Evaluating Hope Brings You Home's Effect on Low-Income Homeownership in Nevada

Annabella Stoll-Dansereau

December 2023

Introduction

This research was initially driven by the desire to *do something* with the wealth of Mortgage data made publically available through Fannie Mae and Freddie Mac under disclosure requirements. This publication was largely influenced by the Federal Housing Finance Agency regulations and mandates by Congress specifically, the Housing and Economic Recovery Act of 2008.

Policy in question

The research focuses on the "Hope Brings You Home Down Payment Assistance Program" in Nevada, a policy designed to aid homebuyers. This initiative provided a maximum of \$20,000, or 10% of the home's purchase price, as a forgivable loan with no interest charged in the initial years and full forgiveness after three years of residency. Eligibility criteria included a household income below \$98,500 and the purchase of a home within designated 'distressed' zip codes. In 2018, Nevada's Hardest Hit Fund allocated \$36 million to this program, which was depleted fairly quickly, indicating high demand and no carry-over into 2019. Although the program resumed in 2020, it falls outside the scope of our time frame. Figure 1 illustrates the program's coverage visually.

Simplified Diagram of Determining Casualty (DAG)

Here is a simple model, shown by Figure 2, of how I am identifying the causality of this policy. Conditional on the relative income level of zip code and the minority percentage (as White individuals own the most homes) we can isolate the effects of the policy on mortgage rates acting as a proxy for homeownership rates under lower-income households without the funds to outright buy a house.

Research Question and Relevance

This research seeks to examine the effectiveness of the *Hope Brings You Home Down Payment Assistance Program* in Nevada defined by an increase in homeownership rates within these

