Did Pandemic Relief Fraud Inflate House Prices?*

John M. Griffin[†] Samuel Kruger[‡] Prateek Mahajan[§]

May 28, 2024

Abstract

Pandemic fraud is geographically concentrated and stimulated local purchases with effects on prices. Fraudulent PPP loan recipients significantly increased their home purchase rate after receiving a loan compared to non-fraudulent PPP recipients, and house prices in high fraud zip codes increased 5.8 percentage points more than in low fraud zip codes within the same county, with similar effects after controlling for other explanations for house price appreciation during COVID. Zip codes with fraud also experience heightened vehicle purchases and consumer spending in 2020 and 2021, with a return to normal in 2022.

JEL classification: G21, G23, G28, H12, R21
keywords: Fraud, House Price Growth, Government Spending

*We thank Jim Conklin, Mike Faulkender, William Fuchs, Caitlin Gorback, Cristoph Herpfer, Elena Loutskina, Vrinda Mittal, Lin Peng, Alex Priest, Marius Ring, Laura Starks, Constantine Yannelis, Eric Zwick, industry participants, and others for helpful comments. We are also thankful for comments from seminar and conference participants at the American Real Estate and Urban Economics Association (AREUEA) national conference, the Financial Intermediation Research Society (FIRS) conference, the Office of the Comptroller of the Currency (OCC) Emerging Risks Symposium, the Society of Financial Studies (SFS) Cavalcade, the University of Kentucky Finance Conference, the Federal Reserve Bank of Dallas, the University of North Carolina, the University of Pennsylvania, and the University of Texas at Austin. We also thank the University of Texas at Austin and the Integra Research Group for research support. Griffin is an owner of Integra Research Group and Integra FEC, which engage in research, financial consulting, and recovery on a variety of issues related to the investigation of financial fraud.

[†]McCombs School of Business, University of Texas at Austin. John.Griffin@utexas.edu
[‡]McCombs School of Business, University of Texas at Austin. Sam.Kruger@mccombs.utexas.edu
[§]McCombs School of Business, University of Texas at Austin. Prateek.Mahajan@mccombs.utexas.edu

Did Pandemic Relief Fraud Inflate House Prices?

May 28, 2024

Abstract

Pandemic fraud is geographically concentrated and stimulated local purchases with effects on prices. Fraudulent PPP loan recipients significantly increased their home purchase rate after receiving a loan compared to non-fraudulent PPP recipients, and house prices in high fraud zip codes increased 5.8 percentage points more than in low fraud zip codes within the same county, with similar effects after controlling for other explanations for house price appreciation during COVID. Zip codes with fraud also experience heightened vehicle purchases and consumer spending in 2020 and 2021, with a return to normal in 2022.

JEL classification: G21, G23, G28, H12, R21
keywords: Fraud, House Price Growth, Government Spending

Are the costs of financial fraud largely confined to funds that are stolen, or does fraud have other, perhaps unanticipated, distortive effects? Akerlof and Romer (1993) propose that fraud can have large unintended externalities that are often greater than the direct costs, including in the form of distorted asset prices. In this paper we seek to test this theory by examining the use of fraudulent funds stolen from pandemic relief programs. Are the proceeds of the fraud used, at least in part, to purchase assets and do they thereby put upward pressure on asset prices?

In the traditional Becker (1968) crime model, the optimal amount of resources to devote to the prosecution of crime and the nature of the punishment depend on both the direct and indirect costs of the crime. For financial fraud, direct cost estimates typically range between three and nine percent of GDP (Gee and Button, 2019). Indirect costs are even harder to detect and quantify. However, federal COVID relief spending of over $4.5 trillion may provide a setting to examine potential spillovers and externalities of fraud.[1] Examining the extent to which this spending may have influenced the price of goods is difficult since most spending programs were largely designed to offset losses of income due to the pandemic and thus are cross-sectionally proportional to population and income. One potential source of regional variation in government spending is fraudulent transfers. There is growing evidence that a sizeable portion of the funds distributed by the Paycheck Protection Program (PPP), Economic Injury Disaster Loan (EIDL) program, and unemployment insurance programs may have been fraudulent (Griffin, Kruger, and Mahajan, 2023a; ProPublica, 2021; SBA OIG, 2023; Associated Press, 2023). From the ground level, one former U.S. Attorney described it as: "nothing like this has ever happened before…it is the biggest fraud in a generation" (NBC News, 2022).

In this paper, we examine the effects of fraud in the PPP and other pandemic relief programs on local housing markets and consumer spending. Fraud in the PPP was widespread

[1]See here for details on pandemic relief spending. Some of the largest categories include COVID-19 unemployment benefits at $872 billion, the Paycheck Protection Program (PPP) at $793 billion, and the Economic Injury Disaster Loan (EIDL) program at $384 billion.

and ramped up quickly in part due to lax standards by some FinTech lenders (Griffin, Kruger, and Mahajan, 2023a). Additionally, Griffin, Kruger, and Mahajan (2023b) find that some zip codes have almost no PPP fraud while others have in excess of 40% of their loans flagged as suspicious, that this fraud was highly correlated with suspicious lending in the EIDL and unemployment insurance, and that information about how to obtain fraudulent loans spread rapidly on social media and through social connections. This resulted in highly concentrated geographic pockets of PPP, EIDL, and unemployment insurance fraud in the same geographic areas.

Determining the effect of pandemic relief programs on asset prices is challenging because most programs were designed to offset lost income, and, if implemented correctly, would not provide an income shock to the borrower. For example, PPP loans were designed to cover business expenditures, such as employee payrolls, for businesses that were struggling due to lost pandemic revenue. PPP loans that offset revenue declines brought on by the pandemic would not generate excess cash for the business owner, but instead would (perhaps only partially) make up for business expenses.[2] In contrast, individuals who received funds through fraudulent PPP loans, fraudulent EIDL advances and loans, or fraudulent unemployment insurance claims potentially gained a windfall of new wealth that they could either spend or save. We first examine how pandemic relief fraud may have had unintended consequences in terms of affecting an immovable regional good–housing.

Before examining house prices, we first investigate whether individuals who received pandemic relief payments through fraudulent means were more likely to purchase homes. We match a random sample of 250,000 individual PPP recipients to property ownership records from PropertyRadar, and also utilize data on house purchases from LexisNexis and data on address changes from Melissa Data. All three sources indicate a sizable upward shift in house purchase probability for recipients of flagged PPP loans as compared to non-flagged

[2]Loans and grants made by the EIDL program were also designed to replace lost revenue and unemployment insurance is designed to cover at least part of income lost due to unemployment.

recipients. In particular, in a difference-in-differences framework, we find that the probability that an individual purchased a house increased by 17% after receipt of a suspicious loan.

Given the higher rate of house purchases by recipients of fraudulent PPP loans and the concentration of pandemic relief fraud, it is possible that house purchases by recipients of fraudulent funds from the PPP and other government programs could have distorted local house prices. We use information on PPP loans for most of our analysis since more detailed information is available for them, but the results from Griffin, Kruger, and Mahajan (2023b) suggest that the effects are likely due to the same geographic areas receiving fraudulent funds through multiple pandemic relief programs, including EIDL loans, EIDL Advance, and unemployment insurance claims. To control for macro factors that may have influenced regional house price growth, our analysis focuses on the zip code level with county fixed effects. We also control for house price growth over the prior two years and other variables that previous research has found to be associated with house price growth during the COVID-19 pandemic. We find that zip codes with high suspicious lending per capita experienced house price growth that is 5.8 percentage points (ppt) higher compared to zip codes with low suspicious lending per capita.[3] This is a sizable 22.5% of the 25.9 ppt average increase in house prices during 2020 and 2021. Additionally, the monthly timing of the coefficient on fraudulent loans in a zip code is consistent across the period; the coefficient first becomes significant in April 2020 and persists until June 2022. After June 2022, areas with higher PPP fraud have lower house price growth, suggesting that the inflationary impact of pandemic fraud on house prices was temporary. Non-fraudulent PPP lending has no effect on house prices, consistent with these funds offsetting legitimate expenses and lost revenues as opposed to providing extra stimulus.

This effect is large relative to other proposed factors explaining house price growth during the COVID period, and a horse race between PPP fraud and other factors shows that

[3]Zip codes with high and low suspicious lending per capita are those in the top and bottom decile, respectively.

PPP fraud is robust to controlling for all of these other factors. Considering all proposed factors together using a common framework is important for identifying which factors are most important and how they may relate to one another.[4] Under both Bayesian Model Averaging and variable selection using the Bayesian Information Criteria, PPP fraud consistently emerges as one of the strongest predictors of house price growth, along with land unavailability. Teleworking, previous (2018–2019) housing price growth, net migration, and remote work are also consistently selected predictors but with smaller economic magnitudes.

A potential concern with our baseline analysis is that PPP fraud is not randomly assigned. Perhaps zip codes with high PPP fraud had pre-existing house price momentum or omitted characteristics related to house price appreciation during the COVID period. To account for these possibilities, we use four additional strategies. First, we use synthetic control method to create a control group with almost identical house price trends during 2018 and 2019. These price trends remain identical during the first months of 2020 and only diverge in the summer of 2020 when pandemic fraud is most likely to have started affecting house prices. Second, we use social connections to fraud in distant parts of the country as an instrument for fraud in a given zip code. The instrumental variables (IV) estimates are consistent with, and even somewhat larger than, our baseline estimates. Restriction to connections that are as far as 500 miles away reduces concerns about migration, local omitted variables, or regional shocks that could violate the exclusion restriction. Further, overidentification tests using connections in different distances bands imply that any effect that social proximity to fraud has on house price growth directly, or indirectly through omitted variables, must be the same over different distances. Third, in addition to controlling for demographic characteristics, we consider interactions between PPP fraud intensity and demographic characteristics. In particular, we estimate the effect of PPP fraud in zip codes with above- and below-median values of income, poverty rate, population density, minority population share, education, and pre-pandemic employment. The effect of PPP fraud on house prices is similar across

[4]For these comparisons, we adopt the frameworks that Griffin, Kruger, and Maturana (2020) use to assess drivers of house prices during the boom/bust cycle around the financial crisis.

all demographic splits, indicating that the results are not concentrated in a subset of zip codes. Finally, following the logic of Oster (2019), we find that the effect of fraud on house prices is similar with or without extensive control variables, which mitigates omitted variable concerns if the included control variable are at least partially informative about the effects of unobservables.

To examine consumer spending and inflation more generally, we analyze three different sources of data. First, zip codes with one standard deviation higher PPP fraud per capita have a highly statistically significant 2.38% increase in automobile title registrations from March 2020 to December 2021. Second, utilizing census tract-year-level data on consumer spending from Mastercard, we find that census tracts with one standard deviation higher PPP fraud per capita experienced a highly statistically significant 0.595 percentile rank increase in spending per capita in 2020 and 2021 compared to 2019 and then returned to normal in 2022. For both vehicle title registrations and consumer spending, we also perform numerous within-county splits by different demographic variables and find similar effects across these splits, indicating that the results are not driven by a subset of zip codes. Further, both of these analyses include detailed demographic control variables, control for overall PPP loan take-up, and include county $\times$ time fixed effects to control for cross-county COVID-19 policies. Third, CBSAs with high PPP fraud per capita experienced elevated inflation starting in late 2021 and persisting through at least April 2023.[5]

Our paper contributes to three main literatures. First, our findings highlight the importance of understanding potential externalities to fraud. One direct externality of financial fraud is decreased participation in the financial system (Guiso, Sapienza, and Zingales, 2008; Gurun, Stoffman, and Yonker, 2017). Fraud also has the potential to distort asset prices. For example, Akerlof and Romer (1993) describe how insolvent banks can gamble through fraudulent lending, which can create a boom and bust in asset prices, such as the observed

[5]Data on regional inflation is limited to just 23 CBSAs, but results are statistically significant even within this limited sample. Elevated inflation is primarily due to housing costs, but there is also some evidence of smaller inflationary effects on non-housing prices, including vehicles.

boom and bust in commercial real estate assets during the 1980s Savings and Loan Crisis. The rampant non-agency mortgage fraud from 2003 to 2006 distorted house prices and contributed heavily to the 2003 to 2006 boom and the 2007 to 2011 bust in house prices.[6] Kedia and Philippon (2007) show how fraudulent accounting distorts the use of economic resources across firms. To our knowledge, we are the first paper documenting economic spillovers from fraud in government programs.

Second, an emerging literature seeks to understand the forces that drove home price appreciation during the COVID period. Gupta et al. (2022) find that house prices and rents increased in areas farther from city centers and that the effect was stronger in MSAs with more remote workers. Suburban areas (Ramani and Bloom, 2021), higher proportions of stay-at-home residents (Gamber, Graham, and Yadav, 2023), the percent teleworking in the CBSA (Dingel and Neiman, 2020), remote worker share (Mondragon and Wieland, 2022; Davis, Ghent, and Gregory, 2024), less population density (Liu and Su, 2021), and higher economic impact payments (Lin, 2022) have also been linked to higher house price growth.[7] We extend this literature in two ways. First, we propose a new and strong channel of housing price growth during the COVID period that is independent of any of the other channels proposed in the literature. Second, we estimate detailed within-county zip code-level horse races between carefully constructed proxies from the literature. We are the first to synthesize and carefully compare these alternative explanations. We find that land unavailability and suspicious PPP lending have the strongest relation to house price increases. Previous levels of remote work, teleworkability, migration during 2020 and 2021, and prior housing price growth are also related to house price growth during this period but have quantitatively smaller effects, particularly when considered in multivariate regressions alongside proxies for

[6] Griffin, Kruger, and Maturana (2020) perform comparisons of measures proposed in the literature and find that excess credit from mortgage fraud (Griffin and Maturana, 2016) and excess subprime credit (Mian and Sufi, 2009, 2018) were the largest forces behind the 2003 to 2011 house price boom and bust cycle.

[7] Cherry et al. (2021) find that mortgage forbearance allowed borrowers to enter forbearance and miss \$86 billion in mortgage payments, which led to lower delinquency rates for home mortgages, auto loans, student loans, and other consumer debt. More generally, our paper adds to a literature showing that housing demand shocks can have large effects on house prices (e.g., Favilukis et al. 2012; Badarinza and Ramadorai 2018; Gorback and Keys 2023; Hartman-Glaser, Thibodeau, and Yoshida 2023; Aiello et al. 2023b).

other potential channels.

Third, we contribute to further understanding the general impact of COVID relief spending, as well as its relation to potential inflation. Chetty et al. (2023) find that the cost of each job saved by the PPP was $377,000, Autor et al. (2022) find a cost of $169,000 to $258,000 per job saved, Dalton (2023) finds that about 24% of PPP money went towards wage retention, and Cole (2024) finds a cost of $270,000 per job-year at small firms. Similarly, Granja et al. (2022) find small effects of the PPP on employment.[8] Diamond, Landvoigt, and Sanchez (2023) model and quantify the effects of fiscal and monetary stimulus on inflation and house prices and find that both can lead to inflation and large increases in house prices. De Soyres, Santacreu, and Young (2023) and Jorda and Nechio (2023) examine differences in fiscal spending across countries during 2020 and 2021 and show that countries with higher COVID-related fiscal spending, such as the U.S., are experiencing higher rates of post-COVID inflation.[9] We provide a detailed cross-sectional analysis within the U.S. on the relation between concentrated fraudulent PPP payments and local house prices and consumer spending.

Overall, our findings suggest that fraud in government programs can have unanticipated effects that are much broader and potentially larger than the fraudulent transfers themselves. These effects may have helped people that already owned houses but hurt individuals who bought houses at inflated prices. Consistent with this concern, our findings indicate that areas with large amounts of fraud are experiencing lower house price returns after June 2022 than they would have otherwise. To the extent that fraud can have large and unanticipated effects, more resources should be spent on fraud prevention in government program design and enforcement.

[8] In contrast, Faulkender, Jackman, and Miran (2021) find that the program was much more effective with an average cost per job saved of $34,000 to $37,000. A broader literature has also examined various aspects of differential access to the PPP (e.g., Denes, Lagaras, and Tsoutsoura 2021; Rabetti and Cong 2023; Bartik et al. 2020; Neilson, Humphries, and Ulyssea 2020).

[9] Aiello et al. (2023a) detail the characteristics of cryptocurrency adopters and find that the COVID-19 stimulus payments were in part invested in cryptocurrencies.

1. Data Sources

We use loan-level indicators of suspicious PPP loans developed by Griffin, Kruger, and Mahajan (2023a) aggregated to various geographic levels. These indicators are based on loan-level PPP data released on January 2, 2022 by the Small Business Administration (SBA). This dataset covers all PPP loans issued from the start of the program on April 3, 2020 through the end of the program on June 30, 2021 that had not been repaid or canceled as of January 2, 2022. At the loan-level, the data include business name, address, business type (e.g., corporation, LLC, self-employed, etc.), NAICS code (industry), loan amount, number of employees, date approved, loan draw (i.e., initial, first-draw loan or repeat, second draw loan), and lender for 11,469,801 loans originated by 4,809 different lenders with a total value of $793 billion. The primary suspicious loan indicators are nonregistered businesses, multiple loans at a residential address, abnormally high implied compensation relative to industry by CBSA averages, and large inconsistencies (as large as tenfold) between the jobs reported by borrowers on their PPP applications and jobs reported to the contemporaneous EIDL Advance program, which had a different incentive structure. In addition to these loan-level primary measures, we consider the extent to which PPP lending at the industry-county level exceeds the number of establishments listed for that industry-county pair in U.S. Census data. See Internet Appendix Section A for additional details regarding these measures. Griffin, Kruger, and Mahajan (2023a) extensively validate these measures, including with four secondary measures of fraud and three independent external measures. The findings of Griffin, Kruger, and Mahajan (2023a) are also validated by a detailed Congressional investigation of PPP fraud that focused on many of the same FinTech lenders flagged by Griffin, Kruger, and Mahajan (2023a) (see Congressional report here).

We use the Zillow House Value Index (ZHVI) at both the zip code and county levels, which estimates the typical value for homes in the 35th to 65th percentile and is adjusted for seasonality and compositional changes in sales over time.[10] Data on home purchases for

[10] As of January 2023, Zillow started using neural networks to generate Zestimates and the ZHVI. Addi-

a sample of individual PPP recipients are from LexisNexis and PropertyRadar. Data on address changes are from Melissa Data.

Additionally, we use data from a number of sources to replicate proposed house price drivers, including the percent of individuals who worked remotely in 2015–2019 and population density from the US Census American Community Survey (ACS),[11] teleworkability based on Dingel and Neiman (2020) and US Census LODES, net migration in 2020–2021 from a FOIA request made by Ramani and Bloom (2021) to the USPS, distance to the central business district from Ramani and Bloom (2021), house price growth in 2018–2019 from Zillow, and land unavailability from Lutz and Sands (2022). We also use county-level data on employment, spending, and small business revenue during the pandemic from the Economic Tracker by Opportunity Insights (described in Chetty et al. (2023)). Monthly vehicle title registration data from January 2018 to December 2022 for five states are from Cross-Sell, with additional data for Washington directly from the state. Annual data on vehicles per household are from the US Census ACS. Annual consumer spending data at the census tract level are from Mastercard's Center for Inclusive Growth. Bimonthly regional CPI data are from the Bureau of Labor Statistics. Housing market metrics at the zip code level are from Realtor.com. Demographic data at the zip code and county levels are from the US Census ACS and IRS Statistics of Income.

2. Background

2.1. Geographic Summary

In addition to being widespread, pandemic relief fraud was also highly concentrated geographically. For example, Griffin, Kruger, and Mahajan (2023b) find that whereas over

tionally, Zillow has provided ZHVI data back to 2012 based on their new neural network methodology; this version of our paper uses this data throughout. A previous version (December 4, 2023) of this paper used ZHVI data based on Zillow's previous model-based valuation methodology found similar results. Figure IA.1 compares monthly house price growth based on the neural network based and pre-2023 legacy ZHVI.

