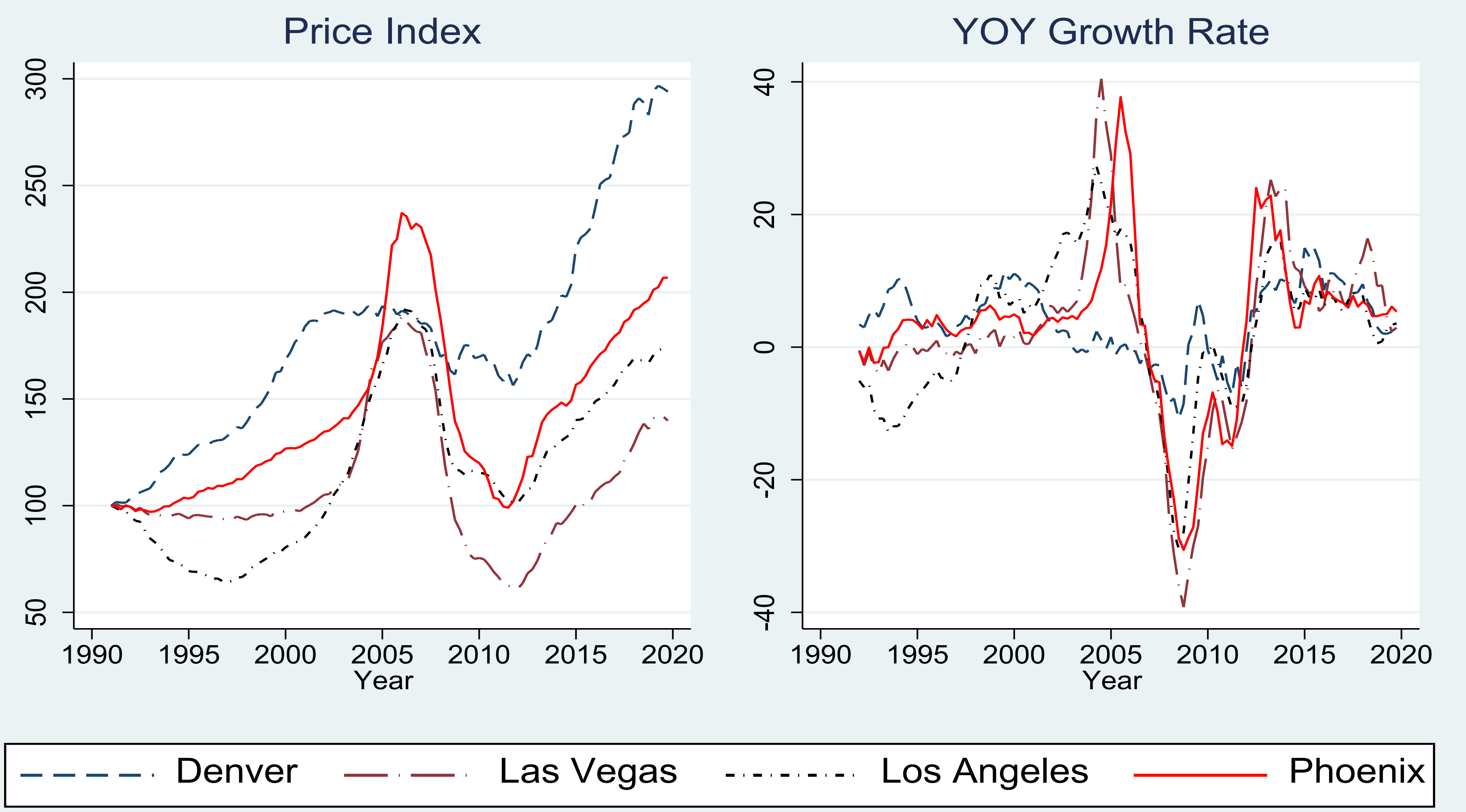
Introduction

- This method is applied using transaction data for single family residential properties from Maricopa County from 2007-2018.
- This was a period including the great recession and its recovery that was particularly impactful in Maricopa County which suffered from significant overbuilding.
- Prices bottomed out in July 2011 at 36% of its January 2007 value.
- It then experienced a steady increase and stood at 66% of its original value in December 2018.
- This significant volatility will challenge the accuracy of land value prediction.

