Market Factors Affecting Housing Prices

A. Laphond, M. Troeger, Mohamed Lashuel, E. Whitehead

2024-12-07

Contents

1 Literature Review **2**

2 ACS Analysis **2**

2.1 Data Description 2

2.2 Methodology 4

3 Analysis of Specific Factors **7**

3.1 Data and Description 7

3.2 Zoning synthetic controls 10

References **13**

1 Literature Review

Kendall and Tulip (2018) quantify the effect of zoning on detached house prices in large cities in Australia. They argue that in the absence of zoning, the market for housing is competitive, as there are low barriers to entry and a large number of firms involved in development and construction. Additionally, they note that in US cities with loose zoning laws, prices are close to costs. Thus, the price of a property can be decomposed into the cost of the structure and land, and zoning representing supply distortions. With this approach, they find large price increases attributable to zoning. For example, in 2016, zoning raised the price of detached houses by 73 percent in Sydney and 69 percent in Melbourne.

2 ACS Analysis

This section used the detailed data from the American Community Survey (ACS), an annual survey conducted by the U.S. Census Bureau that captures key social, economic, housing, and demographic information across the United States[1]. The ACS dataset is a comprehensive resource for understanding the factors influencing rental housing prices. Specifically, the data includes demographic, geographic, and housing information as well as the locations from which the data was collected.

Our analysis seeks to identify the variables that significantly impact rental housing prices and quantify their effect. We focus on demographic factors such as household income, geographic indicators like metropolitan status, and structural features of houses to examine their contributions to rental price variations.

Given the self-reporting status of housing value in the dataset, we concluded that rental prices were a more accurate measure of house value. For this reason, we decided to conduct our analysis using this as our dependent variable. By applying robust statistical methods we provide a deep exploration of the determinants of housing rental prices.

2.1 Data Description

The dataset used in this study includes observations on rental housing units from the most recent ACS. Key variables of interest include:

- Monthly Rental Price: The dependent variable, representing the self-reported cost of renting a housing unit.
- Demographic Variables: These include household income, age of the householder, population density.
- Housing Characteristics: Information about the number of bedrooms, age of the property, kitchen and plumbing inclusion, and room number.
- Geographic Variables: Metropolitan versus rural location, and proximity to major urban centers.

The data was cleaned to ensure accuracy and consistency. The total number of observations was 56,835,321 with 29 features but after removing rows with missing or NA values for rental price or explanatory variables (76.5%) and selecting important features, we were left with 13,341,750 observations with 15 features.

We are not including taxincl and insincl because those columns only have zeros. Likewise, we are excluding debt_incurred since it only has FALSE. Finally, year_from_start was removed given it was perfectly correlated with year.

The dataset contains observations from counties across the United States. However, due to challenges in obtaining complete information and the presence of missing values in the FIPS column, it does not provide a

[1] https://www.census.gov/programs-surveys/acs/

clear representation of individual counties. Instead, we looked at PUMAs (Public Use Microdata Samples), which "are non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each."[2]. Figure 1 highlights the regions from which we have data.