[11]The remote work measure may exclude individuals who only partially work remotely since the ACS survey question only asks about the previous week. Davis, Ghent, and Gregory (2024) use General Social Survey (GSS) data and find that workers that report usually working remotely constitute only about 15% of all workers that ever work remotely. Unfortunately, the GSS sample size of approximately 3,000 respondents every two years is too small to generate zip code-level measures of remote work.

30% of loans in Cook County, IL (Chicago) are suspicious, New York County and Los Angeles County both have suspicious loan rates under 10%. Suspicious loan rates are even more varied at the zip code level, often ranging from close to 0% to over 50%, even within the same county.[12] Griffin, Kruger, and Mahajan (2023b) also find strong geographic correlations between PPP fraud and suspicious activity in other pandemic relief programs, such as the EIDL and unemployment insurance. As a result, economic stimulus from pandemic fraud was highly concentrated geographically and had the potential to have distortive effects on local house prices and purchases of other assets. Additionally, the large amount of variation in suspicious lending within counties indicates the potential for powerful tests even when including county fixed effects to account for regional trends.

2.2. Magnitude of Pandemic Relief Fraud

PPP fraud is clearly geographically concentrated. Is it large enough to meaningfully impact local spending and prices of housing and other goods? Griffin, Kruger, and Mahajan (2023a) estimate that PPP fraud totaled approximately $117.3 billion, which is 14.8% of the program, and this estimate may be conservative given that the paper only uses publicly available data. Moreover, Griffin, Kruger, and Mahajan (2023b) find that fraud in other programs (specifically the EIDL program and unemployment insurance) is highly geographically correlated with PPP fraud. As a result, areas with high PPP fraud also had cash inflows from other pandemic relief fraud. For example, EIDL support to small businesses totaled more than $384 billion in the form of EIDL loans and EIDL Advance grants. A June 27, 2023 report by the Office of the Inspector General (OIG) of the SBA indicates that they have found over $136 billion in loans provided through the COVID-19 EIDL program, which represents 33% of total disbursed funds, to be potentially fraudulent.[13] Expanded

[12] Figure IA.2 replicates Figure 1 from Griffin, Kruger, and Mahajan (2023b) and shows the strong geographic clustering both across counties (Panel A) and across zip codes within counties (Panel B).

[13] The types of fraud indicators used by the OIG include $55.7 billion based on common or suspicious IP addresses; $34.2 billion due to various hold codes placed on loans by the SBA due to the loan being flagged; $20.5 billion based on duplicated or invalid Employer Identification Numbers (EINs); $31.7 billion based on bank accounts receiving multiple loans or individuals changing their bank account from the bank account listed on their application; $5 billion to sole proprietors or independent contractors without EINs; and the rest due to other indicators such as hotline complaints and suspicious phone numbers, physical addresses,

unemployment benefits during COVID amounted to $872.5 billion, and an audit of the UI programs in four large states by the OIG of the Department of Labor found that 20% of Pandemic Unemployment Assistance (PUA) funds were lost to fraud.[14] We may never know the precise magnitudes of pandemic fraud, but if similar fraud rates of 20% fraud hold for other programs, total pandemic relief fraud could be over $900 billion. Even if the rate is half as much, the total would still be economically substantial.

To put pandemic relief fraud in context relative to the U.S. housing market, the total value of all homes sold in the U.S. was approximately $2.2 trillion in 2020 and $2.8 trillion in 2021, but the median home only had a 17% down payment for repeat buyers and 7% down payment for new buyers.[15] At a 15% down payment rate, this would amount to approximately $750 billion in down payments on housing in 2020 and 2021. Thus, if even a small percentage of pandemic fraud was used to purchase houses, it could have a meaningful impact on the housing market.

Another way to gauge the magnitude of the house demand shock is that a one standard deviation shock (6.8 fraudulent loans per thousand people) amounts to approximately $5 million of additional PPP fraud for a typical zip code.[16] For context, the house purchases in a typical zip code in 2020 and 2021 totaled around $120 million.[17] If average down payments are around 15%, this equates to aggregate down payments of $18 million in a typical zip code. Thus, even a small share of $5 million of incremental PPP fraud could have a significant impact. And total pandemic relief fraud in these zip codes is likely even higher given that EIDL and unemployment insurance fraud are highly correlated with PPP fraud.

and email addresses. The OIG report also separately identified $64 billion in potentially fraudulent PPP loans, which represents 8% of total disbursed funds. See report here.

[14] See report here. An analysis by the US Government Accountability Office (GAO) estimates the amount of fraud in pandemic UI programs at between $100 billion and $135 billion (see report here).

[15] Source and source.

[16] The average flagged loan is for $45,293 and the population in an average zip code is 15,881 people. Thus, $0.0068 \times 15{,}881 \times \$45{,}293 = \$4.9$ million.

[17] The average zip code has 4,474 single family housing units that were worth $269,114 in December 2019. Thus, the housing stock in the average zip code was worth $1.2 billion. A turnover of 5% per year implies that the flow of house purchases during 2020 and 2021 was $120 million in the average zip code. The estimated turnover of 5% is based on annual U.S. home sales of approximately 6.4 million units (e.g., see here) and a housing stock of approximately 141 million units (e.g., see here).

3. Did Fraudulent PPP Recipients Buy Houses?

There are two reasons to think that fraudulent funds might lead to different spending patterns than non-fraudulent funds. First, the design of pandemic relief programs was to offset lost revenue. To the extent that the programs followed the intended design, receipts received no income shock and spending practices should not have shifted. However, if someone committed PPP fraud and received a sudden windfall of funds, this could have been a large increase in their income and overall wealth. Second, individuals who were willing to take the risk to commit fraud might be more likely to spend their ill-gotten money on immediate consumption items rather than simply saving the funds for the future. Anecdotal evidence suggests that many recipients of fraudulent PPP loans used the funds they received to purchase expensive houses, cars, and luxury items.[18] We first examine house purchases using property records from PropertyRadar for a random sample of 250,000 individual PPP borrowers. The sample was collected in February 2023 and consists of individual borrowers who received PPP loans during all three rounds of the PPP with data on house purchases through the end of 2022. Round 3 of the PPP ended in June 2021, so we observe at least 18 months of post-PPP house purchase activity for all individuals in the sample. We match names purchasing houses in the PropertyRadar data to names of PPP borrowers, limiting the sample to names that are unique.[19]

Figure 1 plots monthly house purchase rates for PPP loan recipients before and after receiving a PPP loan in event time relative to the date that the PPP loan was approved. Flagged and non-flagged PPP borrowers follow parallel trends before receiving PPP loans, with an average monthly home purchase rate of 0.46% for both flagged and non-flagged PPP recipients. After receiving PPP loans, this purchase rate increased by 6.3 bps (a 14% increase relative to the pre-period house purchase rate) for flagged recipients and remained

[18] For example, here and here.

[19] We determine unique names by using voter rolls from nine states that collectively represent 27% of the US population, and we define a name as unique if it occurs only once across all of the states. While this methodology may add some noise relative to LexisNexis individual-level data (described below), estimates are similar, and individuals in both samples have the same house purchase data 85.3% of the time.

about the same for non-flagged recipients.

Table 1 estimates difference-in-differences regressions. The regression uses monthly observations for each PPP recipient in the sample, and the dependent variable is an indicator variable for whether the PPP recipient purchased a house in that month. The sample starts five years before each PPP loan was approved and ends 18 months after. *Post* is an indicator that takes the value of one starting in the month that the PPP recipient's loan was approved, which can be as early as April 2022 or as late as June 2023, depending on the recipient.

In column (1), the regression controls for loan fixed effects and month of year fixed effects. The main coefficient of interest is an indicator for flagged PPP recipients interacted with the *Post* indicator. All coefficients are multiplied by twelve to represent annual effects. Consistent with Figure 1, the house purchase probability for flagged PPP recipients increases by 0.86 ppt more compared to non-flagged PPP recipients. Non-flagged PPP recipients experience an increase in house purchase probability as indicated by the coefficient of 0.69 ppt on *Post*. The *Flagged* $\times$ *Post* coefficient of 0.86 ppt is large relative to the average annual house purchase rate of 5.01%, reflecting a 17% relative increase in house purchases. Column (2) adds *Post* $\times$ county fixed effects, *Post* $\times$ business type fixed effects, *Post* $\times$ week approved effects, and log loan amount interacted with *Post*. The *Flagged* $\times$ *Post* coefficient remains significant with an unchanged estimate of 0.86 ppt. Consistent with Griffin, Kruger, and Mahajan's (2023a) finding that PPP fraud was concentrated in FinTech loans, we find that the effects are concentrated in FinTech loans, and non-flagged FinTech loans are also associated with increased home purchase rates (see Table IA.1) consistent with these loans also exhibiting high rates of fraud.[20]

As additional examination of house price purchases with a different sample, we examine house purchases using detailed property records from LexisNexis for a random sample of 150,000 individual PPP borrowers used by Griffin, Kruger, and Mahajan (2023a) for their

[20] Griffin, Kruger, and Mahajan (2023a) find that suspicious loan flags are a stronger signal of potential fraud for FinTech loans and through a variety of secondary indicators find that even non-flagged FinTech lender loans contain high rates of fraud.

felony analysis.[21] Columns (3) and (4) report the results, which are essentially the same as the baseline estimates in columns (1) and (2). Monthly home purchase results are also similar for the LexisNexis sample (see Figure IA.3). The advantage of the baseline sample is that it includes a larger sample of PPP borrowers from all three rounds and a longer post-PPP period. The LexisNexis sample is a subset of borrowers and only includes borrowers with loans in rounds 1 and 2 (i.e., in 2020), but this sample has the advantage that it is matched to LexisNexis based on both name and address and uses LexisNexis data on home purchases at the individual level. We also consider an alternative data source on moving based on change of address data from Melissa Data for the same subsample of 150,000 borrowers with equivalent results indicating that suspicious PPP borrowers were more likely to move after receiving a PPP loan compared to recipients of non-flagged PPP laons.[22] Overall, the evidence indicates that fraudulent loans stimulated house purchases.

4. Does PPP Fraud Predict House Price Growth?

Because of the geographic clustering of PPP fraud and geographically correlated fraud in other pandemic relief programs, house purchases by recipients of fraudulent pandemic relief have the potential to distort local house prices. While fraudulent funds could have been used in many ways, the level of fraud combined with its high geographic concentration makes it possible that fraud could have distorted local house prices. We examine whether PPP fraud levels had a discernable effect on zip code-level house prices after controlling for other factors.

In this section, we examine the relation between zip-code level house prices and PPP fraud, as well as other potential factors. To control for macro factors that may influence regional house prices, our analysis is focused on the zip code level with county fixed effects.

[21] The sample consists of individual borrowers who received PPP loans in rounds 1 and 2. The LexisNexis data was collected in March 2021 and includes data on house purchases through the end of 2020. Rounds 1 and 2 of the PPP occurred in April to August 2020, with most loans occurring by the end of May. As a result, we observe at least four months of post-PPP house purchase activity for all individuals in the sample.

[22] See Figure IA.4 and Table IA.2. Table IA.3 shows that the moves by individuals with flagged loans result in larger positive changes in neighborhood quality as compared to moves by individuals with non-flagged loans.

County fixed effects are used in most of our analyses since counties differed dramatically in their COVID policies and county fixed effects can also capture broader regional trends. First, we regress zip code-level house price growth on several measures of potentially fraudulent lending while controlling for demographic variables. Second, we use matching and synthetic control methodologies to generate counterfactuals for zip codes with similar house price trends. Third, we use fraud rates in distant parts of the country that are socially connected to a given zip code as an instrument for local PPP fraud rates. Fourth, we examine and control for variables associated with house price growth in previous research, including measures specific to potential channels influencing house price growth during the COVID pandemic. Finally, we use variable selection procedures to compare variables.

4.1. Weighted Least Squares Regressions

To understand the potential relation between suspicious lending and house prices across zip codes within counties, we regress house price growth on flagged PPP loans per capita with county fixed effects. The regressions are weighted by the zip code's 2019 population to ensure our estimates are nationally representative.[23] Table 2 shows the relation between measures of suspicious lending and house price growth. The measures of suspicious lending are standardized to have a mean of zero and a standard deviation of one, so the coefficients represent the effects of a one standard deviation increase in suspicious lending rates.[24] In addition to county fixed effects, the regressions also control for house price growth between January 2018 and December 2019, PPP loans per capita, population density, housing vacancy rates, number of housing units, and average household income as indicated in the table. Previous house price growth and PPP loans per capita are controlled for non-parametrically using percentile fixed effects, which allow for non-linear relations.[25] The results are also similar when non-flagged loans per capita is used as a control instead of total loans per

[23]Results are similar when ordinary least square (OLS) is used (Table IA.4).
[24]Standardization throughout the paper is based on standard deviations weighted by population.
[25]Tables IA.5 and IA.6 show that the results are also robust to controlling for previous house price growth and loans per capita linearly or with higher order polynomials.

capita (as shown in Table IA.7).

Column (1) estimates a regression of house price growth on flagged PPP loans per capita at the zip code level with county fixed effects and no other control variables. The resulting coefficient of 0.0228 indicates that, on average, zip codes with one standard deviation higher suspicious PPP lending per capita experiences house price growth in 2020 and 2021 that was 1.79 ppt higher. Column (2) reports our baseline estimate, with control variables added for past house price growth, overall PPP lending per capita, and zip code demographic variables. Including these control variables modestly increases the coefficient estimate. A one standard deviation increase in *Flagged Per Capita* is associated with a 2.11 ppt increase in house prices between January 2020 and December 2021. This is a sizable 8.2% of the 25.9 ppt average increase in house prices during this period.[26] The relation between flagged loans and house price growth is concentrated in FinTech loans, consistent with more house purchasing by this group in the previous section and Griffin, Kruger, and Mahajan's (2023a) finding that fraud was more pronounced in FinTech loans and that suspicious loan flags are a stronger predictor of fraud for FinTech loans (see Table IA.8).

Griffin, Kruger, and Mahajan (2023a) show that excess PPP loans relative to establishments and highly similar loans (with nearly identical loan features) are common and highly correlate with other suspicious loan indicators.[27] Results based on these alternative measures in columns (3) and (4) are slightly stronger than the baseline results shown in column (2). Finally, in column (5), we consider a composite measure that is based on a combination of *Flagged Per Capita*, *High Loan-to-Establishment Per Capita*, and *High Similarity Per Capita*. Specifically, it is the ratio of the number of PPP loans that fit the criteria for any

[26]The 2.11 ppt coefficient estimate is also large relative to the standard deviation of house price growth across zip codes during this period, which is 10.5 ppt.

[27]The first measure is the ratio of the number of business PPP loans in the zip code that are in county-industry pairs with excess loans, defined as county-industry pairs with more than twice as many business PPP loans as establishments recorded in U.S. Census County Business Patterns (CBP) data, to population. Establishment counts in the CBP data are at the county-industry pair level. This analysis is restricted to C-corporation, S-corporation, LLC, and sole-proprietorship loans because self-employed and independent contractors are not included in the CBP data. See Griffin, Kruger, and Mahajan (2023a) for more details on the construction of these measures.

of these three measures to population, which we call *Flagged Composite Per Capita*. This more comprehensive measure indicates an even larger impact of suspicious lending on house prices. A one standard deviation increase in *Flagged Composite Per Capita* is associated with a 2.68 ppt increase in house prices, which is 10.4% of the average increase in house prices.

The results in Table 2 are robust to propensity score weighting to control for previous house price growth (Table IA.9), alternative suspicious lending measures focused on the dollar value of suspicious loans and the percent of loans that are suspicious (Table IA.10), alternative fixed effects (Table IA.11), using within-CBSA variation instead of within-county (Table IA.12), controlling for pre-pandemic house price levels (Table IA.13), controlling for COVID mortgage forbearance (Table IA.14), alternative standard error clustering (Table IA.15), restrictions to different subsets of loans (Table IA.16), county-level analysis instead of zip code-level analysis (Table IA.17), restrictions to different subsets of states (Figure IA.5), and alternative house price data including the pre-2023 non-neural-network version of the Zillow Home Value Index (Table IA.18) and house price data from Realtor.com (Table IA.19). Controlling for economic impact payments (i.e., stimulus checks) does not affect results (Table IA.20).[28] Houses in zip codes with more PPP fraud are on the market for fewer days, receive more viewers, and have higher Realtor.com Market Hotness Index value (Figure IA.7). Results using rent growth are positive but smaller than house price growth (Table IA.21). There is no evidence of heterogeneity based on political leanings or exposure to COVID (Table IA.22). In contrast to the predictiveness of suspicious PPP loans, overall PPP lending rates are, if anything, negatively related to house price growth (Table IA.10), which is consistent with legitimate PPP funds being used to replace lost business revenues rather than creating a windfall for owners.

[28] In Figure IA.6, we permutate the suspicious lending measure between zip codes, both within the same county and across the nation, and show that the coefficients on the suspicious lending measures are much smaller in all 10,000 permutations performed (the largest being less than 20% of the true coefficient); this implies the result we found is extremely unlikely to happen by chance.

To graphically illustrate the relation between flagged loans per capita and house price growth between January 2020 and December 2021, the left (right) subpanel of Figure 2, Panel A shows the relations between *Flagged Per Capita* (*Flagged Composite Per Capita*) and house price growth using binscatters. These binscatters include county fixed effects and the same controls used in the regressions reported in Table 2.[29] The relation between suspicious lending and house price growth from January 2020 to December 2021 is close to linear, with a 5.7 ppt difference in house price growth between zip codes in the lowest decile of fraud rates and those in the highest decile.

To further understand the timeseries dynamics of the relation between suspicious lending and house prices, we re-estimate specification (1) of Table 2 for the cumulative house price growth from January 2020 to each subsequent month. The coefficients on *Flagged Per Capita* for each month are reported, with a corresponding 95% confidence interval, in the top subpanel of Figure 2, Panel B. The coefficient first becomes significantly positive in April 2020 and generally strengthens each month until it peaks in June 2022. After June 2022, the effect begins to decrease; as of March 2024, the effect is half the size as it was at its peak. The bottom subpanel of Figure 2, Panel B shows that the cumulative effect of pandemic fraud on house prices over time follow a similar pattern for the broader composite measure of suspicious lending, *Flagged Composite Per Capita*. The general trends in the coefficient match reasonably well with the timing of the PPP and other relief programs that started earlier or ended later, such as the EIDL program and expanded unemployment benefits.[30]

To assess whether results are concentrated in zip codes with particular demographics, we add interactions between PPP fraud and indicator variables for whether the zip code is above or below the national median of different demographic characteristics (also controlling for the demographic indicator variables themselves). Panel C of Figure 2 shows the results.

[29]The binscatters throughout the paper are based on Cattaneo et al. (2024), which highlights and corrects issues related to covariate adjustment in traditional binscatters. The binscatters are weighted by each zip code's 2019 population.

[30]Figure IA.8 shows the monthly, non-cumulative effects.

The first column of Panel C plots separate PPP fraud coefficients for low- and high-income zip codes. Results are almost identical and remain large and highly significant for both subsets of zip codes. The same is true for zip code subsets based on poverty rates, population density, minority population share, educational attainment, and pre-pandemic employment. Effects are also strong across sample splits by specific race and ethnicities and when racially or ethnically homogenous zip codes are excluded (see Table IA.23).[31] Even when coefficients differ a bit, the differences are not statistically significant. Consistent results across diverse zip codes point to a broad-based effect of PPP fraud as opposed to an effect that is concentrated in a particular demographic group. This is reassuring and may alleviate some concerns about omitted variables and non-random assignment of PPP fraud.

4.2. Additional Analysis

One potential concern with our house price regressions is that PPP fraud is correlated with pre-existing house price trends to some extent.[32] We control for these trends in our baseline regressions with flexible percentiles fixed effects for historical house price growth. In this section, we consider the synthetic control methodology to control for potential house price momentum.