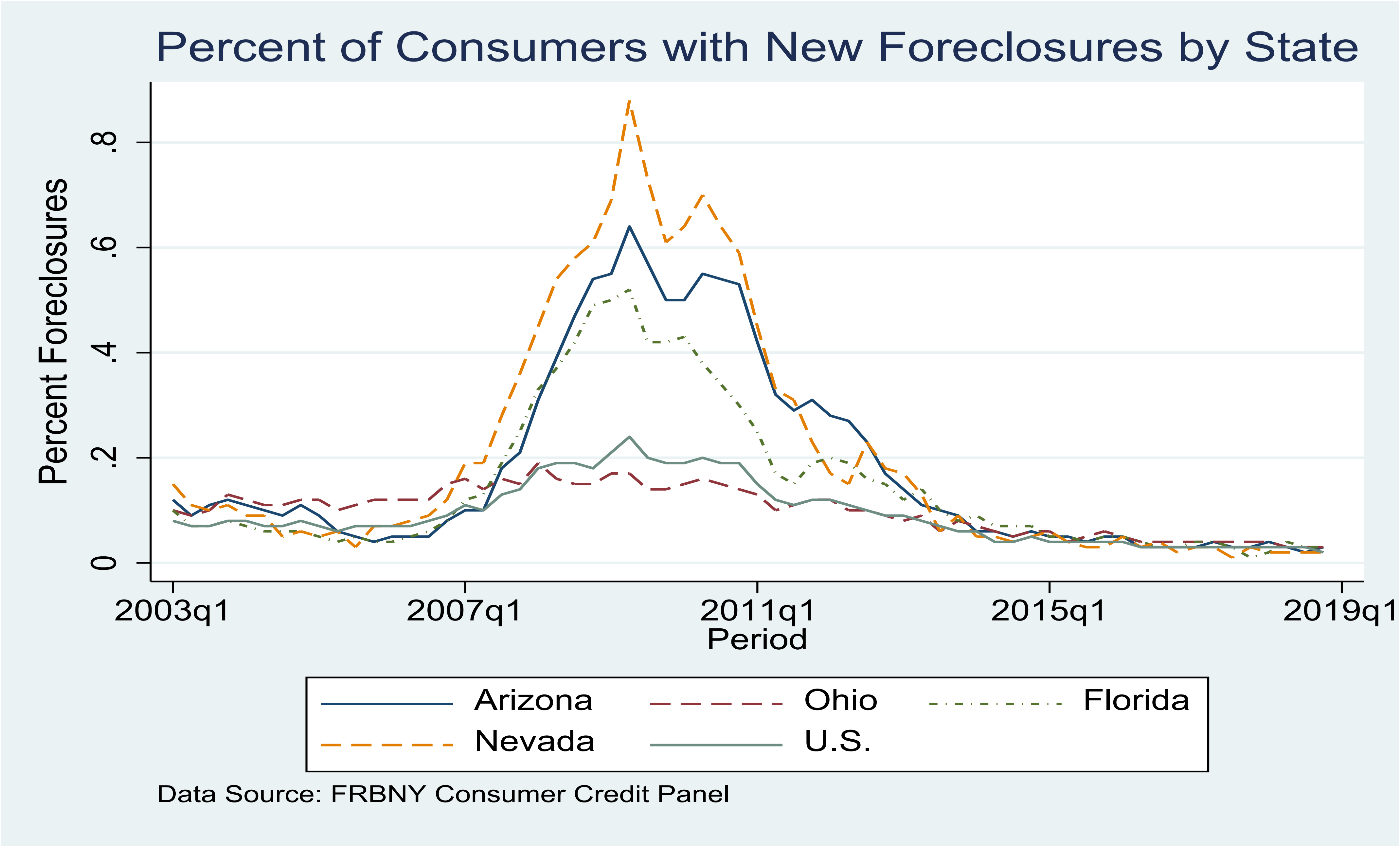


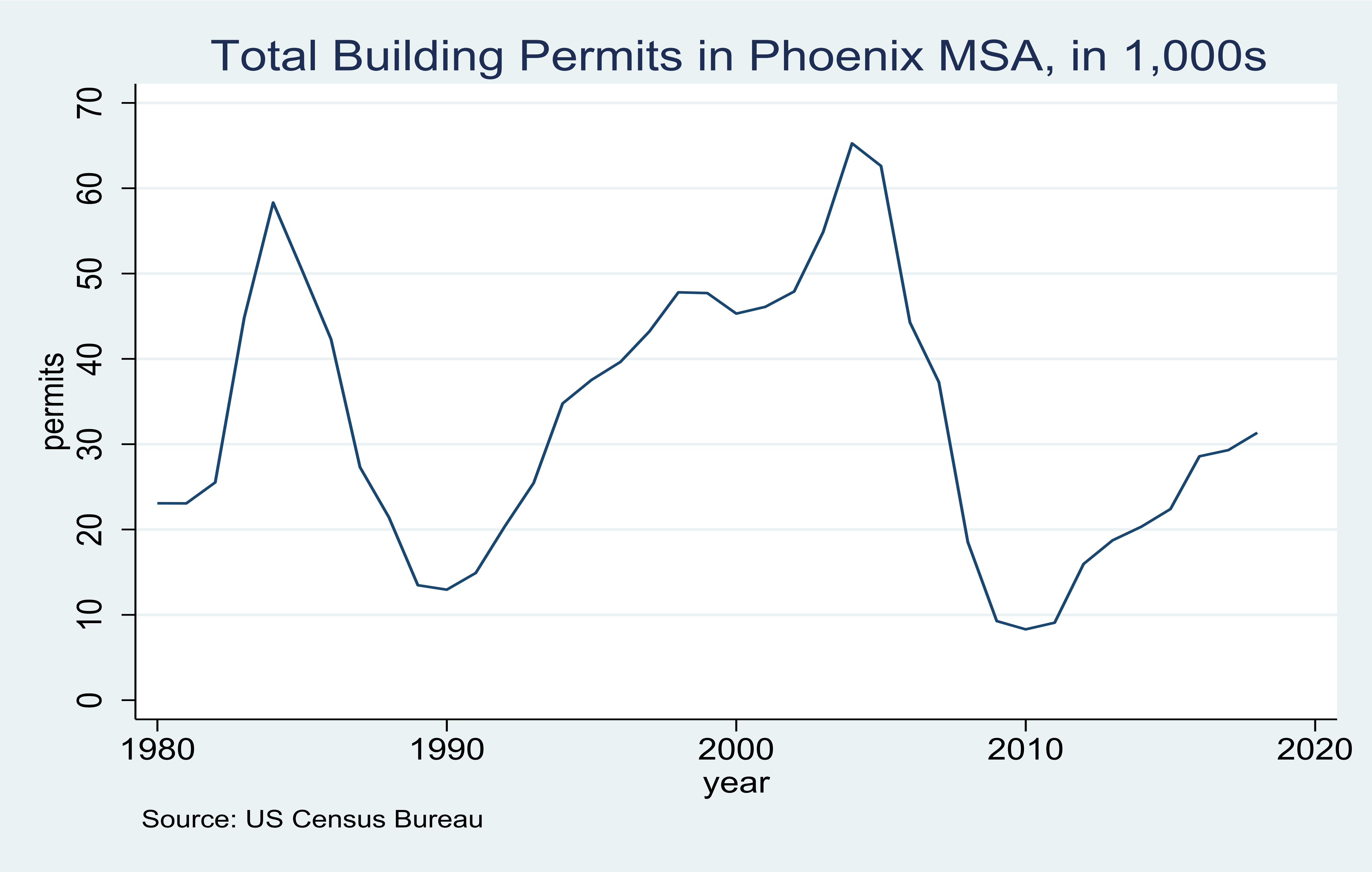
Monthly SFR Real Price Index and Sales
Maricopa County
Price Index
100
80
60
40
sales in 1,000s
10
8
6
4
2
2007
2010
2013
2016
2019
Year
Real HPI
Seasonally Adjusted Sales
Data provided by Lincoln Institute, January 2007 = 100
Real house price index adjusted using CPI less shelter for West Size ClassB/C