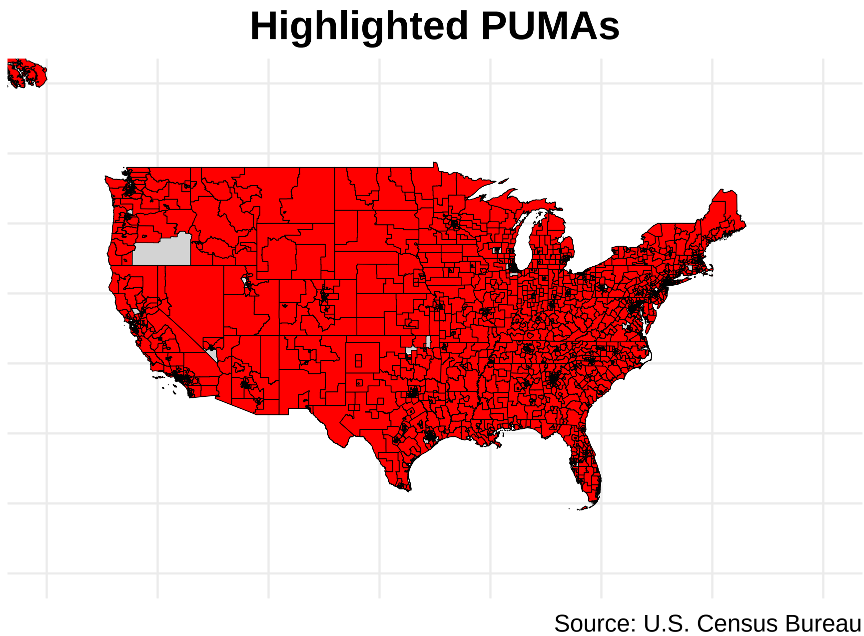

Figure 1: Regions with data present

Table 1 provides an overview of the data distribution, highlighting significant variation in rental prices, income levels, and property characteristics. Such variation underscores the need for robust multivariate analysis to disentangle the effects of different predictors.

Table 1: Summary Statistics for Numeric Variables

Variable	Mean	Variance	Min	Max
density	6761.23	1.906403e+08	2.40	132103.1
hhincome	61772.22	4.512907e+09	-11400.00	1879570.0
rooms	4.88	3.750000e+00	1.00	22.0
builtyr2	5.54	1.456000e+01	1.00	27.0
unitsstr	5.34	6.890000e+00	1.00	10.0
bedrooms	2.37	1.160000e+00	0.00	10.0
age	33.84	4.635100e+02	0.00	97.0
cpi_rent	834.11	2.216091e+05	2.65	5234.8

One variable that caught our attention was age. At first, we were surprised and confused by the presence of ages such as 0 years old or very young ages. However, after further investigation, we discovered that the ACS collects data on every person living in each household to ensure a more comprehensive and accurate measure. As a result, the age variable includes values ranging from 0 to 97.

Table 2 summarizes the factor features and provides further insight into our dataset. As indicated by the FIPS feature, we have data from every state in the country, as shown by the 51 unique values in this feature. California (represented by FIPS code 06) is the most common state in the sample dataset. Additionally, the

[2] https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html

"year" feature contains 11 levels, corresponding to 11 years of data in the sample (2012-2022), with 2014 being the most commonly represented year.

Table 2: Summary Statistics for Factor Variables

	Variable	Number.of.Levels	Most.Frequent.Level	Frequency.of.Most.Frequent.Level
year	year	11	2015	8170
puma	puma	1384	300	1424
statefip	statefip	51	6	14486

2.2 Methodology

To analyze the determinants of rental prices, we adopted a multivariate linear regression model. This method allows us to evaluate the individual contribution of each explanatory variable while controlling for the effects of others. it is important to note that our dependent variable (rent) is the gross rent adjusted by inflation. The steps taken in the analysis are as follows:

Variable Selection: Using domain knowledge and exploratory data analysis, we identified variables with potential relevance to rental price determination.

Model Specification:

The primary model is specified as:

$$Y = \beta_0 + \gamma_i * year_i + \theta_j * state_j + \beta_k * X_k + \epsilon$$

Where:

- Y represents the dependent variable (Rent price).
- $X_1, X_2, X_3, \ldots, X_n$ are the independent variables (e.g., [density, metro, hhincome, rooms, builtyr2, unitsstr, bedrooms, age, married, male, complete_plumbing, has_kitchen]).
- $year_i$ is the dummy variable for the year with $i \in \{1, \ldots, 18\}$ as we're controlling for 18 years of data.
- $state_j$ is the dummy variable for the state with $j \in \{1, \ldots, 51\}$ as we're controlling for all 50 states and the District of Columbia.
- β_0 is the intercept of the regression model.
- $\beta_1, \beta_2, \beta_3, \ldots, \beta_n$ are the coefficients that represent the relationship between each independent variable and the dependent variable.
- γ_i are the coefficients that represent the relationship between each year and the dependent variable.
- θ_j are the coefficients that represent the relationship between each state and the dependent variable.
- ϵ is the error term, capturing unobserved factors.

Note we have elected to go with a two way fixed effects model to increase the accuracy of time and regional trends.

Model Estimation: To estimate the coefficients, we employed Ordinary Least Squares (OLS) regression. All models were estimated with heteroskedasticity-robust standard errors. Diagnostic tests were conducted to validate the model assumptions:

Multicollinearity: Variance Inflation Factors (VIF) were calculated to ensure no strong multicollinearity among the predictors.