We use synthetic controls based on Abadie, Diamond, and Hainmueller (2014) to construct a control group. The treated group is zip codes in the top quartile of the *Flagged Per Capita* measure within each county. The sample is limited to counties where the difference between the top and bottom quartile of *Flagged Per Capita* within the county is at least half the national standard deviation. This requirement is met by 283 counties, which collectively represent approximately 25% of the US population. For each treated zip code, we develop a synthetic control using zip codes in the same county that are in the bottom quartile of

[31] All demographic variables are from the US Census American Community Survey. Poverty rate is the percentage of households with income below the poverty threshold, which varies based on family size and composition. Educational attainment is the percentage of adults with at least a bachelor's degree. Minority population share is the percentage of non-white individuals.

[32] To illustrate the potential concern, Figure IA.9 plots average house prices (indexed to January 2020) in zip codes in the top and bottom quartile of flagged PPP loans per capita within county where the difference between the top and bottom quartile of fraud rates within the county is at least half the national standard deviation.

the *Flagged Per Capita* measure by finding weights such that the squared error in monthly house price growth from January 2018 to December 2019 between the synthetic control and the treated zip code is minimized.

In Figure 3, we plot average house price growth for top and synthetic bottom quartile PPP fraud zip codes. The synthetic controls methodology is designed to generate treatment and control groups with similar overall house price changes from January 2018 to December 2019, which the plots show is clearly true. After January 2020, the treatment and control groups continue to follow similar trends for several months. There is nothing mechanical about this result. The identical trends during these months indicate that the methodologies successfully create control groups with similar price trends. There is also no evidence of any differential impact of COVID during its early development in March and April of 2020. By contrast, treatment and control groups start to differ significantly starting in July 2020. This is around the time we would expect pandemic relief fraud to have an impact since the PPP and other programs ramped up in April and it likely takes a few months to search for and purchase a house. Zip codes in the top quartile of PPP fraud had an average of 28.3% house price growth in 2020 and 2021, compared to an average of 22.8% growth for bottom quartile fraud synthetic control zip codes. The difference of 5.5 ppt is economically large and highly statistically significant, as can be seen from the 95% confidence intervals plotted as dotted lines. Figure IA.10 show similar results based on the *Flagged Composite Per Capita* measure and several other measures and also show that PPP lending in general did not affect house prices.

An alternative to the synthetic control methodology is to directly match high-fraud treated zip codes to low-fraud control zip codes with similar house price growth during 2018 and 2019. For this analysis we match zip codes within CBSAs instead of counties because frequently there are not enough close matches within counties. We also restrict the sample to CBSAs with at least ten zip codes. In a previous version of this paper using legacy data from Zillow (before their adoption of a new neural network methodology), there was clear separation between matched high and low fraud zip codes that is similar to the synthetic controls results. After switching to the new Zillow data, there continues to be a positive difference between house price growth in matched high and low fraud zip codes, but the difference is much smaller than before (see Figure IA.11 for results with both the legacy data and the current Zillow data). All other results are virtually identical after updating the data (e.g., see Table IA.18), but we note the discrepancy in this result.

A second concern is that PPP fraud may correlate with other characteristics related to house price growth during 2020 and 2021. For example, if PPP fraud is concentrated in areas with strong economic growth during this period, the apparent relation between fraud and house price growth could be due to economic fundamentals as opposed to demand due to fraud proceeds. Our baseline specifications address this concern in several ways. First, all analysis includes county fixed effects, which should absorb most differences in economic growth and job opportunities across different areas. Second, the baseline regressions include extensive control variables, including PPP loans per capita, past house price growth, population density, housing vacancy rates, number of housing units, and average household income, and results are robust to different control variable and fixed effect specifications (see Table IA.11). The stability of the effect of fraud on house prices with and without control variables mitigates potential concerns about omitted variables under the logic of Oster (2019).[33] Finally, the lack of heterogeneity in the effect of fraud on house prices across numerous sample splits based on demographic and economic characteristics implies that any potential omitted variables would need to affect a diverse set of zip codes in a similar manner.

[33] In addition to estimating regressions with and without control variables, Table IA.11 also controls for additional potential channels that are discussed in the next subsection. Using the approach proposed by Oster (2019), if omitted variables have the same proportional impact on the flagged per capita coefficient as observed control variables (equal selection assumption) and could potentially increase the R^2 of the regression from 0.831 to 1.0, the coefficient on flagged per capita in column (4) would only decrease from 0.0175 to 0.0170 $[0.0170 = 0.0175 - (1 - 0.831) \times (0.0177 - 0.0175)/(0.831 - 0.763)]$. Relaxing the equal selection assumption and maintaining the assumption that omitted variables could increase the R^2 to 1.0, other omitted variables would need to have approximately 35.2 times the proportional impact of the observed control variables to decrease the coefficient to zero. It is worth noting that these controls include drivers of house price growth that have already been documented by the literature, which should arguably have the largest impact and thus make it even more unlikely for unobserved variables to explain the effect.

To further address at least some of the potential omitted variable concerns, we instrument for PPP fraud using fraud in geographically distant but socially connected zip codes. This identification strategy builds on Griffin, Kruger, and Mahajan (2023b), which shows that the spread of fraud is strongly related to social connections and a zip code's fraud rate is strongly predicted by fraud rates in other zip codes with which the zip code has strong social connectedness. The specific instrument we consider is the average (weighted by social connection strength) *Flagged Per Capita* in other zip codes that a zip code is connected to, where social connectedness of zip codes i and j is measured as the Facebook friendships between users in zip code i and zip code j scaled by the product of Facebook users in zip code i and zip code j, using data from Bailey et al. (2018b, 2020). Additional details on the construction of the instrument are discussed in Internet Appendix A. While social connections to other zip codes are endogenous, this identification strategy has the benefit of isolating variation related to distant social connections as opposed to anything that might jointly influence housing markets and PPP fraud at the local zip code level.

Table 3 shows the relation between suspicious lending rates and house price growth based on the instrumental variable described above. As in our WLS estimates, the measures of suspicious lending are standardized to have a mean of zero and a standard deviation of one, so the coefficients represent the effect of a one standard deviation increase in suspicious lending rates. In addition to county fixed effects, the regressions also control for house price growth in 2018 and 2019, PPP loans per capita, population density, housing vacancy rates, number of housing units, average household income, and the share of friends of Facebook users in the zip code who live within 50 and 150 miles of the zip code. Both the previous house price growth and PPP loans per capita are controlled for non-parametrically using percentile fixed effects, which allows for non-linear relations. The estimates are weighted by the zip code's 2019 population to ensure our estimates are nationally representative. In column (1), we instrument for a zip code's *Flagged Per Capita* using social connections outside of the CBSA where the zip code is located.[34] A one standard deviation increase in instrumented *Flagged Per Capita* is associated with a 3.47 ppt increase in house prices from January 2020 to December 2021. This is a sizable 13.5% of the 25.7 ppt average increase in house prices during this period.

To the extent that distant social connections are less likely to affect house price growth through, for example, migration, local omitted variables, or regional shocks, social proximity to fraud based on only distant zip codes may be more likely to meet the exclusion restriction.[35] To examine this, columns (2), (3), and (4) of Table 3 instrument for a zip code's *Flagged Per Capita* using social connections to zip codes that are at least 100, 250, and 500 miles away, respectively. The results in all four specifications are extremely similar (between 3.36 and 3.52 ppt, with t-statistics above 5.4 and first stage F-statistics of at least 28.5) and cannot be statistically distinguished from one another. In column (5), we include multiple instruments based on social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles). The estimate is again nearly identical, and including multiple instruments allows us to conduct a J-test of overidentifying restrictions. The J-statistic of 2.464 with a p-value of 0.292 indicates that estimates are consistent regardless of which subset of the instruments is used. This implies that any effect social proximity to fraud has on house price growth directly, or indirectly through omitted variables, must be the same over different distances.[36]

Social connections have been shown to affect other economic outcomes and in particular, house price expectations. Bailey et al. (2018a,c) show that individuals whose friends experience house price increases have heightened house price expectations and are more likely to purchase a home and choose lower mortgage leverage. To examine the potential effects

[34] Only zip codes that are located in a CBSA are included in this specification.

[35] This is similar to the logic provided by Bailey et al. (2018a,c) for using house price experiences of out-of-commuting-zone friends as an instrument for house price experiences of all friends. Hu (2021) uses floods in distant but socially connected counties to study flood insurance purchases since distant flooding events are likely orthogonal to an individual's own local flood risk.

[36] As one example of the direct effects of social connections on house prices varies by distance, Table IA.24 shows that the effect of social connections on migration rates is decreasing in distance.

of house price expectations being transmitted through social connections, we construct a measure of social proximity to house price growth in a manner analogous to social proximity to suspicious lending. After controlling for social proximity to house price growth, social proximity to fraud continues to have a strong effect on house price growth (see Table IA.25). This indicates that it is unlikely that social proximity to fraud is capturing house price expectations being transmitted through social connections. It is also worth noting that the effect of social proximity to house price growth falls significantly with distance. This contrasts with the consistent results across different distances that we find for social proximity to fraud and supports the aforementioned logic for using distant social connections and testing overidentifying restrictions.[37]

4.3. Other Proposed Factors Affecting Housing Prices

A growing literature proposes a number of factors that potentially affected house prices during the COVID period. We construct measures for these factors following the literature as closely as possible. The factors considered include prior remote work from 2015 to 2019 (Mondragon and Wieland, 2022; Davis, Ghent, and Gregory, 2024), the percent teleworking in the CBSA prior to the crisis (Dingel and Neiman, 2020), population density (Liu and Su, 2021), net migration during 2020 and 2021 (Ramani and Bloom, 2021), distance to the central business district (Gupta et al., 2022), previous (2018–2019) house price growth, and land unavailability (Lutz and Sands, 2022).[38] All dependent variables are standardized to have a standard deviation of one to allow for easier comparisons of the economic magnitude of the coefficients. We include county fixed effects, past house price growth, PPP loans per capita, vacancy rate, housing units, and average household income as control variables in all regressions. We are able to estimate the specification for a large cross-section of 12,305 zip codes for which we have data for all of the proposed factors.

[37] Additional IV analysis and robustness tests are discussed in Internet Appendix B.

[38] Lin (2022) finds that MSAs with higher economic impact payments (i.e., stimulus payments) and child tax credits per capita experienced higher house price growth. We do not include this channel in our baseline analysis because we do not find support for a positive relation between economic impact payments and house price growth at the zip code level using the standardized regression framework described below (Table IA.20). Controlling for economic impact payments does not affect results for PPP fraud.

The univariate regressions for each of the proposed factors are shown in Table 4. All of the factors are statistically significant, but with varying economic magnitudes.[39] In the multivariate regression (column (9)), the effects of most of the factors remain statistically significant but are attenuated relative to the univariate regressions to varying degrees. The coefficient on *Flagged Per Capita* remains statistically and economically significant, with a coefficient that is close to the univariate coefficient. The coefficient on population density becomes insignificant. To better visualize the relations, Panel A of Figure 4 shows the univariate and multivariate coefficients with 95% confidence intervals. *Flagged Per Capita*, land unavailability, and house price growth in 2018-19 have the largest coefficients at slightly over 2 ppt of house price growth per standard deviation. Teleworking, 2020–2021 migration, prior remote working, and distance to central business district are all also statistically significant in the multivariate specifications, though with considerably smaller coefficients.[40]

We consider alternative specifications, including averaging each of the factors over a 5-mile radius, only including county fixed effects, and using OLS instead of WLS (see Figure IA.12). We also examine heterogeneity in the effects across splits based on land unavailability, previous house price growth, pre-COVID house prices, COVID mortgage forbearances, zip code-level beta with national house prices during 2000–2019, and zip code-level house price volatility during 2000–2019 (see Figure IA.13). Across all of these variations, the effect of suspicious lending on house prices is similar to the baseline results. To understand the effects of each factor on house prices over time, we perform the same monthly analysis as Panel B of Figure 2 for each of the other proposed factors (see Figures IA.14 and IA.15).

Correlation between the factors could complicate the multivariate estimates. To more formally assess which factors most robustly predict house price growth, we apply the Bayesian Model Averaging approach to model selection as suggested by Fernández, Ley, and Steel

[39]The univariate relationships are similar to those documented in the existing literature.

[40]Table IA.26 considers alternative measures of migrations, such as migration during 2020 and 2021 separately and the decomposing net migration into inflows and outflows. The effect of *Flagged Per Capita* is robust across all these variations.

(2001) and Ley and Steel (2009) and following Griffin, Kruger, and Maturana (2020), which examined the determinants of 2003 to 2012 house price growth. Following the assumptions recommended by Ley and Steel (2009) for modeling and prior distributions, the procedure estimates posterior distributions for the probability that a variable is included in a model and the coefficient conditional on inclusion. Posterior coefficient distributions conditional on inclusion in the model are plotted in Panel B of Figure 4. The probability of inclusion in the model is shown by the bars above each of the distributions.[41] The procedure always includes *Flagged Per Capita*, land unavailability, 2018–2019 housing price growth, teleworking, net migration, remote work, and distance to CBD. Population density is only selected in 0.9% of models. Conditional on inclusion, the coefficients are furthest from zero for house price growth in 2018-19, *Flagged Per Capita*, land unavailability, and teleworking, in that order.

Table 5 reports optimal model selection based on the Bayesian Information Criteria. The column number corresponds to the best model if the model is limited to that number of proposed factors (between one and eight). If the model is restricted to one factor, *Flagged Per Capita* is included. Further, *Flagged Per Capita* is consistently included when the model is restricted to any number of factors. The optimal model, across any number of the proposed factors according to the Bayesian Information Criteria, includes seven out of eight of the factors (all but population density).

Overall, the evidence supports many of the proposed factors and shows that regardless of which channels are considered, pandemic fraud continues to be a strong predictor of house price growth during 2020 and 2021. The relation between fraud and house price growth has a magnitude that is at least as high as any other proposed factor, is robust to controlling for any combination of other factors, and has timing that lines up more clearly with the rise in house prices than most other factors.

[41] Column 9 of Table 5 also reports the posterior inclusion probabilities for each proposed factor.

5. Other Spending and Inflation Effects

The stimulus effects of PPP fraud are conceptually not limited to the housing market. In particular, anecdotal evidence suggests that fraudulent PPP recipients also frequently purchased cars and luxury products.[42] In this section, we start by examining the relation between PPP fraud and vehicle purchases at the zip code level. We next examine PPP fraud and general consumer spending at the census tract level. We then close by analyzing how PPP fraud relates to regional differences in inflation.

5.1. Vehicle Purchases

To investigate vehicle purchases, we use monthly data from January 2018 to December 2022 on vehicle title registrations at the zip code level for five large states (California, Texas, Florida, Illinois, and Ohio) from Cross-Sell, supplemented by similar publicly available data for the state of Washington.[43] If recipients of fraudulent PPP loans used their fraud proceeds in part to purchase vehicles, we would expect increased vehicle title registrations in high-fraud zip codes after the start of the PPP. To examine whether this is the case, we estimate a regression of the form:

$$\begin{aligned} Log(VehicleTitleRegistration)_{it} = &\textstyle\sum_{t \neq Feb2020} \beta_t (Month_t \times FlaggedPerCapita_i) \\ &+ ZipCode_i + County_i \times Month_t + \epsilon_{it} \end{aligned}$$

The dependent variable in the regression is the log of the number of vehicle title registrations in zip code i during month t. *Flagged Per Capita* is standardized so that one unit represents one standard deviation. The coefficients of interest are β_t, which estimate effect on vehicle title registrations, relative to the February 2020 baseline, that is associated with a one standard deviation increase in *Flagged Per Capita*. The regressions include zip code fixed effects and county × month fixed effects to isolate differential changes within counties. Standard errors are double clustered by county and month.

[42]For example, here and here.
[43]These six states collectively represent 36.6% of the US population.

The left plot in Panel A of Figure 5 shows the effect of a one standard deviation increase in *Flagged Per Capita* on vehicle title registrations over time. During the pre-COVID period, the effect is on average close to zero (-0.19%), albeit with some months with a positive effect and others with is a negative effect.[44] After the onset of COVID, a one standard deviation increase in *Flagged Per Capita* is associated with a 1.44% in vehicle title registrations over the March 2020 to December 2022 time period. This effect is largely concentrated from March 2020 to December 2021, when a one standard deviation increase in *Flagged Per Capita* is associated with a 2.10% increase in vehicle title registrations. On the other hand, a one standard deviation increase in *Flagged Per Capita* is associated with a 0.23% increase in vehicle title registrations during the January 2022 to December 2022 time period. This is exactly what we would expect from short-term stimulus to vehicle purchases during the PPP in 2020 and 2021.[45]

The right plot in Panel B of Figure 5 shows a binned scatter plot across zip codes of the percentage change in vehicle title registrations versus *Flagged Per Capita*, controlling for county fixed effects. The percentage change in vehicle title registrations on the vertical axis plots the percentage change in the number of vehicle title registrations from the 2018–2019 time period to the 2020–2021 time period. Consistent with the regression results in the left panel, there is a positive relation between the percentage change in vehicle title registrations and *Flagged Per Capita*.[46]

Table 6 collapses the β_t coefficients in Equation 5.1 into a single coefficient for the interaction between *Flagged Per Capita* and an indicator variable for the time period starting in March 2020. Columns (1) to (3) use data from January 2018 to December 2021, and with *Post* define as 1 for March 2020 to December 2021 and 0 otherwise. Standard errors

[44] Because the dependent variable is $log(VehicleTitleRegistration)$, we interpret the coefficients as percentage changes based on the log approximation.

[45] Figure IA.16. Panel A shows the relative vehicle title registration rates across terciles of *Flagged Per Capita*.

[46] Note that vehicle title registrations declined by an average of 9.8% between these periods. This relation is highly statistically significant with a *t*-statistic of 6.25 based on standard errors clustered by county. Figure IA.16, Panel B shows that the results are similar for percentage change in registrations from 2018–2019 to 2020 on its own, as well as a longer post-COVID time period including all of 2020–2022.

are again clustered by county and month, and zip code and county × month fixed effects are included. Column (1) estimates a regression that is similar to the regressions plotted in Figure 5, Panel A. A one standard deviation increase in *Flagged Per Capita* is associated with a 2.28% increase in registrations. In column (2), we add post dummy interacted with total PPP loans per capita. This specification distinguishes the stimulative effect of fraudulent PPP loans from any stimulative effect of PPP loans more generally. The coefficient on *Flagged Per Capita* increases to 3.09% and remains highly statistically significant. A potential concern is that PPP fraud could be correlated with other characteristics that predict vehicle title registration growth during this time period. Column (3) adds detailed demographic data interacted with a post dummy variable to account for this concern. Adding these demographic variables decreases the *Flagged Per Capita* coefficient somewhat (from 3.09% to 2.38%), but the result remains large and highly statistically significant (with a t-statistic of 9.94). Columns (4) to (6) replicate columns (1) to (3) using data from January 2018 to December 2022, and with *Post* defined as 1 for March 2020 to December 2022 and 0 otherwise. The results are similar but with smaller magnitudes as expected based on Panel A of Figure 5. Figure IA.16, Panel C examines heterogeneity in these effects across demographic splits and finds that the effect is present across all of the splits and is generally highly statistically significant.

To examine the effect of PPP fraud on vehicle purchases across the entire country, we also examine data from the American Community Survey (ACS), which is an annual survey run by the US Census. In particular, we use annual data from 2015 to 2022 on the number of vehicles per household over the preceding five years by census tract.[47] We estimate a regression similar to Equation 5.1 using this data, but with the dependent variable being the log of vehicles per household and including census tract and county × year fixed effects.

[47] There are over three times as many census tracts as zip codes. ACS data at the census tract level is only released as five-year estimates. For example, the 2022 estimate is based on data collected during 2018 to 2022, which will cause any effects observed to be understated. Assuming there is no relation between PPP fraud and vehicles per household before 2020, the effect observed in 2020, 2021, and 2022 are one-fifth, two-fifths, and three-fifth of the true effect, respectively.