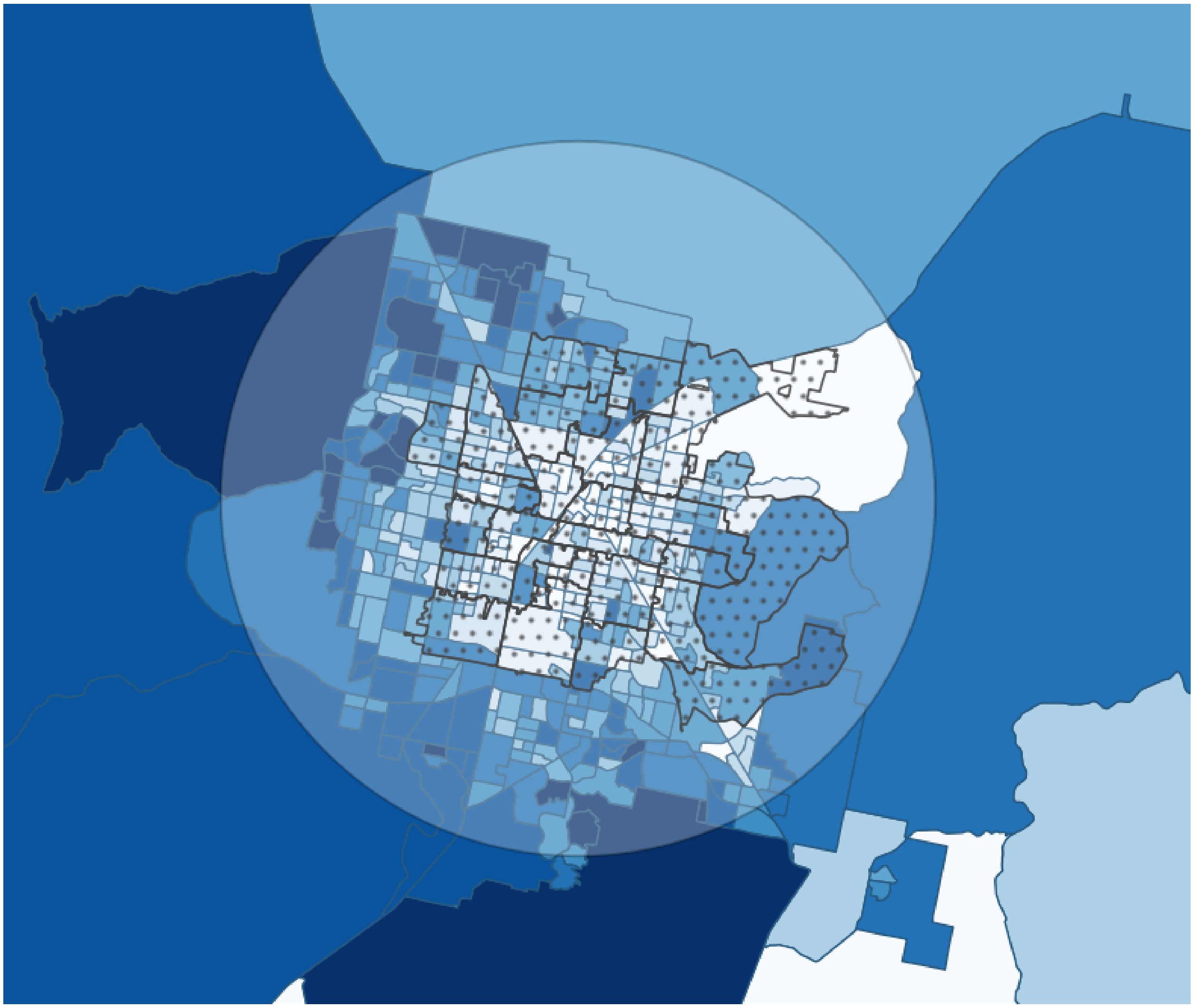

Figure 1: The circle orients you to the greater Las Vegas Area, lighter colours represent lower-income census tracts, and the dotted area shows the zip codes impacted by the policy