Quarterly SFR Real Price Index and YOY Growth Rate



Data from FHFA, for left graph: 1991:1 = 100
Real house price index adjusted using CPI less shelter for West


Percent of Consumers with New Foreclosures by State
Percent Foreclosures
0
.2
.4
.6
.8
2003q1
2007q1
2011q1
2015q1
2019q1
Period
Arizona
Ohio
Florida
Nevada
U.S.
Data Source: FRBNY Consumer Credit Panel


Total Building Permits in Phoenix MSA, in 1,000s
permits
0
10
20
30
40
50
60
70
1980
1990
2000
2010
2020
year
Source: US Census Bureau

Sample Size Restrictions

Description	Sample Size
Initial	**1,218,354**
Limit deed type	895,522
- Drop observations with missing values for age, lot size, and unit size. - Trim top and bottom 1% of price:	877,106
Limit to Residential Zones as classified by the Maricopa County Assessor’s office	603,409
- Exclude “economic zones” that include more than one unit - Limit to the 99% of units that are rated average, good, or very good. - Exclude the few transactions that were recorded as “Barter or Trade” and the very few transactions were the seller is selling a partial interest in a property to the buyer are excluded. - Limit to the 99.5% of sales that are in residential zones with minimum lot size zoning between 4,000 square feet and an acre.	**577,016**

Jurisdictions/Markets

- Maricopa County is divided into 27 “markets” and each market is divided into “neighborhoods”.
- There are 354 neighborhoods: 1 to 44 in each market.

Jurisdictions/Markets

- Maricopa County is divided into 27 “markets” and each market is divided into “neighborhoods”.
- There are 354 neighborhoods: 1 to 44 in each market.
- Market and neighborhood boundaries were created by the Maricopa County Assessor’s office.
  - “A market area is generally a broader grouping of properties and tends to relate to how buyers would view the local real estate market. In some cases, they are based on a jurisdiction.”

Jurisdictions/Markets

- Maricopa County is divided into 27 “markets” and each market is divided into “neighborhoods”.
- There are 354 neighborhoods: 1 to 44 in each market.
- Market and neighborhood boundaries were created by the Maricopa County Assessor’s office.
    - “A market area is generally a broader grouping of properties and tends to relate to how buyers would view the local real estate market. In some cases, they are based on a jurisdiction.”
    - Whereas for neighborhoods “generally the features of homes and the real estate market trends within a neighborhood are considered most similar, especially when compared to other neighborhoods.”

Jurisdictions/Markets

- The market is assumed to be the relevant area for identifying comparables.
- One can think of there being a separate assessor in each market and their job is to assess property values in that jurisdiction.

Land Valuation

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}\beta_1 + f_2\left(\text{lotsize}_{ijt}\,;\beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

Price_{itj}	nominal price of SFR i in jurisdiction j at time t,
X_{it}	structural characteristics
u_j and v_t	jurisdiction and time fixed effects

f_1 and f_2 allow for nonlinear functions of Price and lot size

Land Valuation

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}\beta_1 + f_2\left(\text{lotsize}_{ijt}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

E(price) = structure value + land value

= $X_{it}\beta_2$ + (u_j + f(*lotsize*$_{ijt}$;β_2)) or

E(price) = structure value + lot value + jurisdiction/market value

= $X_{it}\beta_2$ + f(*lotsize*$_{ijt}$;β_2) + u_j.

Note that the structure value, $X_{it}\beta_2$, could be negative for teardowns.

4 Different Land Values

$$f_1\left(\mathrm{Price}_{ijt}\right) = \beta_0 + X_{it}\beta_1 + f_2\left(\mathrm{lotsize}_{ijt};\beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

1. Price_{ijt}: the value of the land and the structure in market j.

4 Different Land Values

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}\beta_1 + f_2\left(\text{lotsize}_{ijt}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

1. Price_{ijt}: the value of the land and the structure in market j.
2. u_j : the “entry” price or market price – this could be measured at the market or neighborhood.
   - It captures the value of all local public goods related to location j.
   - It DOES NOT include amenities available to all residents in the lager housing market such as climate.

4 Different Land Values

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}\beta_1 + f_2\left(\text{lotsize}_{ijt}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

1. Price_{ijt}: the value of the land and the structure in market j.
2. u_j : the "entry" price or market price – this could be measured at the market or neighborhood.
   - It captures the value of all local public goods related to location j.
   - It DOES NOT include amenities available to all residents in the lager housing market such as climate.
3. (lotsize_{it}; β_2): the value of the lot or the price per unit of lot: $f_2(\text{lotsize}_{ijt};\beta_2)$/ lotsize_{ijt}.

4 Different Land Values

$$\mathrm{f}_1\left(\mathrm{Price}_{\mathrm{ijt}}\right) = \beta_0 + \mathrm{X}_{\mathrm{it}}\beta_1 + \mathrm{f}_2\left(\mathrm{lotsize}_{\mathrm{ijt}};\beta_2\right) + \mathrm{u}_{\mathrm{j}} + \mathrm{v}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}}$$

1. Price_{ijt}: the value of the land and the structure in market j.
2. u_j : the “entry” price or market price – this could be measured at the market or neighborhood.
   - It captures the value of all local public goods related to location j.
   - It DOES NOT include amenities available to all residents in the lager housing market such as climate.
3. $(\mathrm{lotsize}_{it}; \beta_2)$: the value of the lot or the price per unit of lot: $f_2(\mathrm{lotsize}_{ijt};\beta_2)/\ \mathrm{lotsize}_{ijt}$
4. $u_j + f_2(\mathrm{lotsize}_{ijt}; \beta_2)$: the value of the land for property i in market j.

Setup and Prediction Procedure for Total Property Valuation

- Assume a subset of properties U have transacted in a housing market (here Maricopa County) in a sub-period T (say 1 or 2 years).

Setup and Prediction Procedure for Total Property Valuation

- Assume a subset of properties U have transacted in a housing market (here Maricopa County) in a sub-period T (say 1 or 2 years).
- The goal is to predict the value of a single target property, Y (not in U) in a jurisdiction in the housing market.