Results for this primary model are displayed as column 1 of Table **??**

The VIF scores were computed to further validate the model in Table 4

Table 3: Results of regressions

	Main specification	Time FEs only	State FEs only	No FEs
	(1)	(2)	(3)	(4)
density	0.004***	0.004***	0.005***	0.005***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
metro	174.867***	174.067***	258.244***	257.347***
	(2.975)	(2.970)	(2.943)	(2.936)
hhincome	0.003***	0.003***	0.003***	0.003***
	(0.00005)	(0.00005)	(0.0001)	(0.0001)
rooms	20.524***	20.607***	12.197***	12.295***
	(1.263)	(1.268)	(1.298)	(1.302)
builtyr2	14.255***	14.654***	11.949***	12.305***
	(0.395)	(0.395)	(0.398)	(0.398)
unitsstr	1.052*	1.106*	3.571***	3.623***
	(0.625)	(0.626)	(0.644)	(0.645)
bedrooms	71.916***	71.243***	78.721***	78.094***
	(2.272)	(2.280)	(2.366)	(2.371)
age	−0.714***	−0.662***	−0.787***	−0.739***
	(0.068)	(0.068)	(0.071)	(0.071)
married	44.374***	42.422***	56.113***	54.325***
	(3.119)	(3.123)	(3.326)	(3.327)
male	−2.108	−2.003	−3.131	−3.020
	(2.439)	(2.444)	(2.592)	(2.596)
Constant	−64.230***	−33.057***	64.204***	94.378***
	(11.053)	(10.609)	(8.498)	(7.832)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: VIF values

	GVIF	Df	GVIF^(1/(2*Df))
year	1.039258	10	1.001927
density	1.952879	1	1.397454
metro	1.328064	1	1.152417
hhincome	1.139373	1	1.067414
rooms	2.819474	1	1.679129
builtyr2	1.147481	1	1.071205
unitsstr	1.622843	1	1.273909
bedrooms	3.038772	1	1.743208
age	1.185043	1	1.088597
married	1.143369	1	1.069284
male	1.009558	1	1.004767
statefip	2.480727	50	1.009127

Before analyzing the results, we note that all estimates are statistically significant, except for the coefficient of *male*. After applying robust standard errors, each estimated coefficient remains significant. Furthermore, we computed the Variance Inflation Factor (VIF) scores to ensure the validity of the model. All VIF scores are below 5, indicating no strong multicollinearity. This confirms that our model is both reliable and robust.

Model Overview: The regression model included a total of 79 predictors and was fitted using a dataset with 13,341,750 observations. The R-squared of 0.4315 indicates that 43.15% of the variance in CPI Rent is explained by the included predictors. The F-statistic ($ 1.257e^5, p < 2.2e^{-16}$) demonstrates that the overall model is statistically significant.

Key Results:

1. Time Trends:

Yearly dummy variables were included to capture temporal trends. Statistically significant coefficients ($p < 0.001$) indicate clear yearly variation in CPI Rent. For example, CPI Rent increased significantly in 2020 compared to the base year($\hat{\gamma_{2020}} = 71.80$, $p < 2.2e^{-16}$).

2. Population Density:

The coefficient for population density was small but highly significant ($\hat{\beta_{density}} = 0.003$, $p < 2.2e^{-16}$) indicating that higher population density is associated with marginal increases in the Consumer Price Index (CPI) Rent. However, given that population density values across most regions in the dataset are not significantly high, this suggests that density has a notable impact primarily in areas with exceptionally high population densities. This means the effect of population density on CPI Rent is significant, but only for a select group of regions where density levels are substantially higher than average.

3. Metro Areas: The presence of metro designation (`metro == TRUE`) substantially increased CPI Rent ($\hat{\beta_{metro}} = 173.7$,$p < 2.2e^{16}$).
4. State Effects: State fixed effects (`fips_state`) showed considerable variation. For example, being in state 15 (Hawaii) significantly increased CPI Rent ($\hat{\theta_{15}} = 552.87$, $p < 2.2e^{-16}$) relative to the baseline.
5. Housing Characteristics: Variables like the number of rooms ($\hat{\beta_{rooms}} = 20.82$, $p < 2.2e^{-16}$) and the presence of complete plumbing ($\hat{\beta_{plumbing}} = 129.4$, $p < 2.2e^{-16}$) were positively associated with CPI Rent. Conversely, older age and inclusion of a kitchen had negative effects.