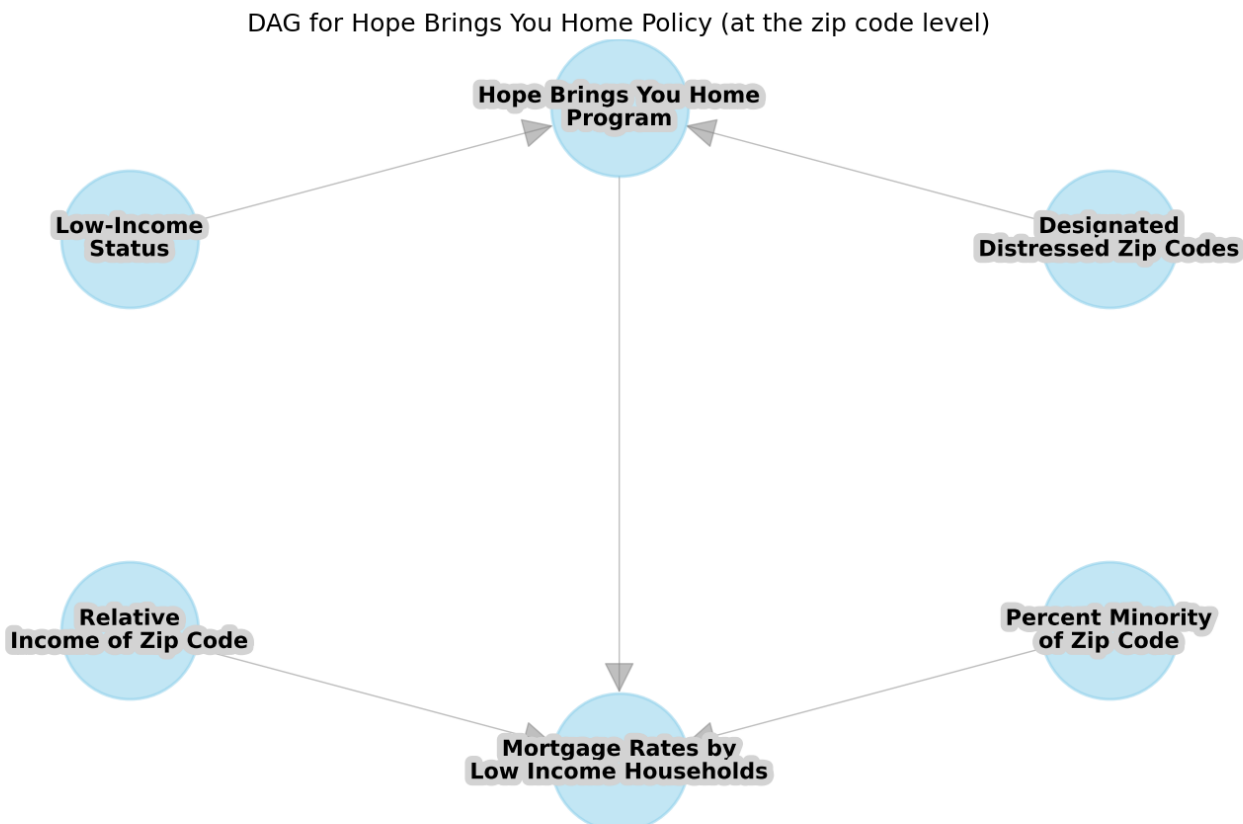

Figure 2: DAG identifying casual path

treated zip codes.

This specific study doesn't have great generalizability, especially given the unique nature of Las Vegas and this policy however it would give the Nevada government valuable feedback. If the program proves to be an efficient use of the funds, evidenced by an increase in homeownership rates among the targeted low-income populations in the designated distressed zip codes, it would justify not only the continuation but also the potential expansion of such initiatives. Conversely, if the analysis reveals a negligible effect, it could suggest the need for the Nevada government to reevaluate the allocation of funds towards more efficacious measures to foster low-income homeownership in the future.

Literature Review

Mortgage and Credit Constraints

Credit constraints represent a significant barrier to homeownership. After the 2008 financial crisis restrictions tightened on lower-income households who may not have as high credit scores, higher debt-to-income ratios and other negative financial indicators. They would find it increasingly harder to find an institution willing to grant them a mortgage (Acolin et al., 2023). There has been research done on the optional credit constraints on consumers (Eberly and Krishnamurthy, 2014) however what might be optimal for an institution can be harmful to general homeownership rates amongst the lower income households.

There have been many governmental policies put into place to increase access to homeownership. Some of which have been subject to economic literature research. One example is the *Help to Buy* program in the UK by Carozzi et al. (2020) which found policies expanding credit restrictions are ineffective in an unaffordable and already tightly supply-constrained area. This shows that merely increasing access without further help doesn't have much of an impact on the market. Another study by Grinstein-Weiss et al. (2013) looks at the implications of a downpayment assistance program. The original findings suggest an immediate positive impact in homeownership rates but this value decreases in the long run to insignificant levels. These two papers show how policies introduced into the housing market with the intention to help increase the accessibility of homeownership is not as simple as my DAG diagram suggests.

However when creating a DAG this paper Carliner (1974) was useful in determining covariates. Many important indicators were mentioned relating characteristics to homeownership rates. Among other variables race and income levels play a large role in homeownership. This study on the direct impacts of downpayment assistance by the U.S. Department of Housing and Urban Development (Herbert and Tsen, 2005) also finds the primary factor that restricts access to homeownership is income and thus downpayment assistance programs targeting lower-income individuals will have a significant impact on homeownership rates in this demographic.

Matching Methods in Policy Analysis

The methodological motivation came directly from the online textbook *Casual Inference for the Brave and True* by Alves (2022). *Chapter 13- Differences-in-Differences, Chapter 15- Synthetic Control*, and *Chapter 25- Synthetic Differences-in-Differences* gave me the inspiration for this topic. Particularly the author used the example of the policy on Tobacco in California (Abadie et al., 2010) where the synthetic control was created from the other states. I figured this could be extended to analyze the mortgage policy described earlier. The code used to implement my methodological section was adapted from this textbook. While there are many packages available the transparency of this simple method was appealing.

Summary

The existing literature underscores the potential of mortgage relief programs to address credit constraints and promote homeownership among vulnerable populations. By employing matching methods, particularly synthetic control, this study contributes to the understanding of such programs' effectiveness through a case study. Aligning with the insights from Abadie et al. (2010) and methodical help from Alves (2022), this research will help assess the impact of Nevada's *Hope Brings You Home Down Payment Assistance Program*, offering an addition to the discourse on policy efficacy in the realm of housing economics.

Data

Source Data

The data, as mentioned earlier, is from the publicly available mortgage datasets released by Fannie Mae and Freddie Mac. These datasets are provided as a Microsoft Access database. Specifically, I used the "Single-Family Census Tract File" from 2012-2019. The variables included throughout the years changed for example in later years we see information such as credit scores and additional tract-level characteristics plus lots of borrower characteristics which were unhelpful for our analysis.