During the pre-COVID years, there is no relation between PPP fraud and vehicles per household. However, a one standard deviation increase in *Flagged Per Capita* is associated with a 0.23% increase in vehicles per household during 2020, a 0.38% increase during 2021, and a 0.44% increase during 2022 (see Figure IA.17).[48]

5.2. Consumer Spending

Next, we examine the effects of PPP fraud on consumer spending more broadly. Consumer spending data at the census tract level is from Mastercard's Center for Inclusive Growth and is based on anonymized and aggregated transactions on the Mastercard network. For further privacy, Mastercard ranks each census tract's consumer spending per capita each year in the national distribution and only releases the percentile rank of the tract for each year. If fraudulent PPP loans stimulated consumer spending, we would expect elevated spending in census tracts with higher PPP fraud per capita. To examine whether this is the case, we estimated regressions of the form:

$$\begin{aligned} SpendingPerCapita_{it} = \textstyle\sum_{t\neq 2019} \beta_t(Year_t \times FlaggedPerCapita_i) \\ + Tract_i + County_i \times Year_t + \epsilon_{it} \end{aligned}$$

The dependent variable in the regression is the tract's percentile rank of spending per capita. *Flagged Per Capita* is standardized so that one unit represents one standard deviation. The coefficients of interest are β_t, which estimate differences in percentile rank of spending per capita relative to the 2019 baseline that are associated with a one standard deviation increase in *Flagged Per Capita*.[49]

[48] The right plot in Figure IA.17 shows the within-county relationship between *Flagged Per Capita* and the percentage change in vehicles per household from 2019 to 2022. Consistent with the results in the left panel, there is a strong positive relation (t-stat=6.11). We also find evidence of increased auto debt in MSAs with high PPP fraud (Figure IA.18). The pattern is consistent with PPP fraud recipients initially paying down their debts and then eventually increasing their auto debts as they used PPP funds for down payments on vehicle purchases.

[49] The regressions include census tract fixed effects and county × year fixed effects to isolate differential changes within counties. In subsequent analysis, we add interactions with demographic characteristics and consider heterogeneous effects across census tracts with different demographic characteristics. Standard errors are clustered by county. Double clustering by county and year is not feasible due to having only six years of data. Purely cross-sectional results based on changes in spending (discussed below) are also highly significant.

The left plot in Panel B of Figure 5 shows the effect of a one standard deviation increase in *Flagged Per Capita* over time. During the pre-COVID period, *Flagged Per Capita* has no relation to spending per capita. After the onset of the PPP, a one standard deviation increase in *Flagged Per Capita* is associated with a 0.87 and 0.94 percentile rank increase in spending per capita in 2020 and 2021, respectively, compared to 2019. Spending levels then return to normal in 2022. This is exactly what we would expect from short-term stimulus during the PPP in 2020 and 2021.[50]

The right plot in Panel B of Figure 5 shows a binned scatter plot across census tracts of the change in spending per capita percentiles versus *Flagged Per Capita*, controlling for county fixed effects. The vertical axis plots the change in consumer spending per capita percentiles from 2019 to the average of 2020 and 2021. Consistent with the results in the left panel, there is a positive relation between spending growth and *Flagged Per Capita*.[51]

Table 7 collapses the β_t coefficients in Equation 5.2 into a single coefficient for the interaction between *Flagged Per Capita* and an indicator variable for the PPP years (2020 and 2021). The sample starts in 2017 and ends in 2021 to compare PPP years with the previous years. Column (1) estimates a regression that is similar to the regressions plotted in Panel B of Figure 5. A one standard deviation increase in *Flagged Per Capita* is associated with a 0.879 percentile rank increase in 2020 and 2021 consumer spending relative to 2019, which is nearly identical to the effects estimated in Panel B of Figure 5 for 2020 and 2021. In column (2), we add post dummy interacted with total PPP loans per capita. This specification distinguishes the stimulative effect of fraudulent PPP loans from any stimulative effect of PPP loans more generally. The coefficient on *Flagged Per Capita* increases to 1.036 and remains highly statistically significant. A potential concern is that PPP fraud could be correlated with other characteristics that predict spending growth during this time period. Column (3)

[50]The stimulative effects of PPP fraud on consumer spending in 2020 and 2021 are similar for tracts that are above and below median values of various demographics, which provides reassurance as to the consistency of the effect (Figure IA.19).

[51]This relation is highly statistically significant with a t-statistic of 8.26 based on standard errors clustered by county.

adds detailed demographic data interacted with a post dummy variable to account for this concern. Adding these control variables decreases the *Flagged Per Capita* coefficient somewhat (from 1.036 to 0.595), but the result remains large and highly statistically significant (with a t-statistic of 4.74). Column (4) replaces the year $\times$ county fixed effect with a year fixed effect and shows that the effect of PPP fraud on consumer spending is even larger in this case, which indicates that PPP fraud also predicts consumer spending differences across counties.

The results in Panel B of Figure 5 and Table 7 consistently point to elevated consumer spending in census tracts with higher levels of PPP fraud. As with the previous analysis of the housing and vehicle markets, PPP fraud is not randomly assigned, which means we do not have a perfect shock for causal interpretation. Nonetheless, elevated spending in 2020 and 2021, with a return to normal in 2022, is what we would expect from a short-term stimulus like PPP fraud and is consistent with the effects observed in the housing market and vehicle markets.

5.3. Inflation

Finally, we examine the effects of PPP fraud on regional inflation. The BLS releases bi-monthly 12-month regional consumer price indices (CPI) by CBSA, but only for 23 CBSAs. Within this limited data, we examine how PPP fraud may have affected overall price levels at the CBSA level by regressing CBSA-level inflation on PPP fraud with regressions of the form:

$$\begin{aligned} 12\text{-}monthInflation_{it} = & \textstyle\sum_{t \neq JanFeb2020} \beta_t (Bi\text{-}month_t \times FlaggedPerCapita_i) \\ & + CBSA_i + Bi\text{-}month_t + \epsilon_{it} \end{aligned}$$

The dependent variable in the regression is the CBSA's 12-month inflation, calculated on a bi-monthly basis based on regional CPI data from the BLS. The coefficients of interest are β_t, which estimate differences in inflation rates relative to the January/February 2020 baseline that are associated with a one standard deviation increase in *Flagged Per Capita*.[52]

Figure 6 shows the results. During the pre-COVID period from 2010 to 2019, regional inflation rates had no relation to *Flagged Per Capita*. There is also little relation between PPP fraud and inflation growth from March to August 2020. Inflation then starts to pick up in CBSAs with high PPP fraud in September/October 2020, with statistically significant differences by January/February 2021, and larger and significant effects throughout 2021 and most of 2022, peaking in July/August 2022. The effects are positive but smaller and insignificant from November/December 2022 to January/February 2024. These patterns correspond closely to overall inflation, which first increased above 2% in March 2021 and has remained elevated since then with a peak annual inflation of 9.06% in June 2022. We also separately consider the housing and non-housing components of CPI with evidence of increased housing inflation and a smaller impact on non-housing inflation (see Figure IA.21, Panel A).[53]

Table 8 collapses the β_t coefficients in Equation 5.3 into a single coefficient for the interaction between *Flagged Per Capita* and an indicator variable for the post-PPP time period, starting in March/April of 2020.[54] Column (1) reports results for overall inflation. Consistent with Figure 6, a one standard deviation increase in *Flagged Per Capita* is associated with 0.43 ppt higher 12-month inflation during the post-PPP time period. This is a large increase and is highly statistically significant with a t-statistic of 3.40.[55] In column (2), we

[52]12-month inflation is determined as the current CPI divided by the CPI 12 months earlier. We consider inflation from January 2010 to April 2024. *Flagged Per Capita* is standardized so that one unit represents one standard deviation. The regressions include CBSA and bi-month fixed effects. Standard errors are double clustered by CBSA and bi-month.

[53]Our evidence in Section 5.1 suggests that recipients of fraudulent PPP were more likely purchase vehicles during March 2020 to December 2021. Given that automobiles are a movable good and the automobile index depends on composition of car purchases (which was affected by supply issues in late 2020 and 2021), it is not clear if any effect would be detectable in regional car prices. In Panel B of Figure IA.21, we find that the vehicle component of regional CPI is somewhat elevated in high PPP fraud CBSAs in mid-2020 before returning to normal by the end of 2020.

[54]Bi-monthly data from January 2010 to December 2022 is used in these regressions.

[55]Figure IA.20 plots the CBSA-level data with particularly high fraud and elevated inflation in Atlanta. Regression results are robust to excluding any single CBSA. Excluding Atlanta results in a coefficient of 0.35 ppt with a t-statistic of 2.74. The t-statistics reported in parentheses are based on standard errors that are double clustered by CBSA and bi-month. Additionally, since the dependent variable is based on overlapping periods, we report t-statistics based on Newey-West standard errors with 6 lags in square brackets.

add the interaction between the post indicator and overall PPP loans per capita to assess whether inflation is coming from PPP loans in general as opposed to fraudulent PPP loans. The coefficient for overall PPP loans per capita is close to zero, and the coefficient on *Flagged Per Capita* increases slightly after controlling for overall PPP loans per capita.

Column (3) of Table 8 repeats the same regressions for housing inflation with an even larger effect of 0.71 ppt. Controlling for overall PPP loans per capita in column (4) again slightly increases the result for flagged PPP loans. Columns (5) and (6) focus on non-housing inflation with an insignificant effect in column (5) and a significant though smaller effect of 0.25 ppt in column (6).

Overall, the results in Figure 6 and Table 8 indicate a strong relation between PPP fraud and inflation, even with relatively limited data for only 23 CBSAs. This inflation is primarily concentrated in housing but is apparent to a smaller extent in non-housing CPI components. Combined with analysis on vehicle purchases and consumer spending in the previous subsections, this suggests that PPP fraud broadly stimulated consumer spending beyond the housing market. Differential price pressure across CBSAs is likely most pronounced for housing because this is an immovable good with distinct local markets.

6. Conclusion

The U.S. government responded to the COVID-19 pandemic with massive relief spending and minimal fraud safeguards. The result was hundreds of billions of dollars in pandemic relief fraud, much of which flowed to highly concentrated geographic markets. Our analysis highlights that recipients of fraudulent funds were more likely to purchase houses. As a result, high fraud zip codes experienced a 5.7 percentage point larger increase in house prices, indicating a sizeable distortionary effect on home prices. The effect holds on monthly data beginning in May 2020 and continues through May 2022. Matching and synthetic control analyses show that zip codes with high PPP fraud experience divergent growth starting in mid-2020, compared to low fraud zip codes within the same CBSA or county that have similar house price trends pre-COVID. When compared to other variables that have been examined in the literature using variable selection and Bayesian Model Averaging methods, flagged loans per capita and land unavailability are the economically largest home price predictors, but remote work, teleworking, migration, and past price growth are also statistically significant indicators. Additionally, vehicle purchases and consumer spending are elevated in areas with high PPP fraud, and pandemic fraud appears to have had a meaningful effect on overall inflation.

Our findings support the idea that unintended fraud externalities, including in the form of distorted asset prices, can be much broader and more costly than the direct costs (Akerlof and Romer, 1993). Our findings also indicate that rent-seeking in the financial system (Zingales, 2015) can have large spillovers into real markets. Given that our findings show that fraudulent transfers are wealth shocks associated with distortions that are not present in normal transfers, future government program designs should take more proactive steps on the front end to prevent fraud, and authorities should allocate more targeted resources for back-end auditing and prosecution as it is thought that the pandemic fraud may have overwhelmed government prosecutors. Additional research could consider other externalities of fraud such as encouraging future criminal behavior.

References

Abadie, A., A. Diamond, and J. Hainmueller. 2014. Comparative politics and the synthetic control method. *American Journal of Political Science* 59:495–510.

Aiello, D., S. Baker, T. Balyuk, M. Di Maggio, M. Johnson, and J. Kotter. 2023a. Cryptocurrency investing: Stimulus checks and inflation expectations. Working Paper.

Aiello, D., S. R. Baker, T. Balyuk, M. Di Maggio, M. J. Johnson, and J. D. Kotter. 2023b. The effects of cryptocurrency wealth on household consumption and investment. Working Paper.

Akerlof, G. A., and P. M. Romer. 1993. Looting: The economic underworld of bankruptcy for profit. *Brookings Papers on Economic Activity* 1993:1–73.

Associated Press. 2023. The great grift: How billions in COVID-19 relief aid was stolen or wasted. Richard Lardner, Jennifer McDermott, and Aaron Kessle, June 12.

Autor, D., D. Cho, L. Crane, M. Goldar, B. Lutz, J. Montes, W. Peterman, D. Ratner, D. Villar, and A. Yildirmaz. 2022. The $800 billion Paycheck Protection Program: Where did the money go and why did it go there? *Journal of Economic Perspectives* 36:55–80.

Badarinza, C., and T. Ramadorai. 2018. Home away from home? Foreign demand and London house prices. *Journal of Financial Economics* 130:532–55.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018a. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126:2224–76.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong. 2018b. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* 32:259–80.

Bailey, M., E. Dávila, T. Kuchler, and J. Stroebel. 2018c. House Price Beliefs And Mortgage Leverage Choice. *The Review of Economic Studies* 86:2403–52.

Bailey, M., P. Farrell, T. Kuchler, and J. Stroebel. 2020. Social connectedness in urban areas. *Journal of Urban Economics* 118:103264–.

Bartik, A. W., M. Bertrand, Z. Cullen, E. L. Glaeser, M. Luca, and C. Stanton. 2020. The impact of COVID-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences* 117:17656–66.

Becker, G. S. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 76:169–217.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng. 2024. On binscatter. *American Economic Review* 114:1488–514.

Cherry, S., E. Jiang, G. Matvos, T. Piskorski, and A. Seru. 2021. Government and private household debt relief during COVID-19. *Brookings Papers on Economic Activity* Fall Edition:141–99.

Chetty, R., J. N. Friedman, N. Hendren, and M. Stepner. 2023. The economic impacts of COVID-19: Evidence from a new public database built using private sector data. *The Quarterly Journal of Economics* qjad048.

Cole, A. 2024. The impact of the paycheck protection program on (really) small businesses. Working Paper.

Dalton, M. 2023. Putting the paycheck protection program into perspective: An analysis using administrative and survey data. *National Tax Journal* 76:393–437.

Davis, M. A., A. C. Ghent, and J. Gregory. 2024. The Work-From-Home Technology Boon and its Consequences. *The Review of Economic Studies* rdad114.

de Soyres, F., A. M. Santacreu, and H. L. Young. 2023. Demand-supply imbalance during the COVID-19 pandemic: The role of fiscal policy. *Federal Reserve Bank of St. Louis Review* First Quarter 2023:21–50.

Denes, M., S. Lagaras, and M. Tsoutsoura. 2021. First come, first served: The timing of government support and its impact on firms. Working Paper.

Diamond, W., T. Landvoigt, and G. Sanchez. 2023. Printing away the mortgages: Fiscal inflation and the post-Covid housing boom. Working paper.

Dingel, J. I., and B. Neiman. 2020. How many jobs can be done at home? *Journal of Public Economics* 189:104235–.

Faulkender, M. W., R. Jackman, and S. Miran. 2021. The job preservation effects of Paycheck Protection Program loans. Working Paper.

Favilukis, J., D. Kohn, S. C. Ludvigson, and S. Van Nieuwerburgh. 2012. *International capital flows and house prices: Theory and evidence*, 235–99. University of Chicago Press.

Fernández, C., E. Ley, and M. F. Steel. 2001. Benchmark priors for bayesian model averaging. *Journal of Econometrics* 100:381–427.

Gamber, W., J. Graham, and A. Yadav. 2023. Stuck at home: Housing demand during the COVID-19 pandemic. *Journal of Housing Economics* 59:101908–.

Gee, J., and M. Button. 2019. *The financial cost of fraud 2019: The latest data from around the world*. United Kingdom: Crowe UK.

Gorback, C., and B. Keys. 2023. Global capital and local assets: House prices, quantities, and elasticities. Working Paper.

Granja, J., C. Makridis, C. Yannelis, and E. Zwick. 2022. Did the Paycheck Protection Program hit the target? *Journal of Financial Economics* 145:725–61.

Griffin, J. M., S. Kruger, and P. Mahajan. 2023a. Did FinTech lenders facilitate PPP fraud? *The Journal of Finance* 78:1777–827.

———. 2023b. Is fraud contagious? Social connections and the looting of COVID relief programs. Working Paper.

Griffin, J. M., S. Kruger, and G. Maturana. 2020. What drove the 2003–2006 house price boom and subsequent collapse? Disentangling competing explanations. *Journal of Financial Economics* 141:1007–35.

Griffin, J. M., and G. Maturana. 2016. Did dubious mortgage origination practices distort house prices? *The Review of Financial Studies* 29:1671–1708–.

Guiso, L., P. Sapienza, and L. Zingales. 2008. Trusting the stock market. *The Journal of Finance* 63:2557–600.

Gupta, A., V. Mittal, J. Peeters, and S. Van Nieuwerburgh. 2022. Flattening the curve: Pandemic-induced revaluation of urban real estate. *Journal of Financial Economics* 146:594–636.

Gurun, U. G., N. Stoffman, and S. E. Yonker. 2017. Trust busting: The effect of fraud on investor behavior. *The Review of Financial Studies* 31:1341–76.

Hartman-Glaser, B., M. Thibodeau, and J. Yoshida. 2023. Cash to spend: IPO wealth and house prices. *Real Estate Economics* 51:68–102.

Hu, Z. 2021. Social interactions and households' flood insurance decisions. *Journal of Financial Economics* 144:414–32.

Jorda, O., and F. Nechio. 2023. Inflation and wage growth since the pandemic. *European Economic Review* 156:104474–.

Kedia, S., and T. Philippon. 2007. The Economics of Fraudulent Accounting. *The Review of Financial Studies* 22:2169–99.

Ley, E., and M. F. Steel. 2009. On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* 24:651–74.

Lin, L. 2022. Fiscal stimulus payments, housing demand, and house price inflation Working Paper.

Liu, S., and Y. Su. 2021. The impact of the COVID-19 pandemic on the demand for density: Evidence from the U.S. housing market. *Economics Letters* 207:110010–.

Lutz, C., and B. Sands. 2022. Highly disaggregated land unavailability. Working Paper.

Mian, A., and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis. *The Quarterly Journal of Economics* 124:1449–1496–.

———. 2018. Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. *The Review of Financial Studies* 30:1832–1864–.

Mondragon, J. A., and J. Wieland. 2022. Housing demand and remote work. Working Paper.

NBC News. 2022. 'Biggest fraud in a generation': The looting of the Covid relief plan known as PPP. Ken Dilanian and Laura Strickler, March 28.

Neilson, C., J. E. Humphries, and G. Ulyssea. 2020. Information frictions and access to the Paycheck Protection Program. *Journal of Public Economics* 190.

Oster, E. 2019. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37:187–204.

ProPublica. 2021. How unemployment insurance fraud exploded during the pandemic. Cezary Podkul, July 26.

Rabetti, D., and L. W. Cong. 2023. Does relationship lending discipline disclosure? Evidence from bailout loans. Working Paper.

Ramani, A., and N. Bloom. 2021. The donut effect of Covid-19 on cities. Working Paper.

SBA OIG. 2023. COVID-19 pandemic EIDL and PPP loan fraud landscape.

Zingales, L. 2015. Presidential address: Does finance benefit society? *Journal of Finance* 70:1327–63.

Figure 1. Housing Purchases and Moving

This figure shows the relationship between individuals receiving a flagged PPP loan and their likelihood of housing purchase. Data from PropertyRadar for a sample of 250,000 loans is used to determine home purchases from one year before the individual received their PPP loan to 18 months after. The sample of non-flagged loans is reweighted to match the distribution of timing of PPP loan approval of the sample of flagged loans. The horizontal lines are the average monthly likelihood of an individual in each group buying a house during the pre-/post-period.