6. Variance Inflation Factor (VIF) Analysis: The VIF values for all predictors were below 3, except for bedrooms (GVIF = 3.019), indicating no severe multicollinearity. This strengthens the robustness of the model's coefficient estimates.

7. Residual Analysis: The residual standard error was 353, indicating the average deviation between observed and predicted values. Residuals ranged from $-6,089.9$ to $4,835.5$, with a median near zero, suggesting an overall balanced fit.

8. Limitations and Future Research: Although the model captures significant variance, 56.85 of CPI Rent variability remains unexplained. Including additional predictors like economic trends, housing policies, and neighborhood amenities could enhance model performance.

Additionally, to further justify the use of a two-way fixed effects model, we conducted additional regressions on house rents using three alternative models:

- Time fixed effects only (no state fixed effects),
- State fixed effects only (no time fixed effects), and
- A model without either time or state fixed effects.

The results of these models are also summarized in Table **??**.

Focusing on the adjusted R-squared values, we observe a significant drop when removing the state fixed effects, which suggests that much of the variation in the dependent variable is explained by differences across states. While removing time fixed effects doesn't result in a similar reduction in R-squared, this does not mean that time fixed effects are not important. There may be a number of possible explanations for this, the most likely is that time fixed effects may just capture less variation in the outcome compared to state fixed effects, or that their contribution to the variation is already explained by state effects or other independent variable in the model.

Thus, while the adjusted R-squared metric is an important metric, including time fixed effects is justified by their ability to control for temporal unobservable trends that would otherwise bias the model. Therefore, a two-way fixed effects model ensures a robust and reliable estimation of house rents across the U.S.

3 Analysis of Specific Factors

3.1 Data and Description

For rents, we use the data set of Fair Market Rents provided by the U.S. Department of Housing and Urban Development, which provides annual data on the 40th percentile of rent on a county level from 1995 onwards. Data on sales prices come from the Survey of Construction, administered by the U.S. Census Bureau and U.S. Department of Housing and Urban Development, and come in the form of quarterly median sales price of houses sold in the United States, from the start of 1963 until July 2024. Construction price indices for houses sold and houses under construction are also sourced from the Survey of Construction data. Indices for houses under construction are monthly from 1963, while indices for houses sold are quarterly from 1963. Both are Laspreyes price indices, with the houses sold index being based on the cost of building the house and the value of the land, and the houses under construction index being based on only the cost of building the house. The CPI series used is the index for all urban consumers and all items, and comes from the Bureau of Labor Statistics, and is available on a monthly basis until September 2024. We also use the Realtor.com data set as explained earlier

3.1.1 Analyzing Supply Costs

Traditional economic models state that the price of a good is affected by the demand and supply functions. A change in the supply function for housing from, for example, an increase in construction costs or the price of land would increase the price of housing. We can control for these factors with the Survey of Construction's price indices. We divide median price of purchased houses by the houses sold index to yield the variable sold_corrected: the price of houses relative to cost of construction and land, thus controlling for the cost of supply.

As shown in Figure 2, House price data shows a significant break around 2011, in the aftermath of the Great Recession. Given the massive effect the Great Recession had on the housing market and the large government response, the sharp increase in housing prices may represent a structural break in the data. For this reason we prefer to limit our data range to 2012 onwards.