Setup and Prediction Procedure for Total Property Valuation

- Assume a subset of properties U have transacted in a housing market (here Maricopa County) in a sub-period T (say 1 or 2 years).
- The goal is to predict the value of a single target property, Y (not in U) in a jurisdiction in the housing market.
- Suppose that Y has values of $X = X_Y$.

Setup and Prediction Procedure for Total Property Valuation

- Assume a subset of properties *U* have transacted in a housing market (here Maricopa County) in a sub-period T (say 1 or 2 years).
- The goal is to predict the value of a single target property, *Y* (not in *U*) in a jurisdiction in the housing market.
- Suppose that *Y* has values of $X = X_Y$.
- A subsample of comparable properties in *U* will be identified that are "similar" to *Y* to estimate the market value of *Y*. These comparables will have values of *X* that are "similar" to X_Y (discussed below).

Setup and Prediction Procedure for Total Property Valuation

- This is done separately for (a random sample of) transactions in each market since the approach requires running a separate regression based on comparable transactions in the previous two years in the same market to get the predicted (assessed) values.

Setup and Prediction Procedure for Total Property Valuation

- This is done separately for (a random sample of) transactions in each market since the approach requires running a separate regression based on comparable transactions in the previous two years in the same market to get the predicted (assessed) values.
- There are three stages:
  1. Choosing variables and estimating weights for choosing comparables
  2. Choosing comparables (matching) and
  3. Prediction.

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{ijt}\right)^{s} = \beta_0 + X_i^s \beta_1 + f\left(\mathrm{lotsize}_{it}^s; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

- There are many structure characteristics in the data.
- These need to be reduced to a much smaller number:

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{\mathrm{ijt}}\right)^{\mathrm{s}} = \beta_0 + \mathrm{X}_{\mathrm{i}}^{\mathrm{s}}\beta_1 + \mathrm{f}\left(\mathrm{lotsize}_{\mathrm{it}}^{\mathrm{s}};\beta_2\right) + \mathrm{u}_{\mathrm{j}} + \mathrm{v}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}}$$

- There are many structure characteristics in the data.
- These need to be reduced to a much smaller number:
    1. Practical: we want this to be a reasonable number for application

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{\mathrm{ijt}}\right)^{\mathrm{s}} = \beta_0 + \mathrm{X}_{\mathrm{i}}^{\mathrm{s}}\beta_1 + \mathrm{f}\left(\mathrm{lotsize}_{\mathrm{it}}^{\mathrm{s}};\beta_2\right) + \mathrm{u}_{\mathrm{j}} + \mathrm{v}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}}$$

- There are many structure characteristics in the data.
- These need to be reduced to a much smaller number:
  1. Practical: we want this to be a reasonable number for application
  2. Technical: OLS tends to overfit when there are many explanatory variables. So, it is important to reduce the number of characteristics to improve out-of-sample prediction.
- Superscript "s" indicates the variable has been standardized so that the estimated coefficients are comparable.

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\text{Price}_{ijt}\right)^{s} = \beta_0 + X_i^s \beta_1 + f\left(\text{lotsize}_{it}^{s}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{ijt}\right)^{s} = \beta_0 + \mathrm{X}_i^s \beta_1 + \mathrm{f}\left(\mathrm{lotsize}_{it}^{s}; \beta_2\right) + \mathrm{u}_j + \mathrm{v}_t + \varepsilon_{ijt}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).
- Estimate above equation using the essential characteristics and month by year fixed effects.

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{\mathrm{ijt}}\right)^{\mathrm{s}} = \beta_0 + \mathrm{X}_{\mathrm{i}}^{\mathrm{s}}\beta_1 + \mathrm{f}\left(\mathrm{lotsize}_{\mathrm{it}}^{\mathrm{s}};\beta_2\right) + \mathrm{u}_{\mathrm{j}} + \mathrm{v}_{\mathrm{t}} + \varepsilon_{\mathrm{ijt}}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).
- Estimate above equation using the essential characteristics and month by year fixed effects.
- Regress the residual from this regression on the remaining variables.

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\text{Price}_{ijt}\right)^{s} = \beta_0 + X_i^s \beta_1 + f\left(\text{lotsize}_{it}^{s}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).
- Estimate above equation using the essential characteristics and month by year fixed effects.
- Regress the residual from this regression on the remaining variables.
- To reduce the number, I use **stepwise regression** with a strict cutoff of a p-value = 0.00001

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\text{Price}_{ijt}\right)^{s} = \beta_0 + X_i^s \beta_1 + f\left(\text{lotsize}_{it}^{s}; \beta_2\right) + u_j + v_t + \varepsilon_{ijt}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).
- Estimate above equation using the essential characteristics and month by year fixed effects.
- Regress the residual from this regression on the remaining variables.
- To reduce the number, I use **stepwise regression** with a strict cutoff of a p-value = 0.00001
- Then I sort the remaining variables by the absolute value of the point estimates and choose a small number with the highest absolute values greater than 0.02. Call this subset of chosen variables X^C.

Stage 1: Choosing Variables and Estimating Weights for Choosing Comparables

$$\ln\left(\mathrm{Price}_{ijt}\right)^{s} = \beta_0 + \mathrm{X}_i^s \beta_1 + \mathrm{f}\left(\mathrm{lotsize}_{it}^s ; \beta_2\right) + \mathrm{u}_j + \mathrm{v}_t + \varepsilon_{ijt}$$

- Essential characteristics: lot size, unit size, age, number of bathrooms, and the presence of a pool (since the data are from Arizona).
- Estimate above equation using the essential characteristics and month by year fixed effects.
- Regress the residual from this regression on the remaining variables.
- To reduce the number, I use **stepwise regression** with a strict cutoff of a p-value = 0.00001
- Then I sort the remaining variables by the absolute value of the point estimates and choose a small number with the highest absolute values greater than 0.02. Call this subset of chosen variables X^C.
- The cutoff p-value of 0.00001 and 0.02 are arbitrary and robustness checks need to be run to see how sensitive these choices are.