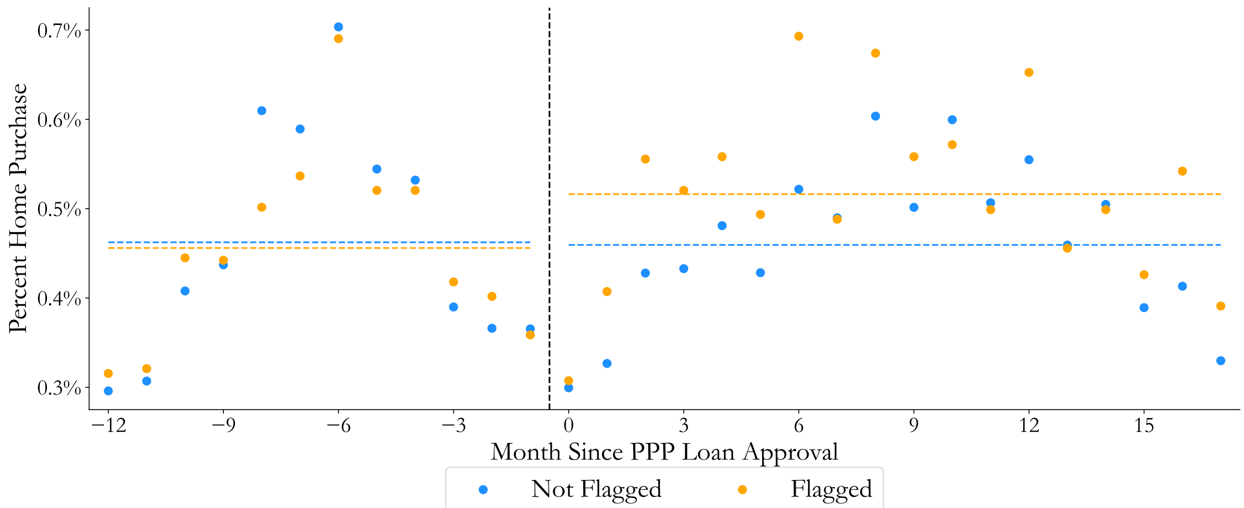
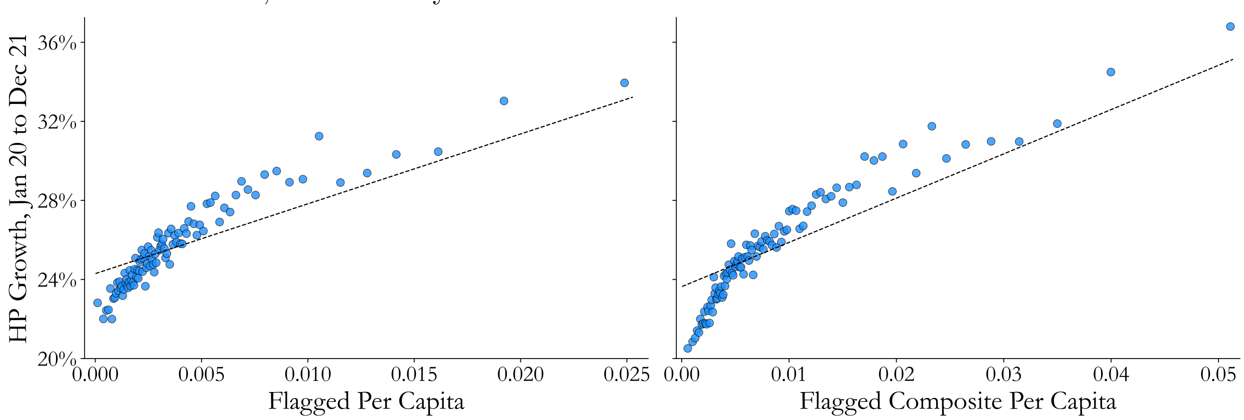


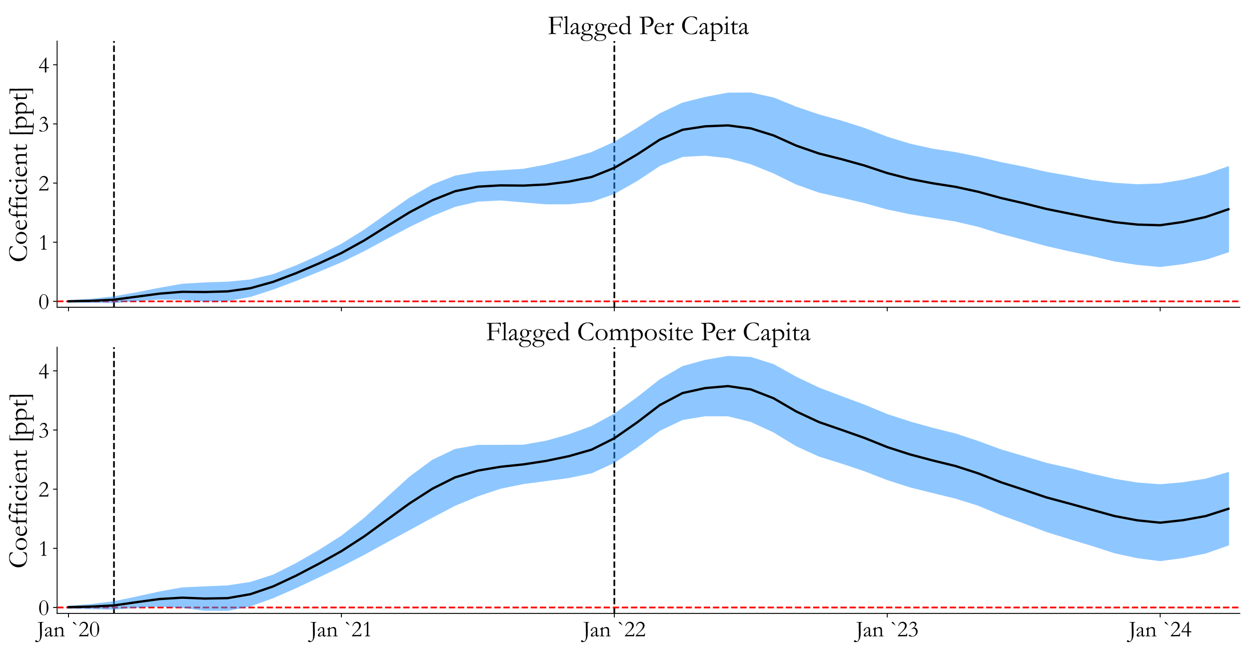
Figure 2. Effect of Suspicious Lending on Housing Prices

This figure shows the relationship between suspicious lending and house prices. Panel A shows the relationship using binscatters. Panel B shows the relationship over time. Panel C examines heterogeneity in the relationship across demographics. All three panels are estimated using zip code-level data, include county fixed effects, and control for house price growth in 2018 to 2019 and loans per capita using percentile fixed effects. Further, they control for log population density, vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. The left subpanel of Panel A and the top subpanels of Panels B and C are based on *Flagged Per Capita*. The right subpanel of Panel A and the bottom subpanels of Panels B and C are based on *Flagged Composite Per Capita*. *Flagged Per Capita* is a ratio of the number of flagged PPP loans in the zip code to the zip code's population. *Flagged Composite Per Capita* is based on the number of loans that are either flagged, in county-industry pairs where there are more than two times as many PPP loans as establishments, or in county-lender pairs with high levels of similarity. See Griffin, Kruger, and Mahajan (2023a) for additional details about these measures. In Panel B and C, the measures of suspicious lending are standardized, so the coefficients represent the house price effect of a one standard deviation change in suspicious lending. The splits in Panel C are based on the median value of the demographic. To have a nationally representative estimate, all three panels use weighted least squares (WLS) regressions with the weight being population of the zip code in 2019. The error bars in Panels B and C represent 95% confidence intervals based on standard errors clustered by county.

Panel A. Binscatters, With County FEs and Controls



Panel B. Effect Over Time



Panel C. Heterogeneity by Demographics

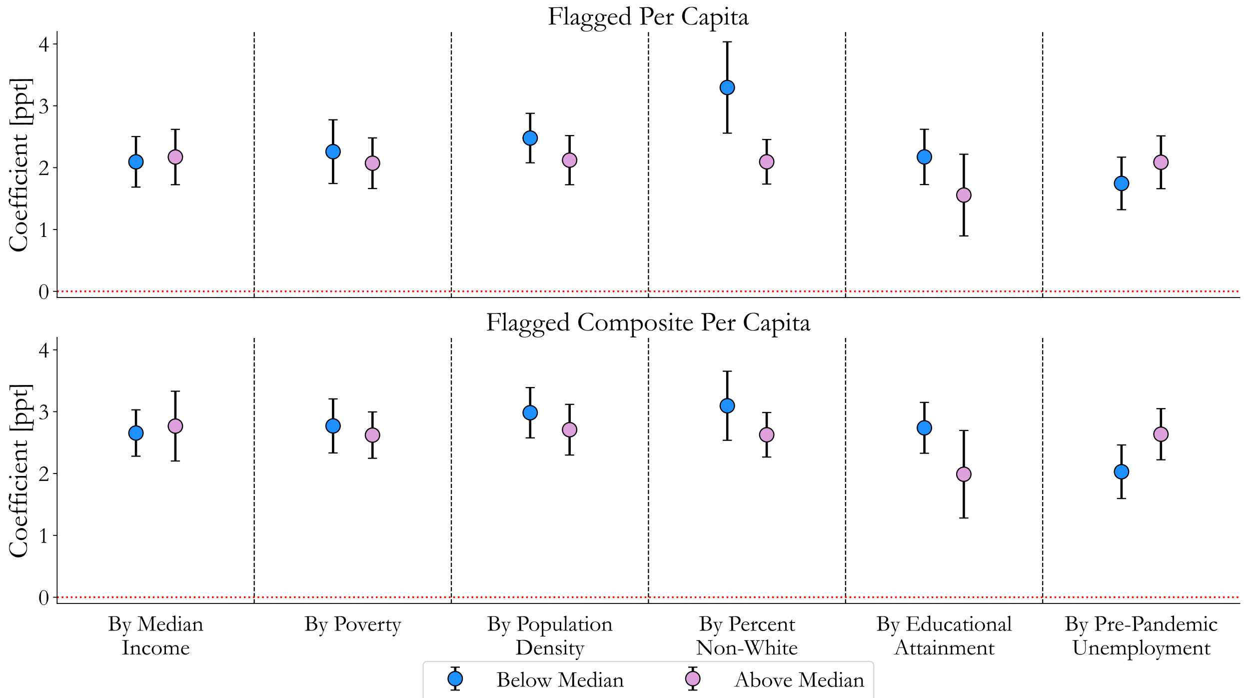


Figure 3. Synethetic Control

This figure shows the effect of suspicious lending on house price growth. A synthetic control method is used to create controls for each zip code in the top quartile of *Flagged Per Capita* using all zip codes in the same county that are in the bottom quartile of *Flagged Per Capita*. Zip codes are split into quartiles within county. Counties where the difference between the 75th and 25th percentile of Flagged Per Capita within the county is at least half as large as the standard deviation across the entire nation are included. The dashed lines are 95% confidence intervals.

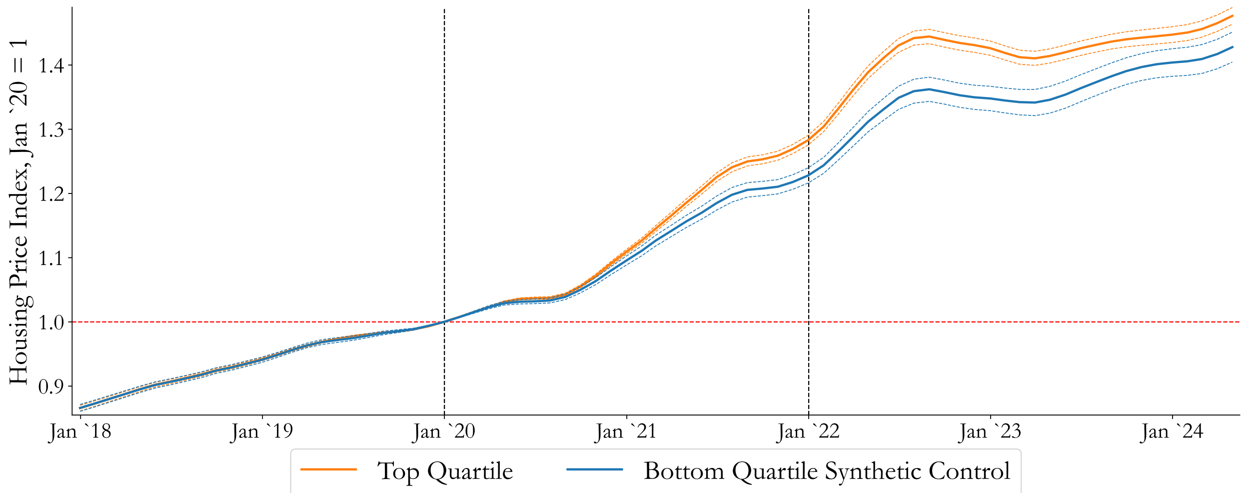
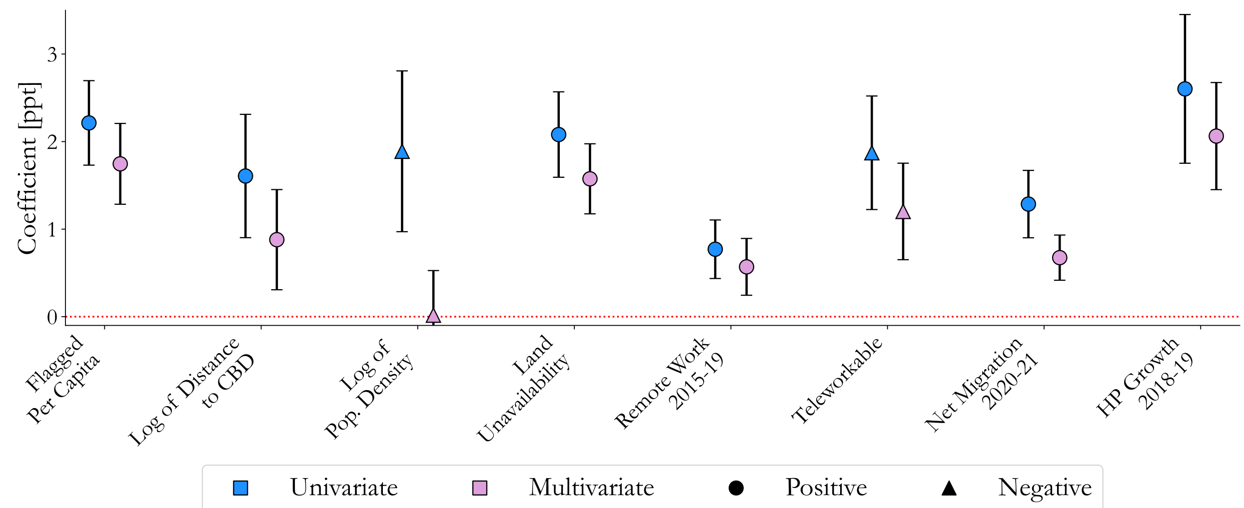


Figure 4. Effect of Other Proposed Variables on Housing Prices
This figure shows the effect of each proposed variable on house prices. Panel A shows univariate and multivariate coefficients using WLS regressions. Panel B shows the posterior coefficient distribution, conditional on inclusion, from multivariate regressions using Bayesian model averaging. The black bars at the top of each distribution plot the posterior inclusion probability. All regressions control for vacancy rate, log housing units, log average household income, and for overall PPP loans per capita and house price growth in 2018-19 using percentile indicator variables to allow for non-linearity, and include county fixed effects. All proposed variables are standardized to have a mean of 0 and a standard deviation of 1. The weighted least squares (WLS) regressions are weighted by the population of the zip code in 2019. Only zip codes for which all proposed variables can be determined are used in both Panels. The error bars in Panel A represent 95% confidence intervals based on standard errors clustered by county. The regressions corresponding to Panel A are shown in Table 4.

Panel A. Univariate and Multivariate Coefficients



Panel B. Bayesian Model Averaging

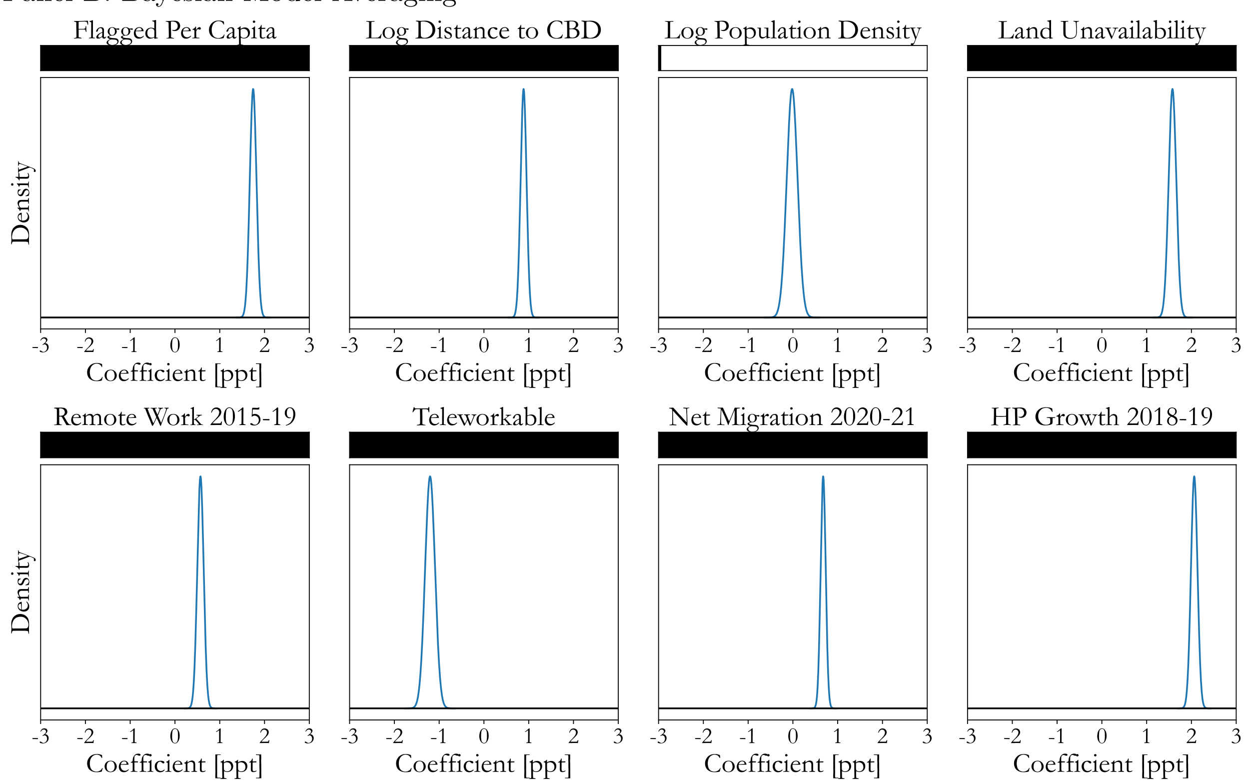
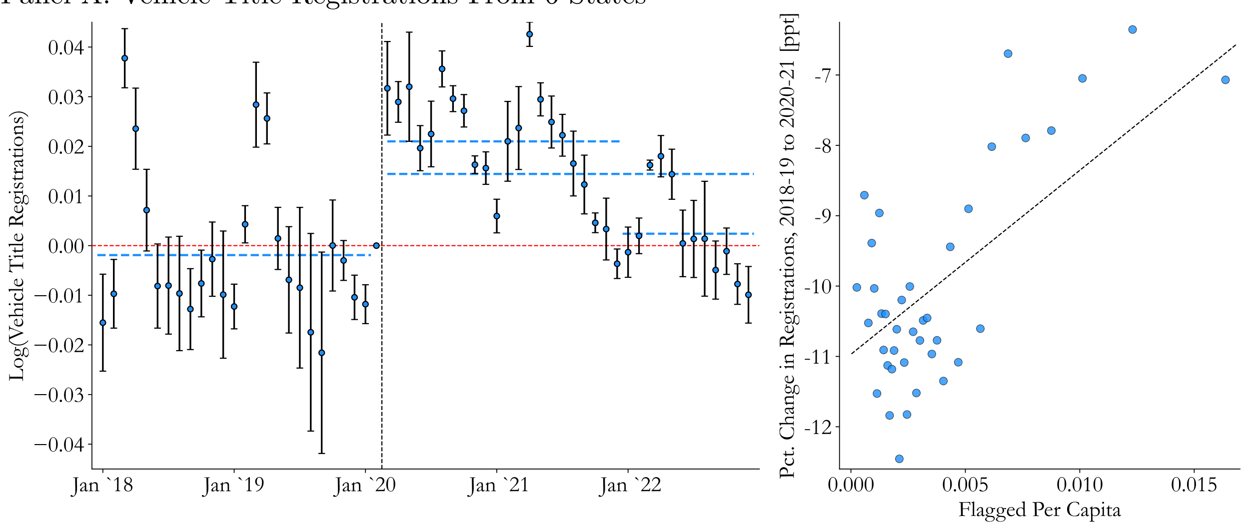


Figure 5. Effect on Vehicle Purchases and Consumer Spending

This figure shows the effect of suspicious lending on vehicle purchases and consumer spending. Panel A uses vehicle title registration data from six states (California, Texas, Florida, Illinois, Ohio, and Washington) at the zip code-month level. Panel B uses annual data for 2017 to 2022 at the census tract from Mastercard's Center for Inclusive Growth. Mastercard ranks each census tract's consumer spending per capita each year in the national distribution and releases the percentile rank of the tract. Data from 2017 to 2022 is used. The left subpanels examine the effects of a one standard deviation change in the number of flagged PPP loans per capita. The left subpanel of Panel A includes zip code and month $\times$ county fixed effects, and the left subpanel of Panel B includes census tract and year $\times$ county fixed effects. The error bars in the left subpanels represent 95% confidence intervals based on standard errors double clustered by county and month in Panel A and clustered by county in Panel B. In the left subpanel of Panel A, the horizontal blue lines represent the average coefficient over the period spanned by each line. The right subpanels examine the within-county effects of flagged PPP loans per capita on percentage change in vehicle registrations between 2018-19 to 2020-21 in Panel A and percentile change in spending per capita between 2019 and 2020-21 in Panel B. To have a nationally representative estimate, both panels use weighted least squares (WLS) regressions with the weight being the zip code's (census tract's) population as of 2019 in Panel A (Panel B).