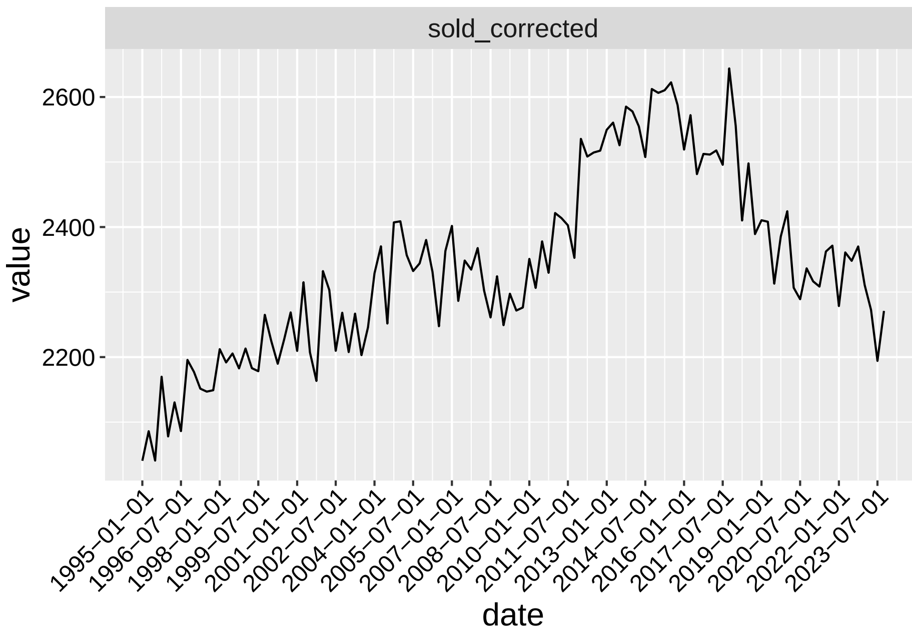

Figure 2: Graph of sold corrected 1992-2023

In figure 3 We contrast prices controlled for inflation only with CPI (cpi_mspus) to prices controlled for with the houses sold index. The most striking feature of this figure is that while cpi_mspus sees a sharp increase in prices starting at roughly 2020, sold_corrected remains stable. Additionally, while cpi_mspus increases steadily until roughly 2018 and stabilizes until 2020, sold_corrected starts stable and then decreases.

While the methodology does not imply a causal interpretation, it seems highly likely that trends in the housing price are caused by trends in construction costs. Supply costs are a major factor in the cost of almost all goods, and it is hard to conceive of any confounding variable correlated with price and supply cost that could dwarf the effect of supply. Additionally, we graph inflation-controlled housing prices (cpi_mspus) and the inflation-controlled houses sold index (cpi_sold) in Figure 4 to show that trends in both coincide.

One factor that could affect house sales price on the demand side is the interest rate: a lower interest rate increases the demand for housing due to more favorable terms on mortgages. To eliminate this factor, we also use the price of houses purchased with cash (full price up front with no mortgage or loan) which should be unaffected by the interest rate. This specification is presented in Figure 5 After controlling with the houses sold index, we see the same general trend, albeit with much more volatility. The average price is stable until 2018, starts decreasing until 2020, then seems to stabilize again.

Additionally, we can control using the houses under construction index instead, which only accounts for