Stage 2: Matching

$$D(Y, k|w) = \sum_{x \in X^C} w_x \left(x_k^S - x_Y^S\right)^2$$

- Use the Mahalanobis distance metric to find a subset of transacted properties that are close matches to the target unit based on the characteristics X^C.

Stage 2: Matching

$$D(Y,k|w) = \sum_{x \in X^C} w_x \left(x_k^S - x_Y^S \right)^2$$

- Use the Mahalanobis distance metric to find a subset of transacted properties that are close matches to the target unit based on the characteristics X^C.
- These differences are weighted by the importance of each variable in predicting price from Stage 1, w_x.

Stage 2: Matching

$$D\left(Y, k \middle| w\right) = \sum_{x \in X^C} w_x \left(x_k^S - x_Y^S\right)^2$$

- Use the Mahalanobis distance metric to find a subset of transacted properties that are close matches to the target unit based on the characteristics X^C.
- These differences are weighted by the importance of each variable in predicting price from Stage 1, w_x.
- The sample of potential comparables will be limited to sales in the previous two years from the time of sale.

Stage 2: Matching

$$D(Y, k|w) = \sum_{x \in X^C} w_x \left(x_k^S - x_Y^S\right)^2$$

- Use the Mahalanobis distance metric to find a subset of transacted properties that are close matches to the target unit based on the characteristics X^C.
- These differences are weighted by the importance of each variable in predicting price from Stage 1, w_x.
- The sample of potential comparables will be limited to sales in the previous two years from the time of sale.
- The properties that qualify as comparables for *Y* will also be limited to the market where *Y* is located.

Stage 2: Matching

$$D(Y, k|w) = \sum_{x \in X^C} w_x \left(x_k^S - x_Y^S\right)^2$$

- Use the Mahalanobis distance metric to find a subset of transacted properties that are close matches to the target unit based on the characteristics X^C.
- These differences are weighted by the importance of each variable in predicting price from Stage 1, w_x.
- The sample of potential comparables will be limited to sales in the previous two years from the time of sale.
- The properties that qualify as comparables for *Y* will also be limited to the market where *Y* is located.
- Call this set of comparables C_Y.

Stage 3: Prediction using Hedonic Estimation

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}^{C}\beta_1 + f_2\left(\text{lotsize}_{it};\beta_2\right) + u_j + v_t + \varepsilon_{ijt}, \qquad t \in T,\ i \in C_Y\left(p\right)$$

- Separate regressions will be run for each transaction

Stage 3: Prediction using Hedonic Estimation

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}^{C}\beta_1 + f_2\left(\text{lotsize}_{it};\beta_2\right) + u_j + v_t + \varepsilon_{ijt}, \qquad t \in T,\ i \in C_Y\left(p\right)$$

- Separate regressions will be run for each transaction
- Regression is weighed by time of sale

Stage 3: Prediction using Hedonic Estimation

$$f_1\left(\text{Price}_{ijt}\right) = \beta_0 + X_{it}^{C}\beta_1 + f_2\left(\text{lotsize}_{it};\beta_2\right) + u_j + v_t + \varepsilon_{ijt}, \qquad t \in T,\ i \in C_Y\left(p\right)$$

- Separate regressions will be run for each transaction
- Regression is weighed by time of sale
- Linear and log price will be used to see which one produces more accurate predictions of land value
  - ln(price): distribution closer to a normal distribution and this regression will likely fit the data better.
  - But the disadvantage is that estimates need to be translated into levels since the assessed value is in levels.

Stage 3: Prediction using Hedonic Estimation

Given that f_1(Price) is linear, the predicted value for *Y* is

$$\hat{\mathrm{P}}\mathrm{rice}_{\mathrm{Yjt}} = \hat{\alpha}_0 + \mathrm{X}_{\mathrm{Y}}^{\mathrm{C}}\hat{\alpha}_1 + \mathrm{f}\left(\mathrm{lotsize}_{\mathrm{it}};\hat{\alpha}_2\right) + \hat{\mathrm{u}}_{\mathrm{j}} + \hat{\mathrm{v}}_{\mathrm{t}}$$

Given that f_1(Price) is logarithmic, the predicted value for *Y* is

$$\hat{\mathrm{P}}\mathrm{rice}_{\mathrm{Yjt}}^{\mathrm{ln}} = \mathrm{a}\cdot\exp\left(\hat{\alpha}_0 + \mathrm{X}_{\mathrm{Y}}^{\mathrm{C}}\hat{\alpha}_1 + \mathrm{f}\left(\mathrm{lotsize}_{\mathrm{it}};\hat{\alpha}_2\right) + \hat{\mathrm{u}}_{\mathrm{j}} + \hat{\mathrm{u}}_{\mathrm{j}} + \hat{\mathrm{v}}_{\mathrm{j}}\right)$$

where $\mathrm{a} = \mathrm{n}_{\mathrm{C}_{\mathrm{Y}}(\mathrm{p})}^{-1}\sum_{\mathrm{i}\in\mathrm{C}_{\mathrm{Y}}(\mathrm{p})}\exp\left(\hat{\varepsilon}_{\mathrm{ijt}}\right)$

Stage 3: Prediction using Hedonic Estimation

- Two subsamples of the set of comparables C_Y will be used
  - the closest 10% and 25% of units.
- In each case, the sample will be trimmed to exclude the top and bottom 1%.
- The sample size must be at least 100 to run the regression and carry out the prediction analysis.
- Will compare to results using the full sample

Stage 3: Prediction using Hedonic Estimation

- The predicted values will be compared to actual prices to determine the prediction accuracy of the different approaches

$$100 \cdot \frac{\mathrm{abs}\left(\hat{\mathrm{P}}\mathrm{rice}_{\mathrm{Yjt}} - \mathrm{Price}_{\mathrm{ijt}}\right)}{\mathrm{Price}_{\mathrm{ijt}}} \%$$

or

$$100 \cdot \frac{\mathrm{abs}\left(\hat{\mathrm{P}}\mathrm{rice}^{\mathrm{ln}}_{\mathrm{Yjt}} - \mathrm{Price}_{\mathrm{ijt}}\right)}{\mathrm{Price}_{\mathrm{ijt}}} \%$$

Stage 3: Prediction using Hedonic Estimation

- The predicted values will be compared to actual prices to determine the prediction accuracy of the different approaches

$$100 \cdot \frac{\mathrm{abs}\left(\hat{\mathrm{P}}\mathrm{rice}_{\mathrm{Yjt}} - \mathrm{Price}_{\mathrm{ijt}}\right)}{\mathrm{Price}_{\mathrm{ijt}}}\% \quad \text{or} \quad 100 \cdot \frac{\mathrm{abs}\left(\hat{\mathrm{P}}\mathrm{rice}^{\mathrm{ln}}_{\mathrm{Yjt}} - \mathrm{Price}_{\mathrm{ijt}}\right)}{\mathrm{Price}_{\mathrm{ijt}}}\%$$

- How can we tell if this procedure produces accurate estimates of market value?