Originally I hoped to merge additional covariates from IMPUS data however due to computational difficulties with just these original datasets I chose only to use covariates included here which were percentage minority and relative income levels of the tract to the surrounding area all from the 2010 CPS data. My dependent variable is the total mortgages taken out by low-income households that year (a calculated value but implicitly in the dataset of course).

You may have noticed how my mortgage data's unit of observation is the census tract but the policy was described on the zip code level. To crosswalk between the two I used IMPUS GIS data from 2018 to match these two geographical boundaries. They far from perfectly overlap and as a judgment call I chose to include census tracts which were partially included in the treated zip code rather than only including tracts that were fully included in the zip code. There is a small margin of error introduced by this method however too many census tracts were lost under the fully enclosed method. The crosswalked dataset was derived using

QGIS software and merging multiple layers and converting it to a csv which is attached to the GitHub page.

Transformations to the Data

The biggest challenge in this project was dealing with large datasets which became computationally infeasible for my laptop very quickly. As such I had to make simplifying assumptions to subset my data before I started any of the methodological analysis. Since I was trying to create a synthetic control I decided my control should consist of major cities in Arizona and the Interior of California as those were most similar to my treatment. So I subsetted down to the counties including these cities: Las Vegas, Sacramento, Reno, Tuscan, Phoenix and Fresno.

Once I had these cities another important transformation was needed to make sure our low-income category was comparable across these cities with different costs of living. Rather than creating some IMPUS-inspired cost of living comparison, I used an online tool from Forbes Advisor (2023) which tells you how much you would need to live by equivalent standards in another city with $98,500 in Las Vegas set as the baseline. In the end, I only used Fresno and Tuscan counties as my control pool (which I will describe later in the methodology) again due to computational difficulties.

After reading in and transforming all the data into a manageable dataset we were left with a dataset consisting of the covariate characteristics we described earlier at the census level and total mortgages by census tract for each year.

Methods

To try and identify the causal impact of this policy I will be utilizing synthetic control to match each census tract to a combination of census tracts in Tuscan and Fresno. From here I will run a differences-in-differences specification (DiD) controlling for the census tract to determine the impact of this policy.

Synethetic Control

In its most simple form, synthetic control is attempting to create some $\hat{Y}_{jt}^{N}$ from a collection of $\sum_{j=2}^{J+1} w_j Y_{jt}$ where Y_{jt} is one of our census tract observations and the hat denotes the control created from a series of these weights.

We want to minimize the distance of the control from the true value of the treated in the period before the policy (in our case 2017 and before). This will create a control which closely follows our treatment before the policy then we can analyze what happens after our treatment period.

To achieve this objective of minimizing the distance between our covariates X_{hi} where $i = 1$ denotes the treated covariate in question (h) and $i = j$ is the same covariate h for different tracts in our control dataset. Like the previous equation, we create a series of weights $\mathbf{W} = (w_2, \ldots, w_{J+1})$ that minimize the equation

$$\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\|^2 = \left(\sum_{h=1}^{k} v_h \left(X_{h1} - \sum_{j=2}^{J+1} w_j X_{hj}\right)^2\right)^{\frac{1}{2}}$$

Here v_h represents the importance of each variable when minimizing the difference between the treated and the control. But for this analysis, I took $v_h = 1$. One important thing was normalizing all my data so that the percentages and mortgage units were similar. Without this normalization, we could get disproportionate results as this would essentially be choosing a $v_h \neq 1$ without control and rather randomly based on the underlying format of the data.

Differences-in-Differences

In a DiD set up we are trying to determine the average treatment effect on the treated, in mathematical notation

$$A\hat{T}ET = (E[Y(1)|D = 1] - E[Y(1)|D = 0]) - (E[Y(0)|D = 1] - E[Y(0)|D = 0])$$

Usually, in DiD setups much effort is devoted to justifying that the parallel trends exist before the treatment. But we know from the synthetic control that our trends before the intervention are extremely similar such that the differences after the treatment time-period should be due to the treatment rather than because the control was fundamentally different than the treatment to begin with.