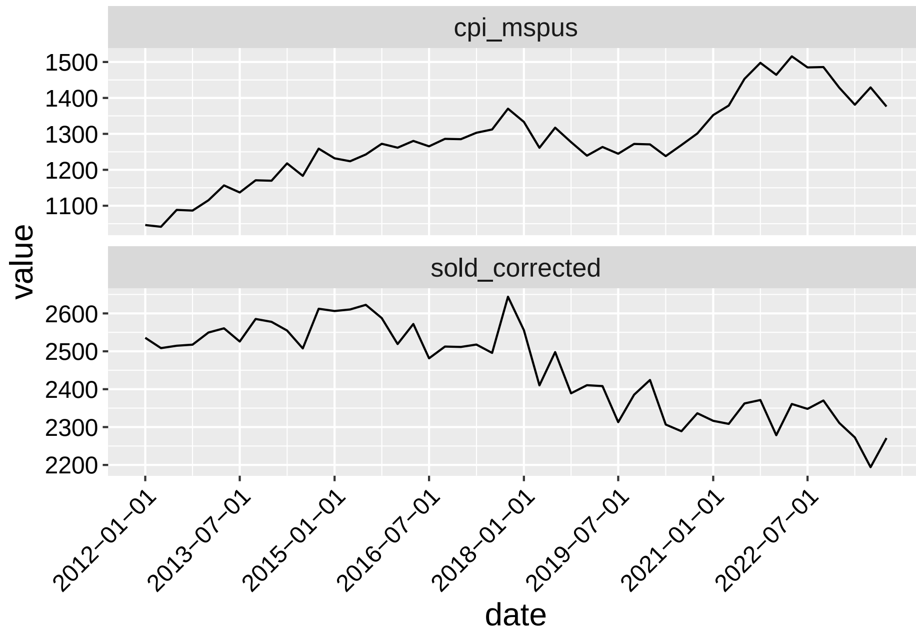

Figure 3: Comparison between cpi mspus and sold corrected

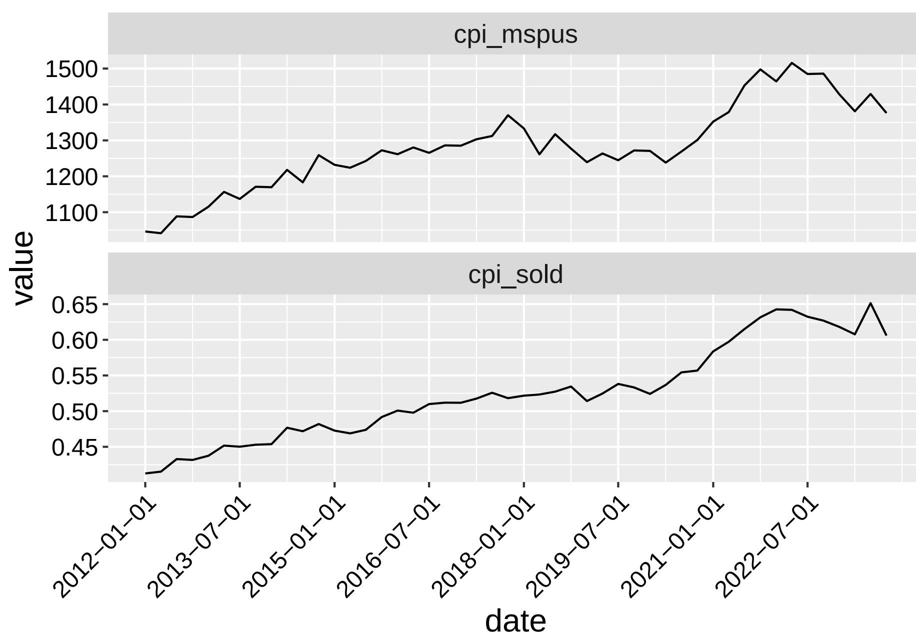

Figure 4: Comparison between cpi mspus and CPI-controlled houses sold index

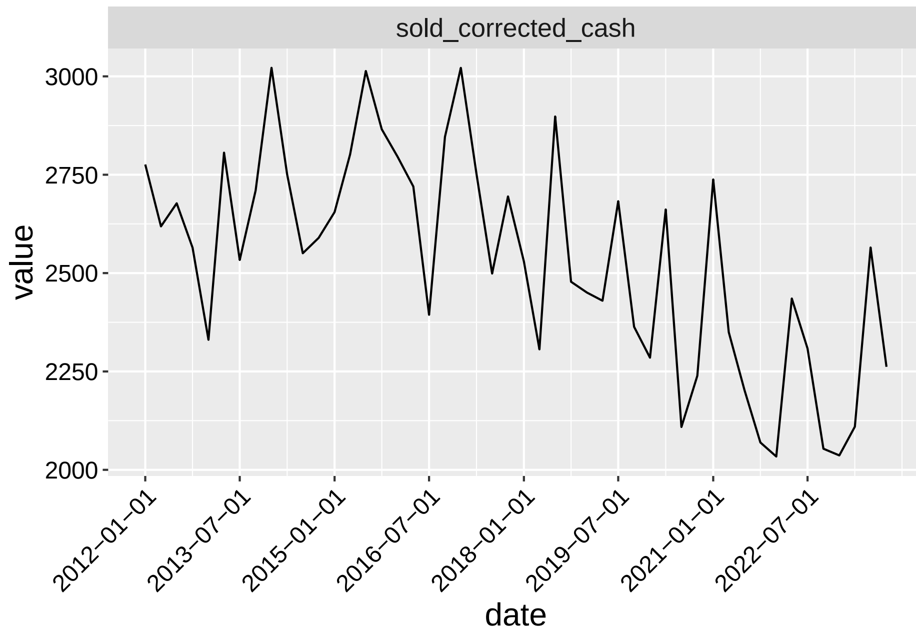

Figure 5: Prices of houses purchased with cash controlled with houses sold index

construction costs, and contrast with the price controlled by houses sold index. We present this specification in Figure 6 Both variables appear near-identical, indicating construction costs as the main driver of the effect we see.