Stage 3: Prediction using Hedonic Estimation

- The predicted values will be compared to actual prices to determine the prediction accuracy of the different approaches
- $100 \cdot \frac{\text{abs}\left(\hat{\text{P}}\text{rice}_{\text{Yjt}} - \text{Price}_{\text{ijt}}\right)}{\text{Price}_{\text{ijt}}}\%$ or $100 \cdot \frac{\text{abs}\left(\hat{\text{P}}\text{rice}^{\text{ln}}_{\text{Yjt}} - \text{Price}_{\text{ijt}}\right)}{\text{Price}_{\text{ijt}}}\%$
- How can we tell if this procedure produces accurate estimates of market value?
- One criterion used by appraisers to signify the accuracy of AVMs is the rule that 70-80% of properties valued will fall within 10% of the realized sales values (Gayler et al. 2015).

Estimating local neighborhood and lot prices

- First, to estimate the price per square foot of lot, I estimate a separate hedonic model for each market

$$\ln\left(\mathrm{Price}^{\mathrm{r}}_{\mathrm{int}}\right) = \beta_0 + \mathrm{X}^{\mathrm{C}}_{\mathrm{it}}\beta_1 + \beta_2 \ln\left(\mathrm{lotsize}_{\mathrm{it}}\right) + \mathrm{u}_{\mathrm{n}} + \varepsilon_{\mathrm{int}}$$

- where u_n is the neighborhood price

- price per square foot of lot = $\beta_2 \cdot \dfrac{\overline{\mathrm{lotsize}}}{\overline{\left(\mathrm{Price}^{\mathrm{r}}\right)}}$

Literature Review

- A similar process was developed by Vandell (1991).
- Rather than using a regression to predict the assessed value of a property, a weighted average of the adjusted sales prices of the comps is used where adjustments are based on the differences in observed characteristics between the assessed property and the comparable.
- I plan on implementing this procedure as a comparison with mine.

Results – Variable Choice

- Essential characteristics: functions of lot size, unit size and age, the number of bathrooms, and the presence of a pool

Results – Variable Choice

- Essential characteristics: functions of lot size, unit size and age, the number of bathrooms, and the presence of a pool

- The additional variables chosen via stepwise regression:
  - the number of stories
  - located on a golf course, lake, or arterial road
  - whether there is storage and if so, the size of the storage space
  - whether there is an attached garage and if so, the size if the attached garage.

Results – Variable Choice

- Also use a smaller number of explanatory variables that includes
  - the levels of lot size, unit size, and age
  - bathrooms
  - the number of stories
  - located on a golf course, lake, or arterial road,
  - whether there is a pool, storage, and attached garage.

Results – Variable Choice

- Also use a smaller number of explanatory variables that includes
  - the levels of lot size, unit size, and age
  - bathrooms
  - the number of stories
  - located on a golf course, lake, or arterial road,
  - whether there is a pool, storage, and attached garage.
- The purpose is to see if a more parsimonious set of explanatory variables produces more accurate predictions of unit price.

Results: Predicting Price

- Predicted prices are generated using a:
  - ln(price) regression,
  - price regression, and
  - straight means or medians

Results: Predicting Price

- Predicted prices are generated using a:
  - ln(price) regression,
  - price regression, and
  - straight means or medians
- While the goal is to predict price, this procedure is much more accurate in predicting ln(price).
  - This could well be because the distribution of ln(price) tends to be closer to a normal distribution than is the distribution of prices
  - It has fewer extreme values and hence has a much lower kurtosis value

Results: Predicting Price

- Takeaways:

1. Taking straight means or medians of the sample results in much larger percent absolute differences.
   - So, there DOES appear to be an advantage to using the regression approach to improve the prediction accuracy of prices.

Results: Predicting Price

- Takeaways:

1. Taking straight means or medians of the sample results in much larger percent absolute differences.
   - So, there DOES appear to be an advantage to using the regression approach to improve the prediction accuracy of prices.
2. Accuracy is best for the 10% sample with the ln(price) regression and the large set of explanatory variables; 9.6%.

Results: Predicting Price

- Takeaways:

3. Using the full sample versus the 25% or 10% sample produces worse predictions.
    - This is support for using the Mahalanobis distance metric to choose comparables.

Results: Predicting Price

- Takeaways:

3. Using the full sample versus the 25% or 10% sample produces worse predictions.
    - This is support for using the Mahalanobis distance metric to choose comparables.
4. While the ln(price) regressions perform better than the price regressions, the differences are small. So, if one is interested in simplicity, using the price regression may be preferred.

Results: Predicting Price

- Regressions are rerun weighting by time of sale so that sales that are closer in time to the unit whose price is being predicted get more weight since one might expect that these prices are more accurate predictors.

Results: Predicting Price

- Regressions are rerun weighting by time of sale so that sales that are closer in time to the unit whose price is being predicted get more weight since one might expect that these prices are more accurate predictors.
- Surprisingly, there is little, if any, gain is accuracy when weighting.
  - This might be the case since the year by month dummies that account for changes in house prices over time are enough to account for the potential inaccuracy in sales further away in time.