I will use a regression to model this $A\hat{T}ET$ including fixed effects for my census tracts. This is to try and absorb some of the heterogeneity between zip codes so that we can see the impact of this policy on the entire treated region rather than focusing on specific census tracts which is a fairly noisy estimate of our outcome variable. Below is the basic model I used:

$$\text{Distressed Area Mortgages} = \beta_0 + \beta_1 \text{Treated} + \beta_2 \text{Post} + \beta_3 \text{Treated} \times \text{Post} + \sum_{k=1}^{K} \gamma_k C(\text{Tract}_k) + \varepsilon$$

The variables outlined are pretty self-explanatory. There are some issues with this methodology mostly resulting from data simplifications and the inability to do robustness checks but these will be addressed in the Future Steps section.

Results

Synthetic Control

First I will demonstrate the results of the matching which will hopefully instill confidence in the prior-to-treatment common trends assumption. On aggregate, we see that the number of low-income mortgage observations we are interested in matched very well see Figure 3.

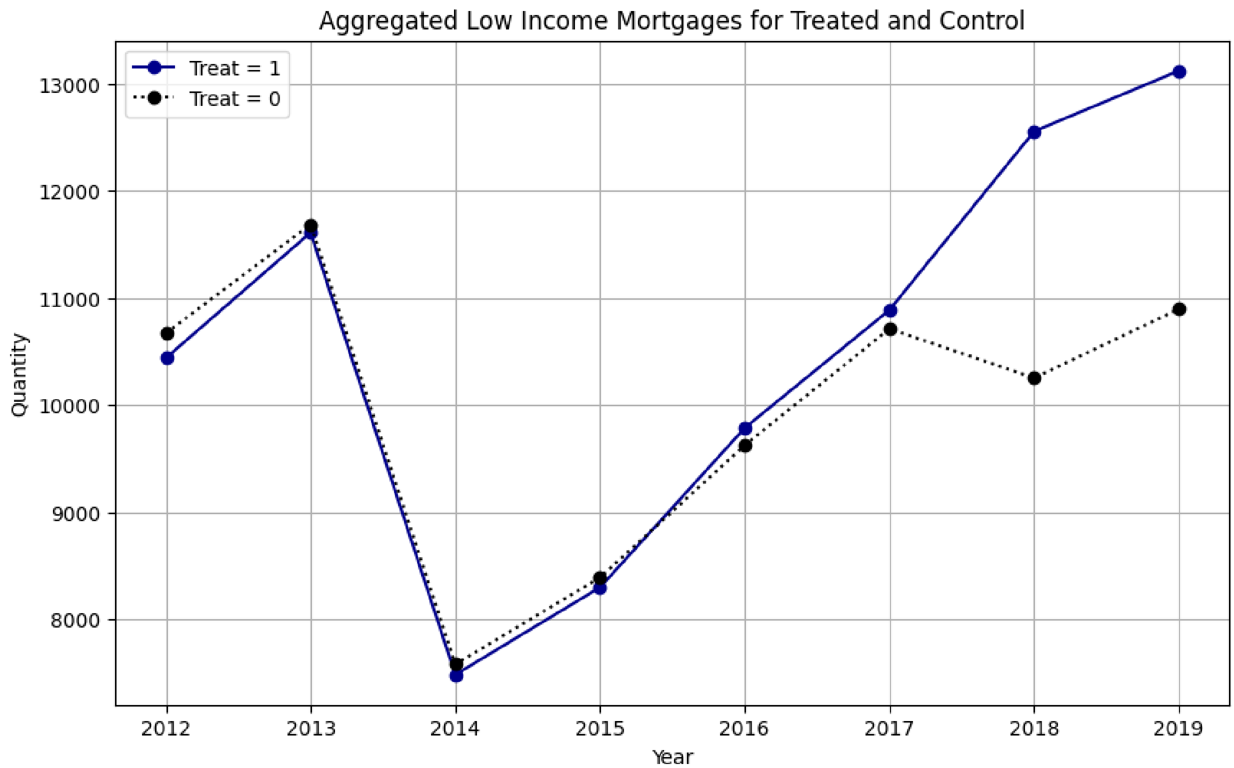

Figure 3: Pre-Trend Matching

Since this matching is comprised of a series of sub-matchings for each census tract it is worth showing a few of the disaggregated matches.