3.2 Zoning synthetic controls

The practice of zoning, or restricting what types of construction can occur on certain areas, is often blamed for driving up the price of housing by artificially limiting supply. One particular form of zoning that receives significant attention is single-family zoning, only allowing the construction of single-family homes. As part of the comprehensive Minneapolis 2040 plan, taking effect January 1, 2020, Minneapolis passed a law ending single-family zoning, a feat few other cities have matched. Additionally, the city implemented inclusionary zoning, requiring developers to have some of their units to be rented at prices below certain percentages of median income in the area.

To assess the effectiveness of these policies in reducing local rents, we utilize a synthetic control approach, as we have no reason to believe the parallel trends assumption should hold. The control group is comprised of metro areas as designated by the U.S. Department of Housing and Urban Development in the states of Minnesota (which contains Minneapolis) and Wisconsin. One limitation of our data is that the metro area containing Minneapolis also contains the slightly smaller neighboring city of St. Paul, which was not directly affected by the laws passed. This should dilute the overall effect, but so long as there are no simultaneous shocks to housing prices in St. Paul, a causal interpretation is still possible. We could also expect to see spillover effects decreasing rent in St. Paul as landlords compete with lower prices. For predictive variables, we use population in 2010 and market factors from 2016.

Our results indicate a increase in rent in the metro area containing Minneapolis and St. Paul compared to the synthetic control. As outlined in Abadie et al. (2010), we compute Fisher's Exact p-value as 0.5714, indicating no signifiant change in rents. The results of the synthetic control model are presented in Table 5 and Figure 7. While the synthetic control fits very well from 2018-2020, the two years before the intervention period, it does not fit as well in the years before, which is a possible limitation of the model.

We observe an increase in rents instead of a decrease in rents, alleviating concerns that insignificance is