Variation in Prediction Accuracy Over Time

Table 4: Yearly Results and Percent <=10% (In)Accuracy

	With Weights			Without Weighs		
Year	Mean	Median	Pct<=10	Mean	Median	Pct<=10
2009	18.2	13.1	38	20.6	12.6	40
2010	17.8	13.9	38	17.0	11.8	42
2011	18.0	14.0	37	16.4	11.9	40
2012	17.9	12.2	43	18.5	13.4	45
2013	15.0	10.7	47	14.7	10.1	49
2014	13.7	9.1	54	13.5	9.5	52
2015	13.4	8.3	57	11.5	8.2	56
2016	11.3	8.0	59	12.9	7.9	59
2017	11.9	7.4	61	11.0	7.1	61
2018	10.7	7.3	63	14.6	7.6	62

Results from Prediction (In)Accuracy Regression

	Full Sample	Trimmed Sample	Median Regression
Individual Characteristics			
Ln(price)	-0.705**	-0.578**	-0.741**
Ln(lot size)	0.407**	0.333**	0.392**
Ln(house size)	0.316**	0.260**	0.399**
Market Characteristics			
Ln(mean price)	0.356	0.175	0.591**
Ln(std dev price)	0.543**	0.676**	0.421*
Ln(mean lot size)	0.110	0.196*	0.172
Ln(mean house size)	-2.011**	-1.956**	-2.275**
Business and Housing Market Characteristics			
Δunemployment rate	0.127*	0.123**	0.094
Absolute value of monthly growth rate	0.085**	0.083**	0.059**
N	2,822	2,769	2,822

* $p<0.05$; ** $p<0.01$

Results: Predicting Lot Prices

- This is controlling for neighborhood, so this is the additional price for adding an extra square foot to the property size conditional on all local amenities that are captured in the neighborhood fixed effects.

Results: Predicting Lot Prices

- This is controlling for neighborhood, so this is the additional price for adding an extra square foot to the property size conditional on all local amenities that are captured in the neighborhood fixed effects.
- The mean elasticity is 0.18 with a range of 0.05 to 0.25.

Results: Predicting Lot Prices

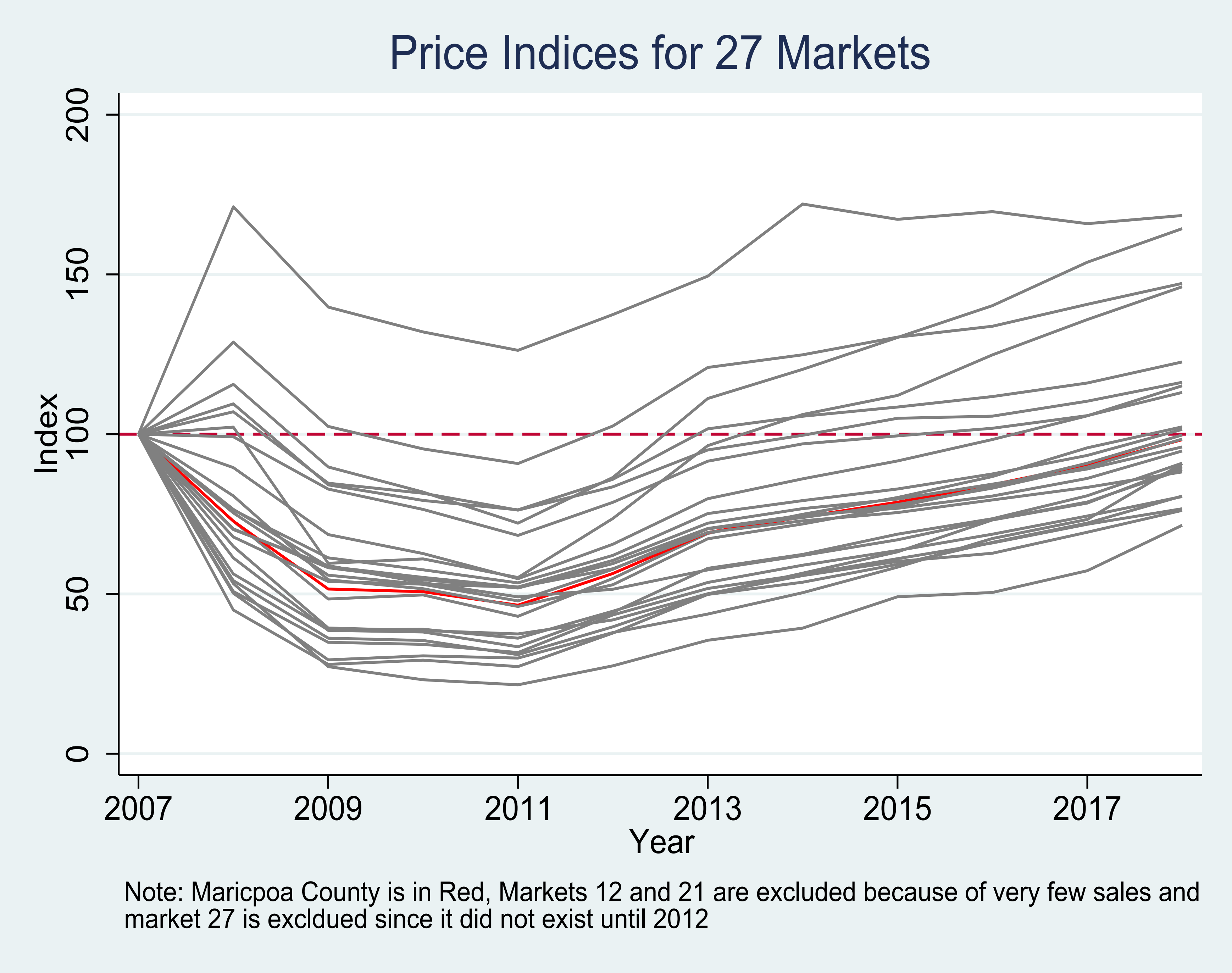
- This is controlling for neighborhood, so this is the additional price for adding an extra square foot to the property size conditional on all local amenities that are captured in the neighborhood fixed effects.
- The mean elasticity is 0.18 with a range of 0.05 to 0.25.
- The mean/std deviation of price per square foot (psf) is \$4.59/\$1.80 psf with a range of \$0.67 to \$7.60 psf. Two of the three lowest values arise because the corresponding markets have larger lot sizes with a median of around 1 acre.