For the following, I set random seeds to grab 3 tracts. In Figure 4 we see an example of perfect matching where they are practically identical. While a lot of the matches look like this from the aggregated Figure 3 graph we know not every zip code was perfectly matching, some look more like the dark blue line in Figure 5. This is somewhat expected since heterogeneity across specific tracts contains a lot of noise.

Additionally, notice how the treatment isn't homogenous in its effects. In both Figure 3 and Figure 4 we see the dark blue census tract is actually decreasing relative to the control. Again we cannot gain that much information from overanalyzing individual tracts since there are so many factors that may have impacted such a small region but it is promising we see a positive divergence in the aggregated Figure 2.

Differences-in-Differences

The results from the DiD regression are shown in Table 1 below. We see that the coefficient we are interested in to measure the $\hat{ATET}$ is significantly positive implying this policy did have a positive impact on the number of mortgages taken out at the tract level for individuals. In the model outputted by my Python code, there are hundreds of fixed effects estimates. These were omitted from this table due to the irrelevance of the specific outputted values

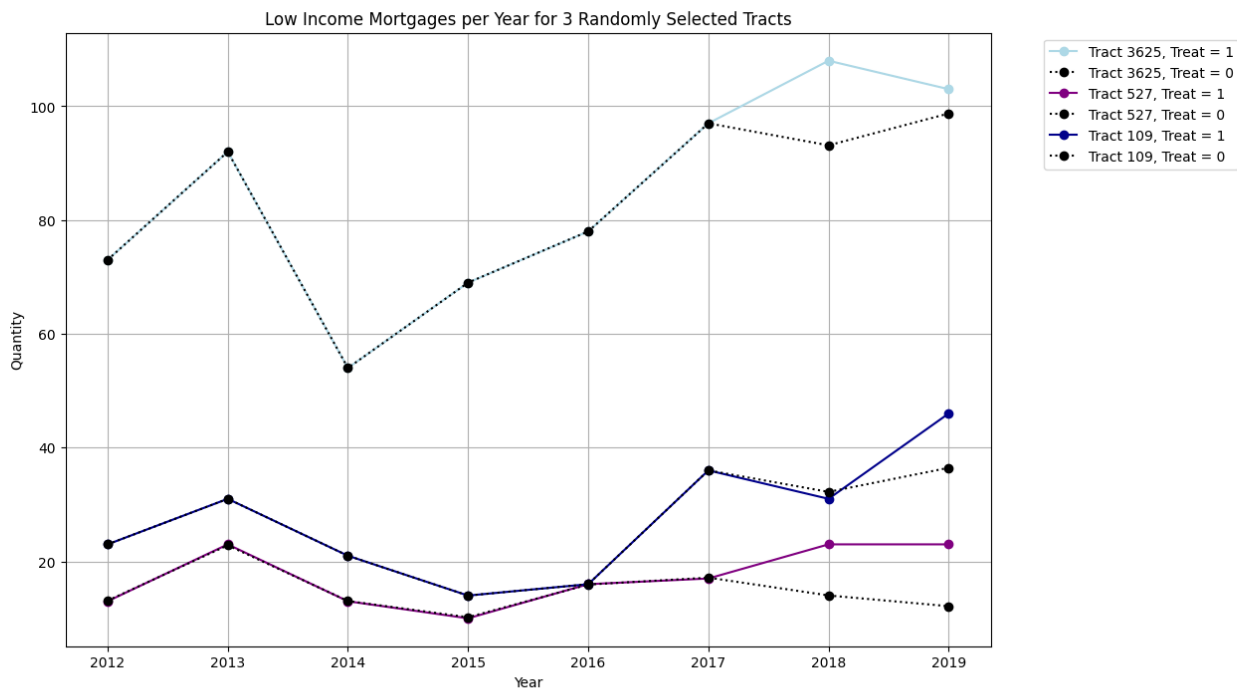

Figure 4: Perfect Matching

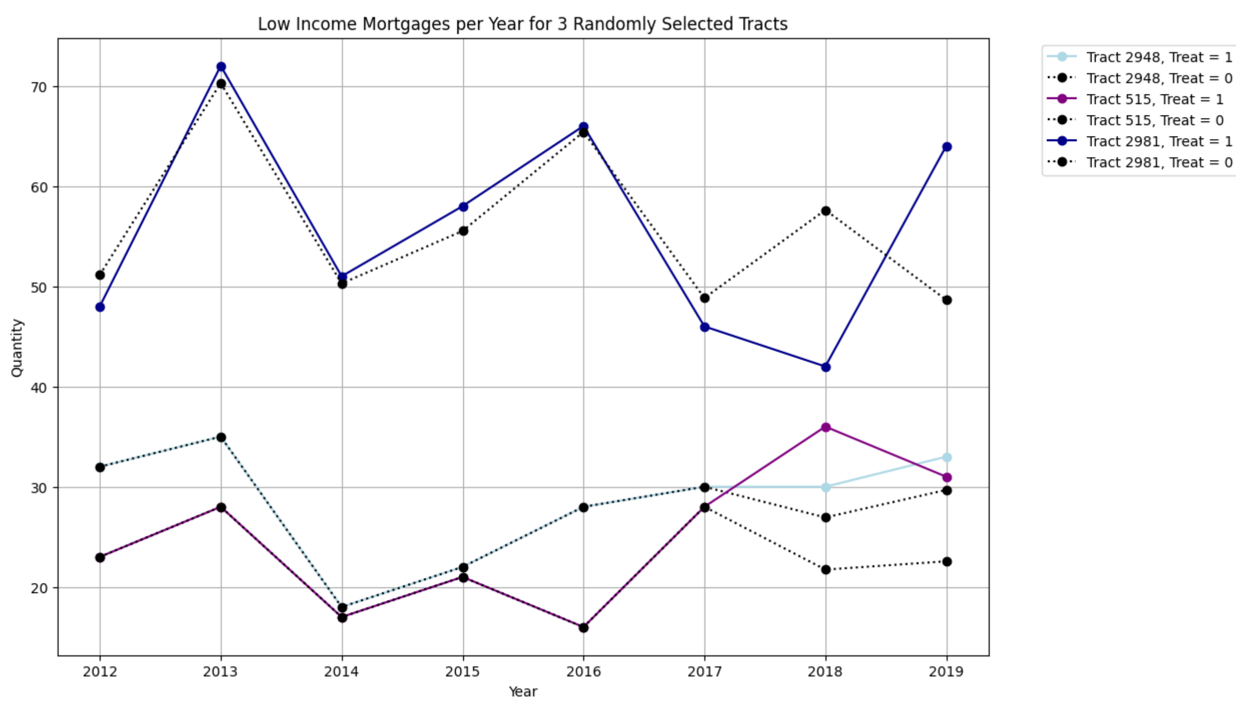

Figure 5: Dark Blue Imperfection

and the sheer magnitude of estimates which would take away from the main point of interest which is the *Treated:Post* variable.