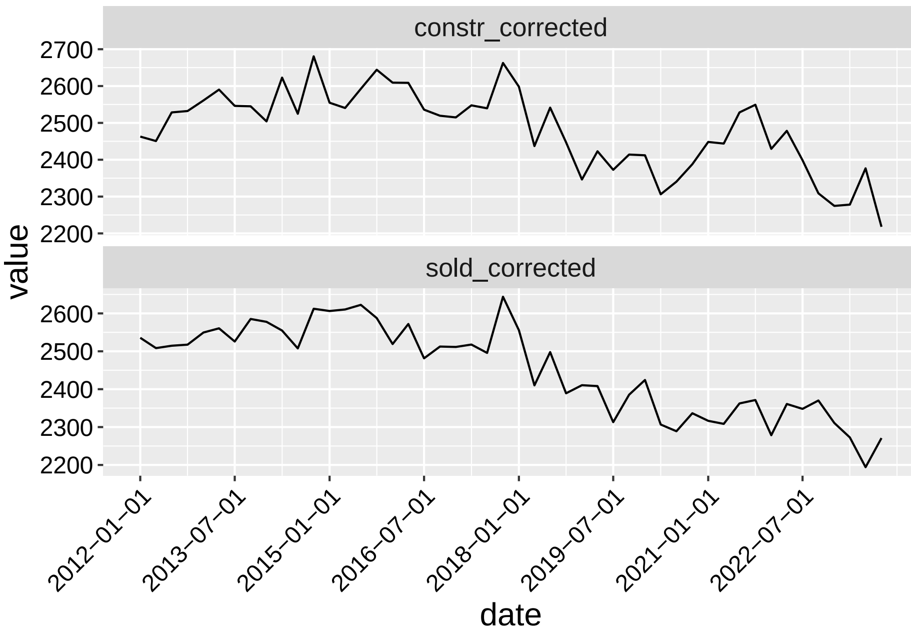

Figure 6: Comparison of houses sold and construction indices

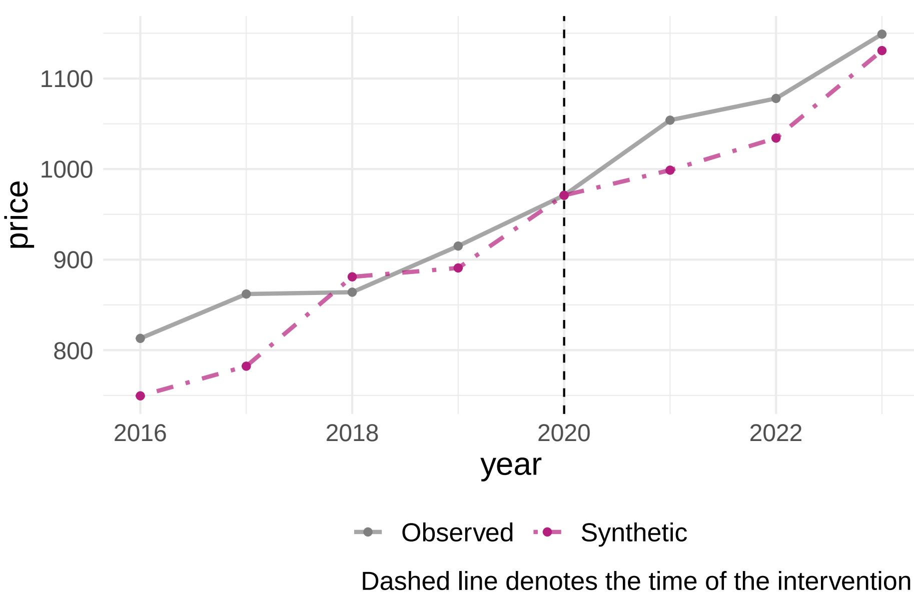

Figure 7: Synthetic vs. observed fair market rent

Table 5: Results of synthetic control

Year	Observed	Synthetic	difference
2016	813	749.5327	63.46727
2017	862	782.2895	79.71053
2018	864	880.9781	-16.97811
2019	915	890.7804	24.21961
2020	971	971.0000	0.00000
2021	1054	998.7758	55.22415
2022	1078	1034.1920	43.80803
2023	1149	1130.8338	18.16621

caused by the breadth of the metro area containing Minneapolis. One reason for the insignificance of the effect to rent is that the elimination of single family zoning was less effective in reality than may have been imagined. Of the roughly 12,600 housing structures authorized in the Minneapolis-St Paul. metro area, only roughly 350 were 2-4 units[3], which could only a house a tiny fraction of the population of the cities and would make a small dent in rent prices.

[3]https://www.census.gov/construction/bps/index.html

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Bureau, US Census. 2019. "Survey of Construction (SOC)." Census.gov. https://www.census.gov/construction/soc/index.html.

———. n.d. "Building Permits Survey (BPS)." census.gov. https://www.census.gov/construction/bps/index.html.

"Early Effects of COVID-19 Pandemic-Related State Policies on Housing Market Activity in the United States." 2022. *Journal of Housing Economics* 57. https://doi.org/https://doi.org/10.1016/j.jhe.2022.101857%7D, author = {Yörük, B.}, .

FRED. 2023. "Consumer Price Index for All Urban Consumers: All Items." Stlouisfed.org. https://fred.stlouisfed.org/series/CPIAUCSL.

Housing, U. S. Department of, Urban Development Office of Policy Development, and Research. 2017. "Fair Market Rents | HUD USER." Huduser.gov. https://www.huduser.gov/portal/datasets/fmr.html.

Kendall, Ross, and Peter Tulip. 2018. "The Effect of Zoning on Housing Prices." *Reserve Bank of Australia Research Discussion Paper*, no. 2018-03.

Michael Manville, Paavo Monkkonen, and Michael Lens. 2020. "It's Time to End Single-Family Zoning." *Journal of the American Planning Association* 86 (1): 106–12. https://doi.org/10.1080/01944363.2019.1651216.

Minneapolis, City of. 2019. "Residential Buildings with up to Three Units." Minneapolismn.gov. https://www2.minneapolismn.gov/business-services/planning-zoning/amendments/adopted-proposed/recently-adopted/residential-buildings-3-units-amendment/.

Minneapolis, City of. 2024. "Inclusionary Zoning Requirements." www2.minneapolismn.gov. https://www2.minneapolismn.gov/government/projects/cped/inclusionary-zoning/.