Panel A. Vehicle Title Registrations From 6 States



Panel B. Consumer Spending

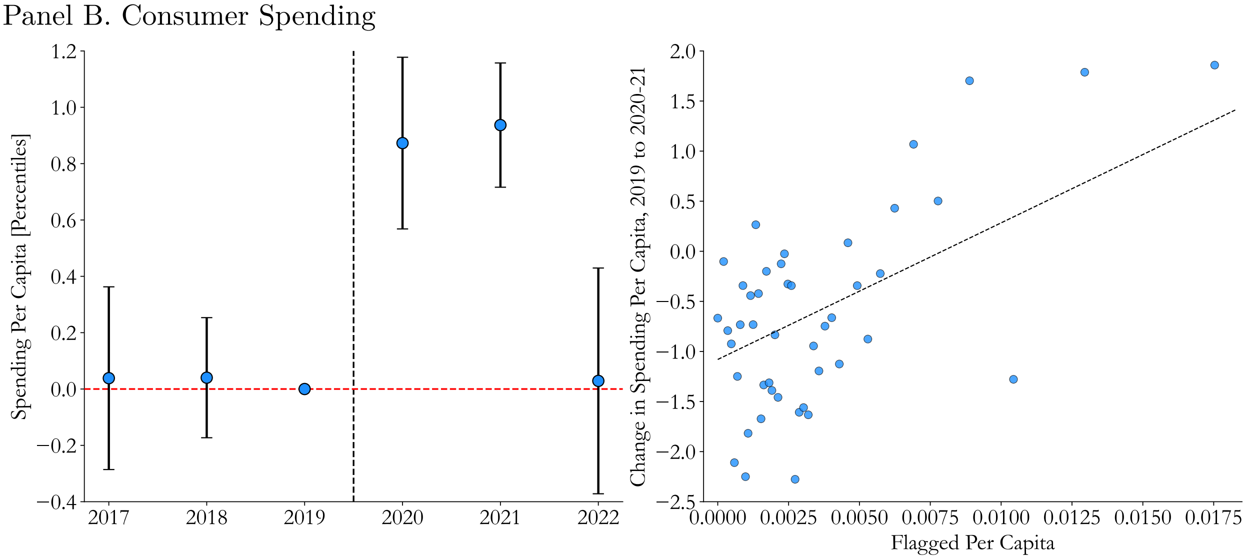


Figure 6. Effect on Regional Inflation

This figure shows the effect of suspicious lending on regional inflation using regional all items consumer price indices (CPI) from the BLS. Data for 23 CBSAs is released bi-monthly. 12-month inflation is determined by dividing the given month's CPI by the CPI 12 months earlier. *Flagged Per Capita* is standardized, so the coefficients represent the inflation effect of a one standard deviation change in suspicious lending. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the CBSA's population as of 2019. The error bars represent 95% confidence intervals based on standard errors double clustered by CBSA and bi-month.

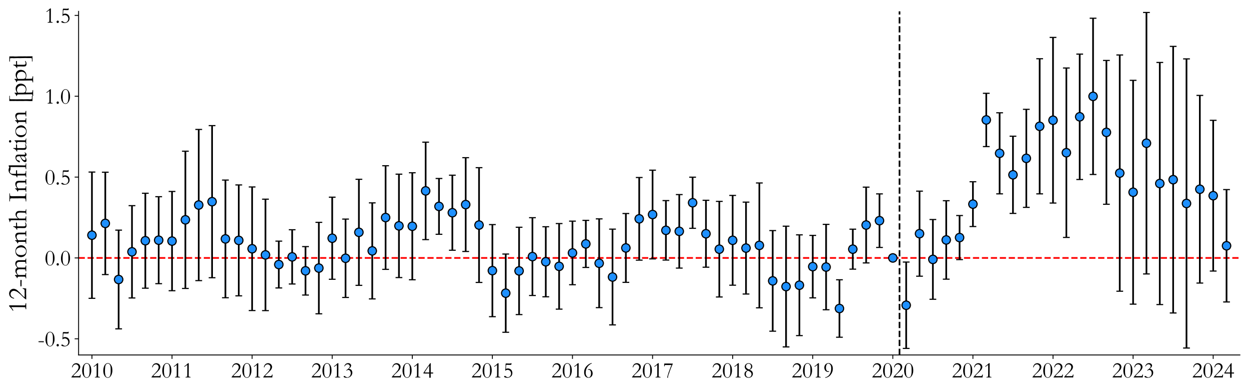


Table 1. Housing Purchases

This table examines whether individuals purchased homes after receiving a flagged PPP loan. Data on home purchases for a sample of 250,000 loans from PropertyRadar is used in columns (1) and (2) and for a sample of 150,000 loans from LexisNexis in columns (3) and (4). For each individual, we include monthly observations for the five years before they received their PPP loan to 18 (4) months after for PropertyRadar (LexisNexis). $1(HousingPurchase)$ takes a value of 12 (multiplied by 12 to annualized) if the individual bought a house during the given month. $1(Flagged)$ takes a value of 1 if the individual received a PPP loan that is flagged by at least one of the primary measures from Griffin, Kruger, and Mahajan (2023a). $1(Post)$ takes a value of 1 if the month is after the individual received their PPP loan. Fixed effects are indicated at the bottom of each column. Robust standard errors are double clustered by PPP loan and month.

Dep. Variable: 1(Housing Purchase) × 12

	(1)	(2)	(3)	(4)
Sample:	PropertyRadar		LexisNexis	
1(Flagged) × 1(Post)	0.00863***	0.00863***	0.0108***	0.00781***
	(5.91)	(5.43)	(6.37)	(5.25)
1(Post)	0.00685**		-0.00515***	
	(2.50)		(-3.94)	
Loan FE	Yes	Yes	Yes	Yes
Month of Year FE	Yes	Yes	Yes	Yes
1(Post) × ln(Loan Amount)	No	Yes	No	Yes
1(Post) × County FE	No	Yes	No	Yes
1(Post) × Business Type FE	No	Yes	No	Yes
1(Post) × Week Approved FE	No	Yes	No	Yes
Observations	19,500,000	19,500,000	9,600,000	9,600,000
R^2	0.0279	0.0279	0.0203	0.0203
Mean of Dep. Variable	0.0501	0.0501	0.0559	0.0559

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 2. Housing Price Growth

This table examines the relationship between house price growth and various measures of suspicious PPP lending. *Flagged Per Capita* is a ratio of the number of flagged PPP loans in the zip code to the zip code's population. *High Loan-to-Est. Per Capita* is based on the number of PPP loans in county-industry pairs where there are more than two times as many PPP loans as establishments per the US Census CBP. *High Similarity Per Capita* is based on the number of PPP loans in county-lender pairs with high levels of similarity in terms of loan amount, jobs reported, and industry. *Flagged Composite Per Capita* is based on the number of loans that are either flagged, in industry-counties where there are more than two times as many PPP loans as establishments, or in lender-counties with high levels of similarity. See Griffin, Kruger, and Mahajan (2023a) for additional details about these measures. *Past HP Growth* and *Loans Per Capita* control for house price growth in 2018-19 and PPP lending intensity, respectively, using percentile fixed effects. The controls included are log population density, vacancy rate, log housing units, and log average household income. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Flagged Per Capita	0.0179*** (10.38)	0.0211*** (9.84)			
High Loan-to-Est. Per Capita			0.0222*** (11.97)		
High Similarity Per Capita				0.0224*** (11.06)	
Flagged Composite Per Capita					0.0268*** (13.34)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth	No	Yes	Yes	Yes	Yes
Loans Per Capita	No	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215	2,215
R^2	0.781	0.828	0.824	0.828	0.827
Mean of Dep. Var.	0.259	0.259	0.259	0.259	0.259

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 3. Housing Price Growth, IV

This table examines the relationship between house price growth and suspicious PPP lending using instrumented variables based on social connectedness between zip codes. Column (1) is based on social connectedness between each zip code and zip codes that are outside the given zip code's CBSA. Columns (2), (3), and (4) are based on social connectedness between each zip code and zip codes that are at least 100, 250, and 500 miles away, respectively. Column (5) includes three instruments based on social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles) at the same time. The J-stat and p-value for an overidentification test are provided at the bottom of column (5). *Past HP Growth* and *Loans Per Capita* control for house price growth in 2018-19 and PPP lending intensity, respectively, using percentile fixed effects. The controls included are log population density, vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions in both stages with the weight being the population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

Instrument:	(1) Outside CBSA	(2) $\geq$ 100 Mi	(3) $\geq$ 250 Mi	(4) $\geq$ 500 Mi	(5) Concentric Rings
Flagged Per Capita	0.0347***	0.0353***	0.0336***	0.0340***	0.0367***
	(6.35)	(6.51)	(6.17)	(5.43)	(6.41)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,725
Num. Counties	1,789	2,215	2,215	2,215	2,213
R^2	0.256	0.247	0.251	0.250	0.243
Mean of Dep. Var.	0.257	0.259	0.259	0.259	0.259
First Stage F-stat	34.33	38.04	34.27	28.48	14.38
Hansen's J-stat					2.465
(p-value)					(0.292)

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 4. Housing Price Growth, Univariate and Multivariate

This table examines the univariate and multivariate relationships between various proposed variables and house price growth. *Past HP Growth* and *Loans Per Capita* control for house price growth in 2018-19 and PPP lending intensity, respectively, using percentile fixed effects. The controls included are vacancy rate, log housing units, and log average household income. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the population of the zip code in 2019. Only zip codes for which all variables can be determined are used. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flagge Per Capita	0.0221*** (8.99)								0.0175*** (7.41)
Log of Dist. to CBD		0.0161*** (4.47)							0.00880*** (3.01)
Log of Pop. Density			-0.0189*** (-4.03)						-0.000170 (-0.07)
Land Unavailability				0.0208*** (8.36)					0.0157*** (7.72)
Remote Work 2015-19					0.00771*** (4.52)				0.00569*** (3.44)
Teleworkable						-0.0187*** (-5.66)			-0.0120*** (-4.27)
Net Migration 2020-21							0.0129*** (6.55)		0.00674*** (5.12)
HP Growth 2018-19								0.0260*** (6.00)	0.0206*** (6.61)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Loans Per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	12,305	12,305	12,305	12,305	12,305	12,305	12,305	12,305	12,305
Num. Counties	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013
R^2	0.814	0.811	0.807	0.810	0.803	0.806	0.810	0.797	0.828
Mean of Dep, Var.	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 5. Variable Selection

This table shows the results of a variable selection process where the optimal, based on the Bayesian Information Criteria, model with between one and eight independent variables are included in the model. Column (9) reports posterior inclusion probabilities for each variable based on Bayesian model averaging. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	Regression Coefficients and t-statistics								BMA Posterior Inclusion Prob.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flagged Per Capita	0.0269***	0.0223***	0.0202***	0.0186***	0.0176***	0.0176***	0.0175***	0.0175***	1.000
	(11.52)	(9.79)	(8.02)	(7.65)	(7.41)	(7.57)	(7.42)	(7.41)	
HP Growth 2018-19		0.0215***	0.0231***	0.0217***	0.0217***	0.0208***	0.0206***	0.0206***	1.000
		(5.96)	(6.93)	(6.72)	(6.74)	(6.78)	(6.60)	(6.61)	
Log of Dist. to CBD			0.0142***	0.0127***	0.00990***	0.00903***	0.00886***	0.00880***	1.000
			(3.94)	(3.97)	(3.20)	(2.95)	(2.89)	(3.01)	
Land Unavailability				0.0163***	0.0161***	0.0159***	0.0158***	0.0157***	1.000
				(8.16)	(7.77)	(7.44)	(7.59)	(7.72)	
Net Migration 2020-21					0.00825***	0.00740***	0.00676***	0.00674***	1.000
					(6.20)	(5.74)	(5.42)	(5.12)	
Teleworkable						-0.0108***	-0.0120***	-0.0120***	1.000
						(-3.78)	(-4.27)	(-4.27)	
Remote Work 2015-19							0.00569***	0.00569***	1.000
							(3.44)	(3.44)	
Log of Pop. Density								-0.000170	0.009
								(-0.07)	
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	12,305	12,305	12,305	12,305	12,305	12,305	12,305	12,305	12,305
Num. Counties	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013
R^2	0.798	0.811	0.818	0.823	0.826	0.827	0.828	0.828	
Mean of Dep. Var.	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257	0.257

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 6. Effect on Vehicle Purchases

This table examines the effect of suspicious lending on vehicle purchases using vehicle title registration data from six states (California, Texas, Florida, Illinois, Ohio, and Washington) at the zip code-month level. *Post* is a dummy variable that takes a value of 1 from March 2020 to December 2021 (2022) in columns (1) to (3) (columns (4) to (6)) and a value of 0 from January 2018 to February 2020 in all columns. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are double clustered by county and month.

Dep. Variable: Log(Vehicle Title Registrations)

	(1)	(2)	(3)	(4)	(5)	(6)
Post Period:	March 2020 to December 2021			March 2020 to December 2022		
Post × Flagged Per Capita	0.0228***	0.0309***	0.0238***	0.0163***	0.0201***	0.0163***
	(5.42)	(3.76)	(9.94)	(4.32)	(2.98)	(9.56)
Post × Loans Per Capita		-0.0150**	-0.00587**		-0.00707	-0.00136
		(-2.23)	(-2.36)		(-1.24)	(-0.56)
Post × Median Income			0.0203**			0.0274***
			(2.36)			(3.40)
Post × Poverty			0.00267			0.00295
			(0.37)			(0.46)
Post × Population Density			0.0173***			0.0146***
			(4.12)			(3.65)
Post × Pct. Non-White			-0.00786			-0.00156
			(-1.57)			(-0.36)
Post × Educational Attainment			-0.0285**			-0.0233**
			(-2.47)			(-2.20)
Post × Pre-Pandemic Unemployment			0.0104**			0.00599
			(2.58)			(1.67)
Zip Code FE	Yes	Yes	Yes	Yes	Yes	Yes
Month × County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	278,400	278,400	278,400	348,000	348,000	348,000
R^2	0.981	0.981	0.981	0.979	0.979	0.979

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 7. Effect on Consumer Spending

This table examines the effect of suspicious lending on consumer spending using annual data at the census tract level from Mastercard's Center for Inclusive Growth. Mastercard ranks each census tract's consumer spending per capita each year in the national distribution and only releases the percentile rank of the tract for each year. Data from 2017 to 2021 is used. *Post* is a dummy variable that takes a value of 1 if the year is 2020 or 2021 and 0 otherwise. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the census tract's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Spending Per Capita Percentile Rank

	(1)	(2)	(3)	(4)
Post × Flagged Per Capita	0.879***	1.036***	0.595***	0.895***
	(8.86)	(9.90)	(4.74)	(8.75)
Post × Loans Per Capita		-0.218***	0.0557	-0.181**
		(-2.66)	(0.44)	(-2.46)
Post × Median Income			-0.0889	-0.512***
			(-0.48)	(-3.23)
Post × Poverty			0.0471	0.170
			(0.37)	(1.36)
Post × Population Density			0.490	0.720***
			(1.53)	(5.30)
Post × Pct. Non-White			0.731***	0.0653
			(3.44)	(0.44)
Post × Educational Attainment			-1.582	-1.782**
			(-1.63)	(-2.09)
Post × Pre-Pandemic Unemployment			0.122	0.414***
			(1.08)	(3.71)
Census Tract FE	Yes	Yes	Yes	Yes
Year × County FE	Yes	Yes	Yes	No
Year FE	No	No	No	Yes
Observations	307,585	307,585	307,585	307,585
R^2	0.348	0.348	0.348	0.305

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table 8. Effect on Regional Inflation

This table examines the effect of suspicious lending on regional inflation using regional consumer price indices (CPI) from the BLS. Data for 23 CBSAs is released bi-monthly. 12-month inflation is determined by dividing the given month's CPI by the CPI 12 months earlier. Data from January 2010 to December 2022 is used. *Post* is a dummy variable that takes a value of 1 if the bi-month is on or after March 2020 and zero otherwise. Columns (1) and (2) are based on the all items regional CPI, columns (3) and (4) are based on the housing component of the regional CPI, and columns (5) and (6) are based on the all items excluding shelter regional CPI. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the CBSA's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. t-statistics based on robust standard errors that are double clustered by CBSA and bi-month are reported in parenthesis. t-statistics based on Newey-West standard errors with 6 lags are reported in square brackets.

Dep. Variable: 12-month Inflation

	(1)	(2)	(3)	(4)	(5)	(6)
CPI Used:	Overall		Housing Component		Excluding Shelter	
Post × Flagged	0.00429***	0.00613***	0.00707***	0.00964***	0.00139	0.00252**
Per Capita	(3.40)	(3.13)	(4.10)	(3.69)	(1.48)	(2.28)
	[3.45]	[3.27]	[3.82]	[3.37]	[1.46]	[1.68]
Post × Loans		-0.00259*		-0.00363		-0.00159
Per Capita		(-1.91)		(-1.21)		(-1.43)
		[-1.49]		[-0.98]		[-1.15]
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes
Bimonthly FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,317	1,317	1,317	1,317	1,317	1,317
Num. CBSAs	23	23	23	23	23	23
R^2	0.895	0.896	0.749	0.751	0.924	0.924
Mean of Dep. Var.	0.0254	0.0254	0.0271	0.0271	0.0236	0.0236

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Internet Appendix for: "Did Pandemic Relief Fraud Inflate House Prices?"