The observed coefficient of approximately 9.7 warrants a cautious interpretation, yet it aligns closely with our expectations. Given that our treatment group has 286 unique tracts, this suggests an aggregate increase of approximately 2775 mortgages. Recall from the introduction that the exhaustion of the fund was projected at 2000 mortgages, assuming maximal utilization of the available resources. This discrepancy suggests the possibility of a minor spillover effect or that the housing prices were sufficiently low to prevent the full utilization of the downpayment assistance per mortgage. The absence of detailed data regarding the individuals who purchased homes through this program poses a challenge in conclusively differentiating between these scenarios. Nonetheless, the congruence of the findings with the expected outcomes, based on our understanding of the policy's specifics, provides a measure of reassurance regarding the validity of the results.

Table 1: Regression Results

	Coef.	Std. Err.	t	$P > \|t\|$	0.025	0.975
Intercept	75.9033	3.313	22.911	0.000	69.408	82.399
Treated	-0.1138	0.496	-0.229	0.819	-1.087	0.859
Post	3.3797	0.702	4.815	0.000	2.004	4.756
Treated:Post	9.7065	0.993	9.779	0.000	7.760	11.653

Observations: 3776 Adj. R-squared: 0.859 Tract Fixed Effects: True

Future Steps

As you hopefully noticed with my hinting throughout this report there were some flaws with the analysis mostly coming down to computational and time restraints. As such I want to address each concern I have in turn. While the inference should be taken with a grain of salt I hope the methodology was clear and provided an idea of what I would've done should the timeframe of this project be extended and I gain access to a server to submit jobs to rather than constantly crashing my jupyter kernel.