Results: Market Prices

- These prices are measured in January 2007 dollars and range from \$89 thousand to \$379 thousand with a mean of \$185 thousand.

Results: Market Prices

- These prices are measured in January 2007 dollars and range from \$89 thousand to \$379 thousand with a mean of \$185 thousand.
- The coefficient of variation, the ratio of the standard deviation and the mean of a variable, is a useful measure of the variation in the neighborhood prices.
  - The value for market prices is 0.36.


Price Indices for 27 Markets
200
150
100
50
0
Index
2007
2009
2011
2013
2015
2017
Year
Note: Maricpoa County is in Red, Markets 12 and 21 are excluded because of very few sales and market 27 is excldued since it did not exist until 2012

Predicted Price/Percent for Structure, Lot, and Market

Component	mean	median	sd	min	max
Predicted Price	415.63	343.38	259.81	170.13	1409.35
Structure					
Price	154.98	107.57	167.33	14.59	770.63
Percent	37.29	31.33			
Lot					
Price	75.47	67.16	47.46	13.45	259.41
Percent	18.16	19.56			
Neighborhood					
Price	185.18	173.17	67.16	89.30	379.31
Percent	44.55	50.43			
24 of 27 markets included					

Results: Neighborhood Prices

- Look at how much variation there is in neighborhood prices within markets.

Results: Neighborhood Prices

- Look at how much variation there is in neighborhood prices within markets.
- For the 20 markets with at least 8 neighborhoods, the mean of the coefficient of variation is 0.16.

Results: Neighborhood Prices

- Look at how much variation there is in neighborhood prices within markets.
- For the 20 markets with at least 8 neighborhoods, the mean of the coefficient of variation is 0.16.
- This is a small value and indicates that there is not a lot variation in neighborhood prices in a market

Results: Neighborhood Prices

- Look at how much variation there is in neighborhood prices within markets.
- For the 20 markets with at least 8 neighborhoods, the mean of the coefficient of variation is 0.16.
- This is a small value and indicates that there is not a lot variation in neighborhood prices in a market
- Recall that the coefficient of variation for market prices is 0.36.

Results: Neighborhood Prices

- Look at how much variation there is in neighborhood prices within markets.
- For the 20 markets with at least 8 neighborhoods, the mean of the coefficient of variation is 0.16.
- This is a small value and indicates that there is not a lot variation in neighborhood prices in a market
- Recall that the coefficient of variation for market prices is 0.36.
- Hence, there is more variation in prices across markets than variation in neighborhoods within markets.

Conclusions

- Developed a new method for choosing comparables using the Mahalanobis distance measure and for generating predicted land values for the total market value, lot, and neighborhood.

Conclusions

- Developed a new method for choosing comparables using the Mahalanobis distance measure and for generating predicted land values for the total market value, lot, and neighborhood.
- This does better in terms of prediction accuracy than just taking the mean or median value of the comparables.

Conclusions

- Developed a new method for choosing comparables using the Mahalanobis distance measure and for generating predicted land values for the total market value, lot, and neighborhood.
- This does better in terms of prediction accuracy than just taking the mean or median value of the comparables.
- Prediction regressions are based on a log and linear price as the dependent variable and large (30) and small (13) sets of explanatory house structure variables and on the full set of comparables and the 10% and 25% closest comparables

Conclusions

- I find that more accurate estimates of market value occur when
    1. Using the log-price regressions versus the price regressions even though the latter need to be transformed for log-price predictions to price predictions, and

Conclusions

- I find that more accurate estimates of market value occur when
  1. Using the log-price regressions versus the price regressions even though the latter need to be transformed for log-price predictions to price predictions, and
  2. restricting the sample to the 10% closest observations particularly compared to using the full set of transactions from the same market over the past two years.

Conclusions

- I find that more accurate estimates of market value occur when
  1. Using the log-price regressions versus the price regressions even though the latter need to be transformed for log-price predictions to price predictions, and
  2. restricting the sample to the 10% closest observations particularly compared to using the full set of transactions from the same market over the past two years.
- Generally, using either the small or large set of explanatory structure variables produce similar results.

Conclusions

- If one is interested in simplicity, the estimators based on the price regression perform nearly as well as those using the ln(price) regressions.

Conclusions

- If one is interested in simplicity, the estimators based on the price regression perform nearly as well as those using the ln(price) regressions.
- This points to using the 10% or 25% sample with the smaller set of explanatory variables and the price regression.

Conclusions

- I also decompose the predicted total market value into the value of the structure, lot, and neighborhood.

Conclusions

- I also decompose the predicted total market value into the value of the structure, lot, and neighborhood.
- This approach seems to work quite well.

Conclusions

- I also decompose the predicted total market value into the value of the structure, lot, and neighborhood.
- This approach seems to work quite well.
- On average the structure, lot, and neighborhood values make up 37%, 18%, and 45% of total predicted price.

Conclusions

- How accurate are the predictions based on this procedure?

Conclusions

- How accurate are the predictions based on this procedure?
- The smallest median percent absolute value of the forecast error is 9.6%.
  - This is within the range of median values for HouseCanary which reports a range of 3.4% to 15.0% across the 50 U.S. States.
  - Zillow reports a range of median values of 3% to 25% for 666 U.S. counties.

Conclusions

- How accurate are the predictions based on this procedure?
- The smallest median percent absolute value of the forecast error is 9.6%.
  - This is within the range of median values for HouseCanary which reports a range of 3.4% to 15.0% across the 50 U.S. States.
  - Zillow reports a range of median values of 3% to 25% for 666 U.S. counties.

- This procedure reaches values above 60% in 2017 and 2018 for the percent absolute value of the forecast error within 10%,
  - HouseCanary also reports a range of values across the 50 U.S. states for the percent absolute value of the forecast error within 10% of 39.3% to 81.5%.
  - Zillow reports a range of 20% to 92% for 666 U.S. Counties (Matysiak 2017).

Conclusions

- This procedure has promise as well as relative simplicity.
- It would be useful to apply this procedure across other areas of the U.S. to see how this affects accuracy