A. Details Regarding PPP Fraud Measures

The loan-level indicators of suspicious PPP loans were developed and validated by Griffin, Kruger, and Mahajan (2023a). To aid in understanding, we briefly describe each of the primary indicators here. For further details, we refer the reader to Griffin, Kruger, and Mahajan (2023a).

Business Registry Flag

Businesses organized as corporations, S-corporations, and LLCs are required to file an article of incorporation or LLC filing with a state, either as a domestic company in their home state or as a foreign company in another state. Further, the SBA required businesses to be "in operation on February 15, 2020…[and] not permanently closed." Based on these requirements, the following conditions are checked for all corporation, S-corporation, and LLC borrowers:

(i). Is there a matching business in the business registry data? ("Missing Business")

(ii). Was the business dissolved and inactive before being approved for a PPP loan? ("Dissolved Business")

(iii). Is the earliest incorporation or initial filing date for the business after February 15, 2020? ("Late Incorporation/Filing")

These three subflags are combined to form an overall business registry flag.

Multiple Loans Flag

While a business owner may have multiple businesses registered to the same address, the presence of multiple loans at an individual residential address during the same draw is also a potential sign of fictitious operations. This flag identifies residential (i.e., nonbusiness, noncentral) standardized addresses with three or more loans within the same draw. Note that the flag uses a cutoff of three loans instead of two loans in order to be more conservative.

High Implied Compensation Flag

PPP loan size is limited to 2.5 times a business's average monthly payroll expenses, including up to $100,000 in annual compensation per employee. PPP loan applications report how many employees the business has based on the same time period used to calculate average payroll expenses (2019 in most cases). Based on loan size and number of reported employees, one can impute implied average annual compensation. Loans for which the implied compensation per job reported is more than three times the industry-CBSA average compensation/receipts are flagged by this indicator.

EIDL Advance Jobs > PPP Jobs Flag

Concurrently with the PPP, the SBA provided businesses and individuals with the ability to receive a forgivable EIDL advance of up to $10,000. For all EIDL advances issued in 2020, the advance amount was calculated as $1,000 per employee (up to the $10,000 maximum). Thus, there was an incentive for borrowers to inflate the number of jobs reported on their EIDL applications. This flag identifies borrowers that appear to have manipulated the number of employees reported on their EIDL applications to exploit this incentive. To make it less likely that differences are driven by reporting or timing differences in the number of employees reported, only loans where the individual claimed three or more jobs extra on their EIDL applicant as compared to their PPP application are flagged by this indicator.

Validation of PPP Fraud Measures

See Griffin, Kruger, and Mahajan (2023a) for validation of the fraud measures including secondary measures of fraud, independent external measures. A Congressional investigation into PPP fraud (see Congressional report here), also validates the findings of Griffin, Kruger, and Mahajan (2023a), particularly with respect to the high fraud rates they find for many FinTech lenders.

B. Additional Details on Social Connections Instrument

Drawing on Griffin, Kruger, and Mahajan (2023b), we instrument for local PPP fraud per capita with fraud per capita in distant counties that are socially connected. Specifically, we define social proximity to fraud as:

$$Social\ Proximity_i = \frac{\sum_{j \neq i} SCI_{i,j} \times Flagged\ Per\ Capita_j}{\sum_{j \neq i} SCI_{i,j}}$$

where $SCI_{i,j}$ is the social connectedness index between zip code i and j and $Flagged\ Per\ Capita_j$ is the ratio of flagged loans in zip code j to the population of zip code j. The Social Connectedness Index (SCI) is from (Bailey et al., 2020), which is calculated as $SCI_{i,j} = \frac{Connections_{i,j}}{FB\ Users_i \times FB\ Users_j}$ where $Connections_{i,j}$ is the total number of Facebook friendship links between Facebook users living in zip code i and Facebook users living in zip code j and $FB\ Users_i$ is the number of Facebook users in zip code i.

Social proximity to fraud is essentially the weighted average fraud per capita in connected zip codes where weights are based on the strength of connections between the zip codes in pairwise Facebook data. Restricting zip codes to different distances (e.g., not in the same county, over 100 miles apart, over 500 miles apart) can generate different versions of social proximity to fraud.

Griffin, Kruger, and Mahajan (2023b) show that social proximity to suspicious lending strongly predicts the rate of suspicious lending in a zip code. For social connections to be used to form a valid instrument, the relevance condition is that the social proximity to suspicious lending of a zip code should predict the rate of suspicious lending in the zip code, and the exclusion restriction is that social proximity to suspicious lending only affects house prices in the zip code through suspicious government spending in the zip code. While the exclusion restriction cannot be directly tested, for reasons described in the text, any direct effect, or indirect effect through omitted variable, of social proximity to fraud would need to have a similar effect across different distances. In contrast, most factors that might influence house prices (like migration, local omitted variables, or regional shocks) likely decrease with distance.

Table IA.27 shows results using each of the instruments based on social connections in the non-overlapping rings separately. Table IA.28 shows that the results are also robust to using social connections outside counties, CBSAs, and states, as well as a combination of all three. Table IA.29 shows the results based on the percentage of loans flagged instead of flagged per capita. Figure IA.22 replicates Panel B of Figure 2 with IV estimates. The resulting cumulative effect is first significant in August 2020, and then strengthen until June 2022 when they peak. Afterwards, the effects are slowly diminished and as of April 2024, are approximating 60% of the effect at its peak. First stage regressions are reported in Table IA.30, and reduced form regressions are reported in Table IA.31. Table IA.32 reports estimates without population weighting.

Finally, Table IA.33 shows estimates based on the *Flagged Composite Per Capita* measure. We instrument for a zip code's *Flagged Composite Per Capita* using the same sets of social connections as in Table 3 and find an even larger effect on house prices. In column (1), which is based on social connections outside the zip code's CBSA, a one standard deviation increase in instrumented *Flagged Composite Per Capita* is associated with a 4.20 ppt increase in house prices between January 2020 and December 2021. This is a sizable 16.3% of the 25.7 ppt average increase in house prices during this period. As in Table 3, the results in all specifications are extremely similar (between 3.96 and 4.32 ppt), significant with t-statistics above 5.6, and have first stage F-stats of at least 30.

C. Supplemental Figures and Tables

Figure IA.1. Zillow Home Price Index (ZHVI)
This figure compares the monthly house price growth between the current neural network based Zillow Home Price Index (ZHVI) and the legacy ZHVI that Zillow provided before January 2023.

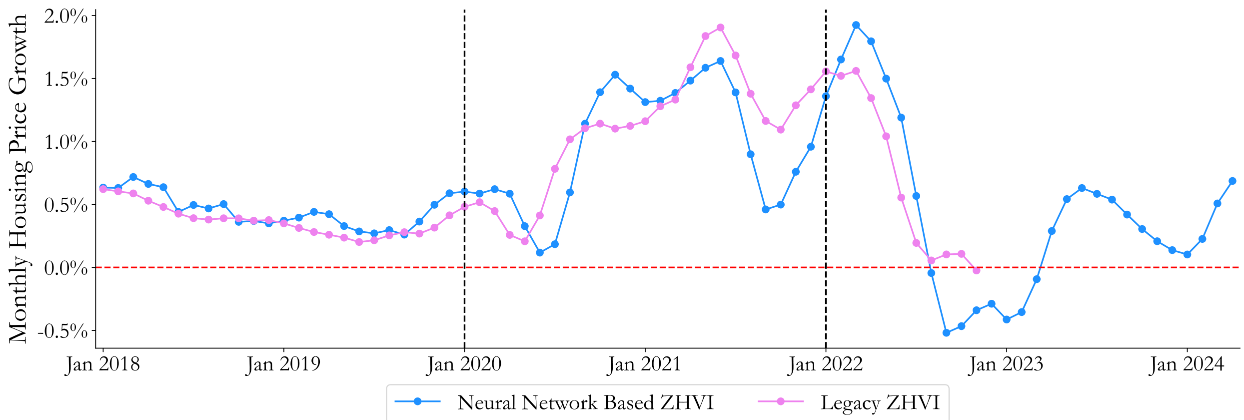
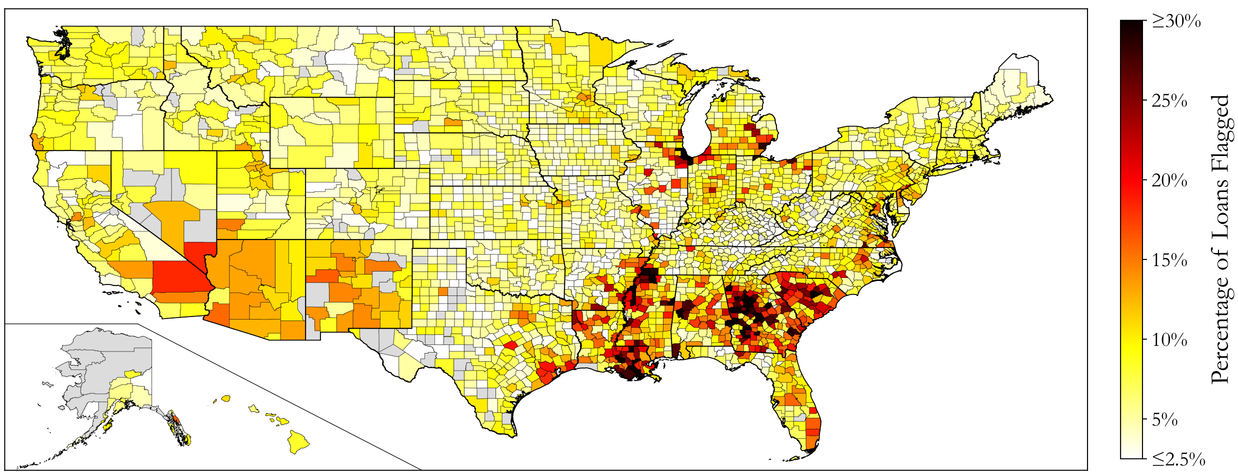


Figure IA.2. Geography of Flagged Loans

This figure shows geographic variation in the percentage of flagged loans. Panel A shows the percentage of flagged loans in each county and Panel B shows within county variation. In Panel A, counties are colored based on the color scheme shown by the bar to the right of the map and counties with fewer than 100 loans are colored grey. Panel B shows the percentage of flagged loans in each zip code on the vertical axis and the percentage of flagged loans in the corresponding county on the horizontal axis. Dots are sized based on the number of loans in the zip code. Zip codes with at least 100 loans are shown. The dashed line is a linear fit and the correlation is shown in the bottom left corner.

Panel A. Percentage of Flagged Loans, by County



Panel B. Within County Variation

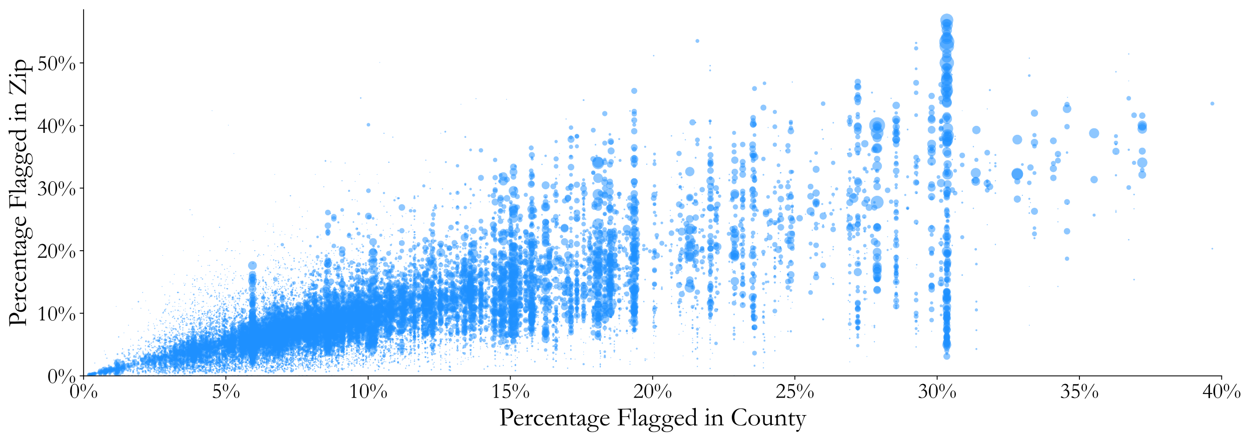


Figure IA.3. LexisNexis

This figure replicates Panel A of Figure 1 using data from LexisNexis for the sample of 150,000 loans used in Griffin, Kruger, and Mahajan (2023a). The sample consists of individual borrowers who received PPP loans in rounds 1 and 2. The LexisNexis data was collected in March 2021 and includes data on house purchases through the end of 2020. Rounds 1 and 2 of the PPP occurred in April to August 2020, with most loans occurring by the end of May. As a result, we observe at least four months of post-PPP house purchase activity for all individuals in the sample. The horizontal lines are the average monthly likelihood of an individual in each group buying a house during the pre-/post-period.

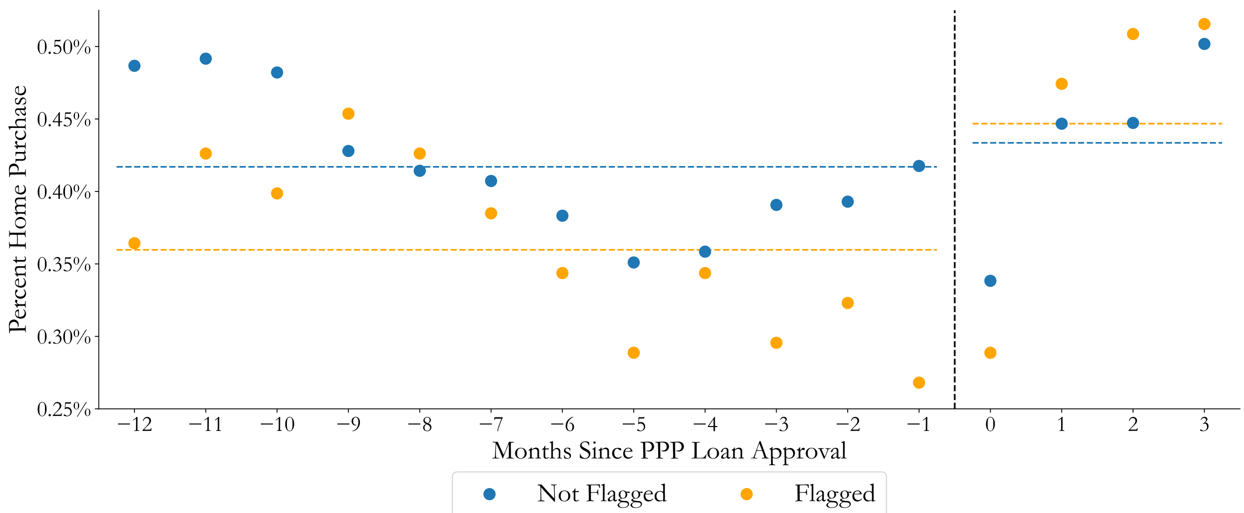


Figure IA.4. Moving

This figure shows the relationship between individuals receiving a flagged PPP loan and their likelihood of moving subsequently. Data on a sample of 150,000 loans from Melissa Data is used to determine whether an individual moved in the two years after they received their PPP loan. The sample of non-flagged loans is reweighted to match the distribution of timing of PPP loan approval of the sample of flagged loans. In the right subpanel, the dashed lines represent 95% confidence intervals.

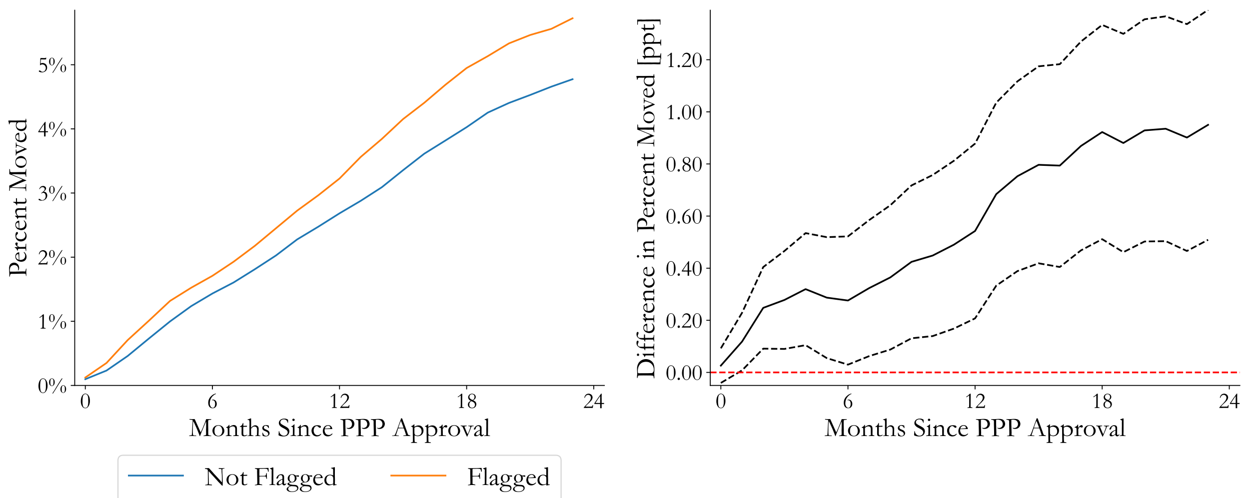
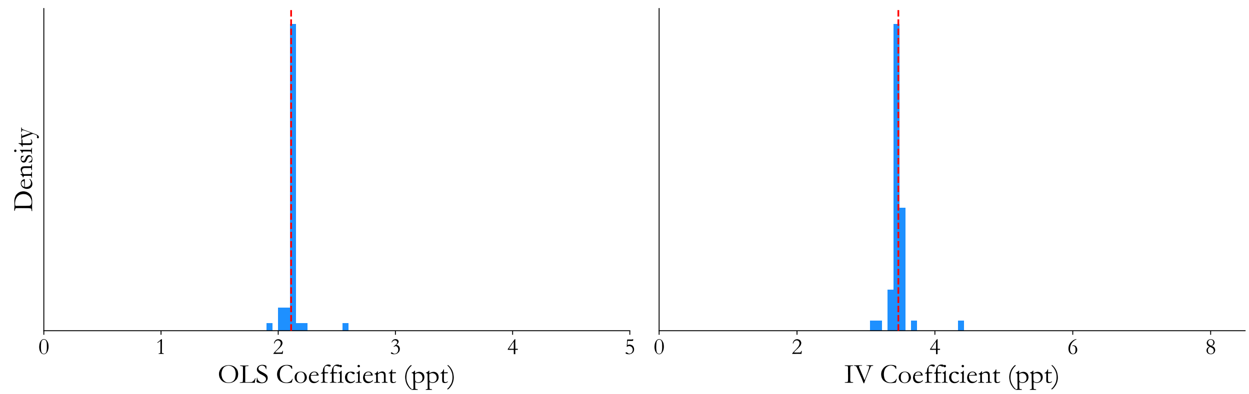


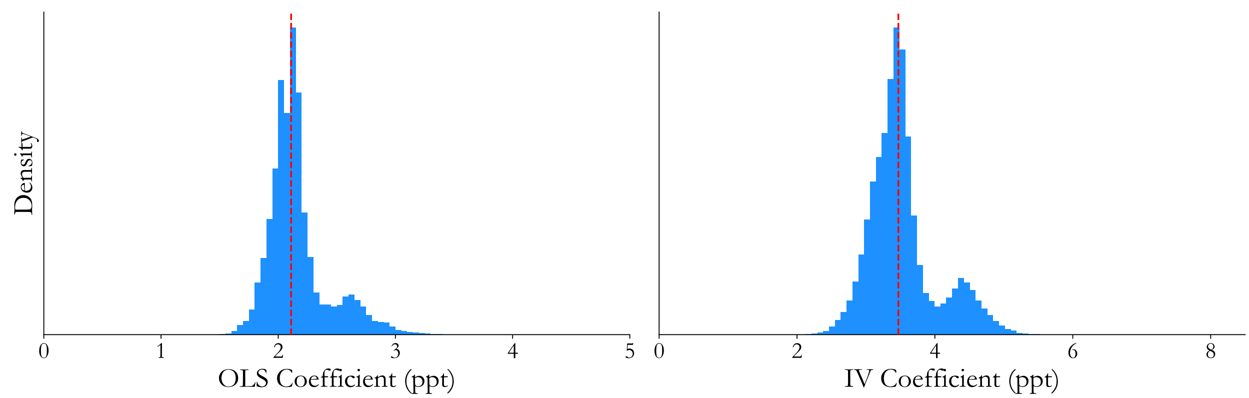
Figure IA.5. Robustness to Excluding States

This figure shows the robustness of column (1) of Table 2 (left subpanels) and column (1) of Table 3 (right subpanels) to excluding states. Panel A shows the distribution of coefficients when each state is excluded one at a time. Panels B and C show the distribution of coefficients using 100,000 bootstraps where 10 and 25 states, respectively, are randomly excluded. The red vertical lines are the coefficients using the full sample.