Covariates

As mentioned throughout I initially hoped to merge a series of covariates I pulled from the IMPUS CPS dataset to enrich my picture of each census tract including items like homeownership rates, educational levels etc. But my computer was struggling just with the data in question so I didn't want to expand my dataset. Under different conditions, if I merged on the covariates I pulled I would have more confidence in the control and treatment matching. There is the potential these results are overfitting too few variables rather than finding truly representative controls.

Subsetting of Control Pool

As mentioned throughout I had to subset my control group to only the counties including Tuscan and Fresno. Ideally, I wanted to include all of Nevada, Arizona and California to match on. I think this approximation was okay however it may have found more similarities and created a better control if it had more tracts to match with. This would have been more important if I included more covariates, part of why I could get away with this subsetting was due to matching on only a few things.

Robustness Checks

Finally, I didn't have any robustness checks particularly to see whether this result could've happened by chance. A technique described by both the textbook and class lectures of perturbing the treated variable so we no longer had a true treatment but randomness that we were matching on to get a treatment effect would have been a good test to run. If we perturbed this say 100 times we could see where our estimates fell in the series of randomness. If the actual treated tracts fell within the top 5 highest treated effects found by the random noise we could say with 95% confidence that our results of significant and this policy had a positive effect on mortgages taken out by low-income households.

Conclusion

The "Hope Brings You Home Down Payment Assistance Program" in Nevada was launched to increase homeownership rates among low-income households. The empirical analysis conducted in this research, utilizing a synthetic control and difference-in-differences approach, suggests that the program had a positive impact on the number of mortgages taken out by low-income households in distressed areas.

The synthetic control method provided a robust pre-treatment match, allowing for a credible estimation of the program's effect. Despite computational limitations that narrowed the control pool, the analysis revealed a significant positive coefficient for the interaction of treatment and post-treatment periods. This finding indicates that the policy likely contributed to a rise in homeownership among the targeted demographic.

However, several limitations of this study must be acknowledged. The computational constraints led to a reduced set of covariates and a limited control pool, potentially affecting the precision of the control match. Furthermore, the absence of robustness checks, such as permutation tests, means that the results should be interpreted cautiously.

While this research presents encouraging results regarding the effectiveness of down payment assistance programs, further research is essential to fully understand their impact and to guide policymakers in the design of interventions that promote equitable access to homeownership. However, just from this research, it is promising to see that this policy achieved its objective and had a positive impact on low-income homeownership.

References

Abadie, A., Diamond, A., and Hainmueller, J. (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.

Acolin, A., Goodman, L., and Wachter, S. M. (2023). Accessing homeownership with credit constraints. Wharton School, University of Pennsylvania, Samuel Zell and Robert Lurie Real Estate Center. Working paper no. 803.

Alves, M. F. (2022). *Causal Inference for the Brave and True.*

Carliner, G. (1974). Determinants of home ownership. *Land Economics*, 50(2).

Carozzi, F., Hilber, C. A. L., and Yu, X. (2020). On the economic impacts of mortgage credit expansion policies: Evidence from help to buy. *Journal of Urban Economics*, 139.

Eberly, J. and Krishnamurthy, A. (2014). Efficient credit policies in a housing debt crisis. *Brookings Papers on Economic Activity*, 45.

Forbes Advisor (2023). Cost of living calculator. https://www.forbes.com/advisor/mortgages/real-estate/cost-of-living-calculator/. Accessed: [insert date of access here].

Grinstein-Weiss, M., Sherraden, M., Gale, W. G., Rohe, W. M., Schreiner, M., and Key, C. (2013). Long-term impacts of individual development accounts on homeownership among baseline renters: Follow-up evidence from a randomized experiment. *American Economic Journal: Economic Policy*, 5(1):122–145.

Herbert, C. E. and Tsen, W. (2005). The potential of downpayment assistance for increasing homeownership among minority and low-income households. Report, U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Cambridge, MA.