Panel A. Exclude Single State



Panel B. 100,000 Bootstraps with 10 States Excluded



Panel C. 100,000 Bootstraps with 25 States Excluded

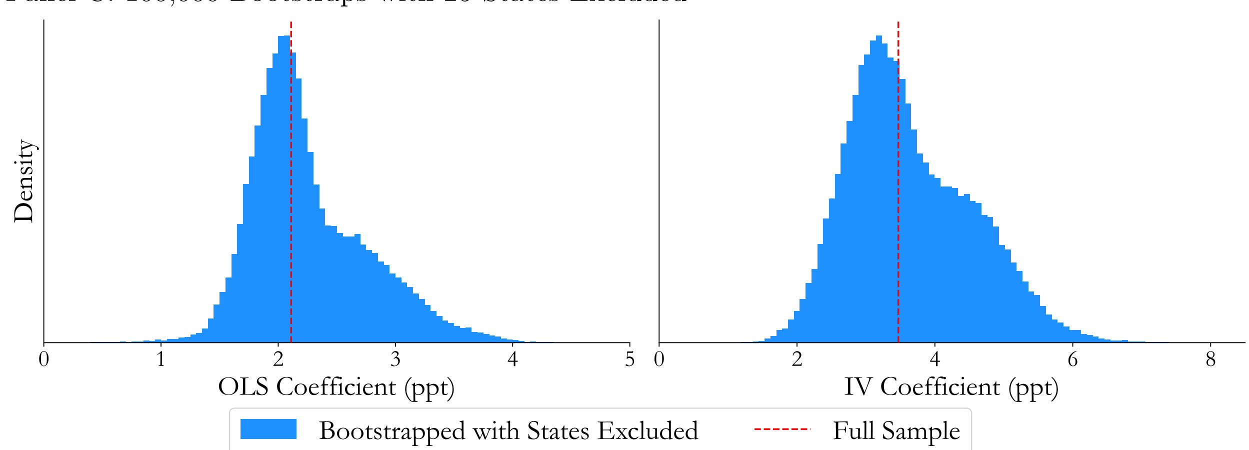
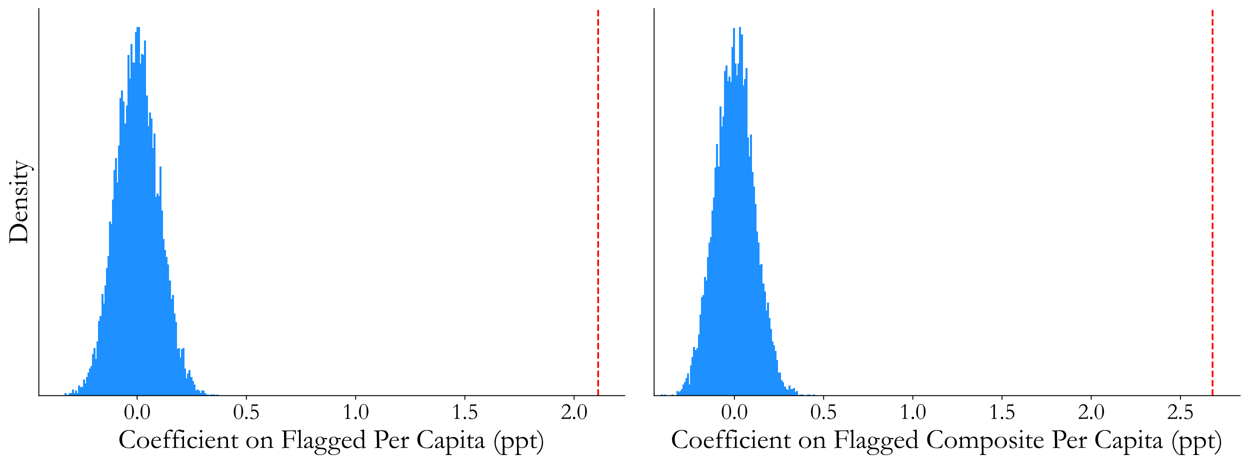


Figure IA.6. Permutation Tests

This figure shows the coefficients when the measures of suspicious lending are permuted between zip codes within the same county (Panel A) and across the nation (Panel B). 10,000 permutations are performed. The left and right subpanels show the distribution of coefficients when columns (1) and (5), respectively, of Table 2 are estimated using the permuted samples. The red vertical lines are the coefficients using the actual sample.

Panel A. 10,000 Permutation of Zip Codes Within County



Panel B. 10,000 Permutation of Zip Codes Across Nation

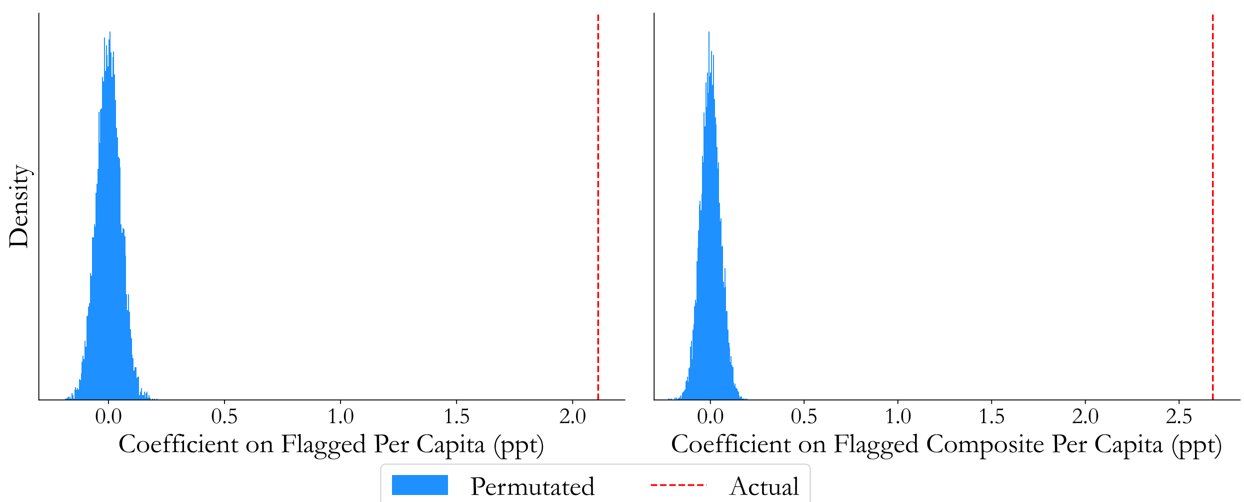
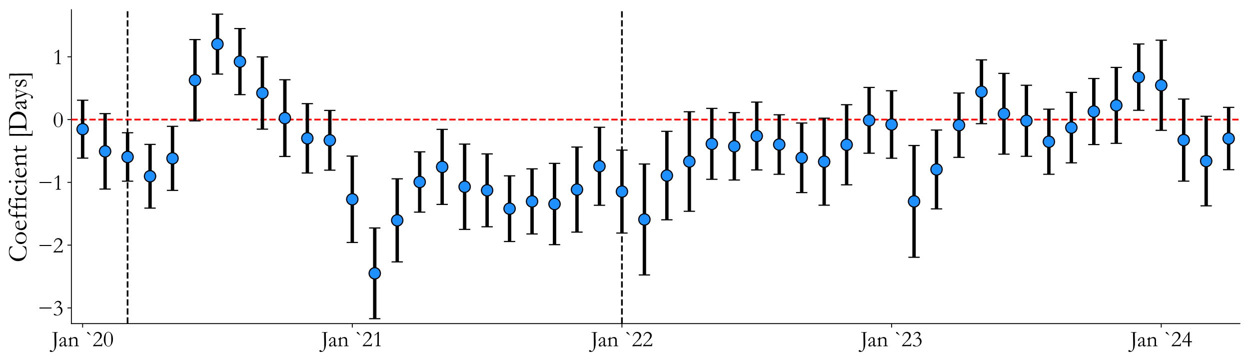
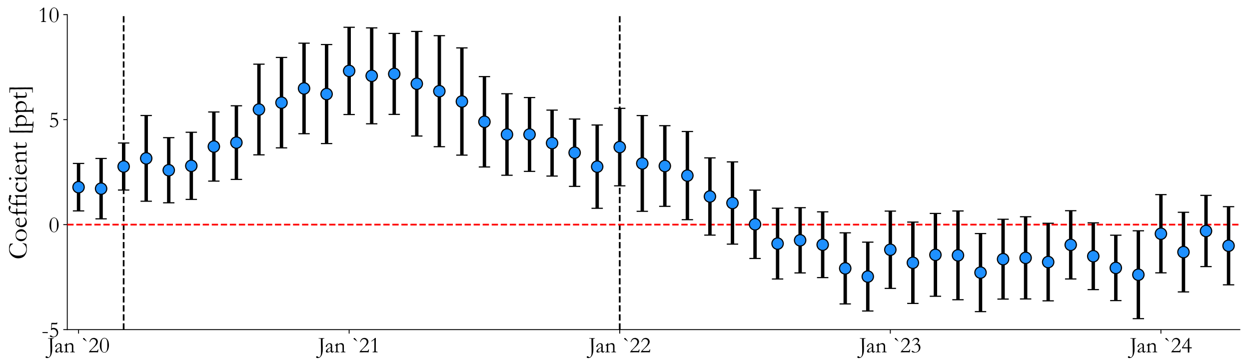


Figure IA.7. Effects on Other Housing Market Metrics
This figure shows the effect of a one standard deviation change in *Flagged Per Capita* on various housing market metrics. Panel A shows the effect on median days on the market, Panel B on unique viewers per property (relative to typical property across US), and Panel C on the Realtor.com Market Hotness Index (ranges from 0 to 100). All three metrics are from Realtor.com. The dependent variable is the difference between the metric during the given month and its average in the same month-of-year in 2018 and 2019. The controls and fixed effects are the same as Column (1) of Table IA.19. The error bars correspond to 95% confidence intervals based on standard errors that are clustered at the county level.

Panel A. Median Days on Market



Panel B. Unique Viewers Per Property



Panel C. Realtor.com Market Hotness Index

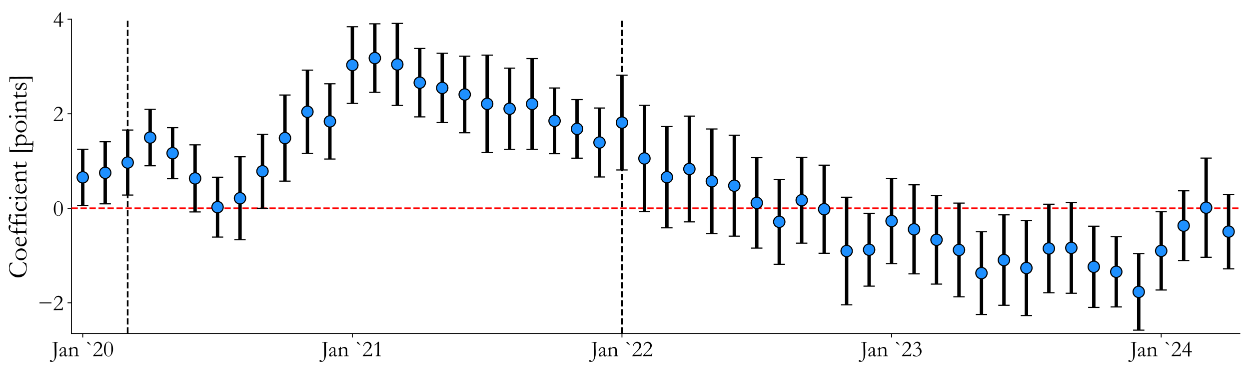


Figure IA.8. Effects Over Time, Monthly Non-Cumulative Effect

This figure shows the effect of a one standard deviation change in *Flagged Per Capita* on monthly house price growth. We estimate the following regression for each month plot β_t:

$$HPGrowth_{i,t} = \beta_t \times FlaggedPerCapita_i + County_k + LoansPerCapitaPercentile_i + HPGrowth\text{2018-19}Percentile_i + Controls_i$$

where i is a zip code, k is the county that zip code i is in, and t is a month. We control for overall PPP loans per capita and house price growth in 2018-19 using percentile indicator variables to allow for nonlinearity and include county fixed effects. The controls included are vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. All proposed variables are standardized to have a mean of 0 and standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being population of the zip code in 2019. Only zip codes for which all proposed variables can be determined are used. The error bars represent 95% confidence intervals based on standard errors clustered by county.

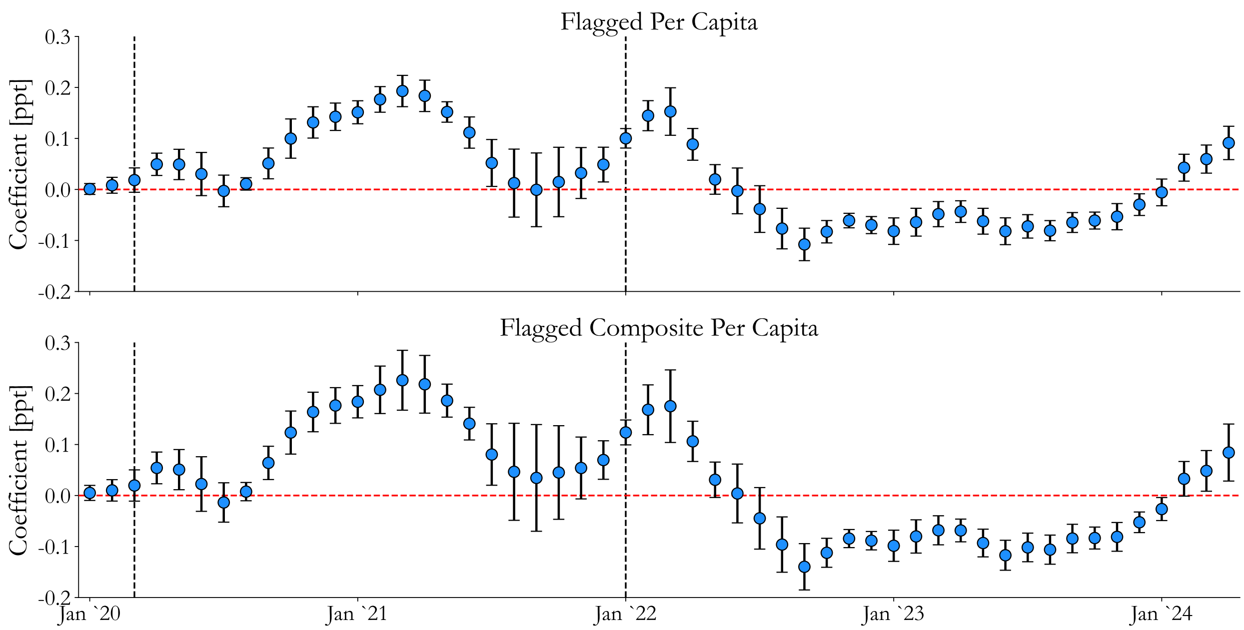


Figure IA.9. Housing Price Growth, Without Synthetic Controls
This figure shows the house price growth from January 2018 to September 2022 in the bottom and top quartile of zip codes based on flagged per capita. Zip codes are split into quartiles within county. Counties where the difference between the 75th and 25th percentile of flagged per capita is at least half as large as the standard deviation across the entire nation are included. The dashed lines are 95% confidence intervals.

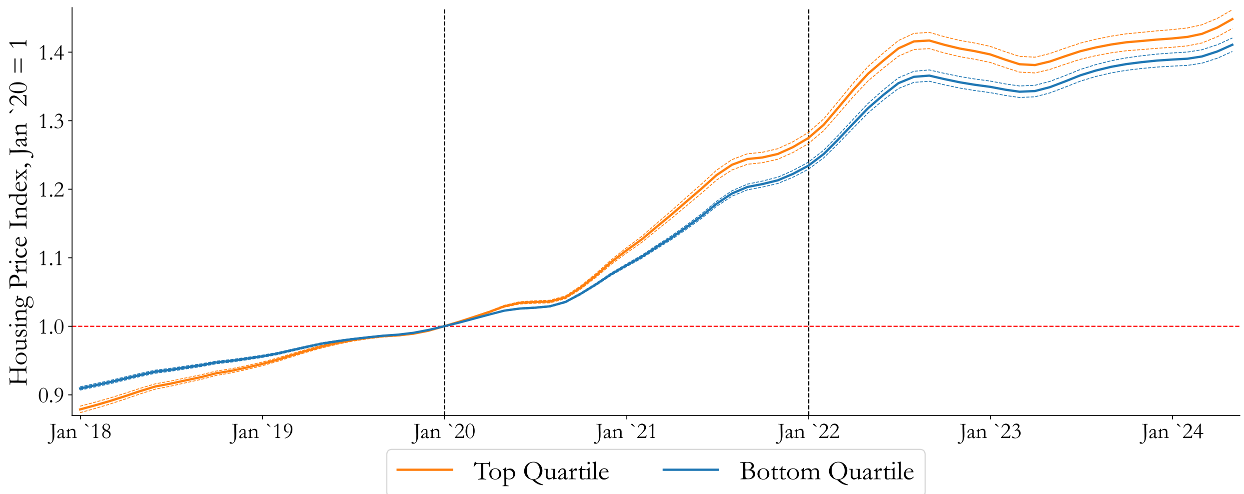
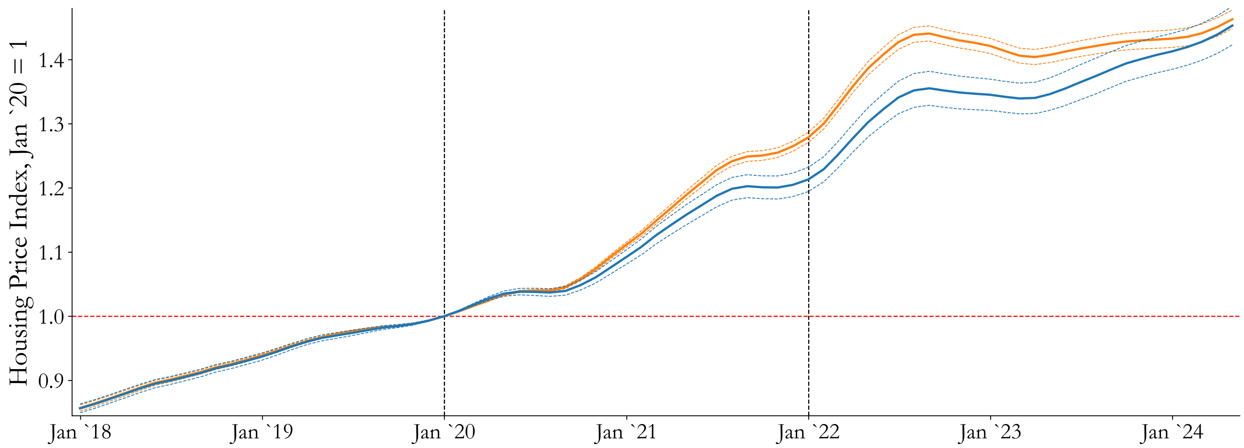
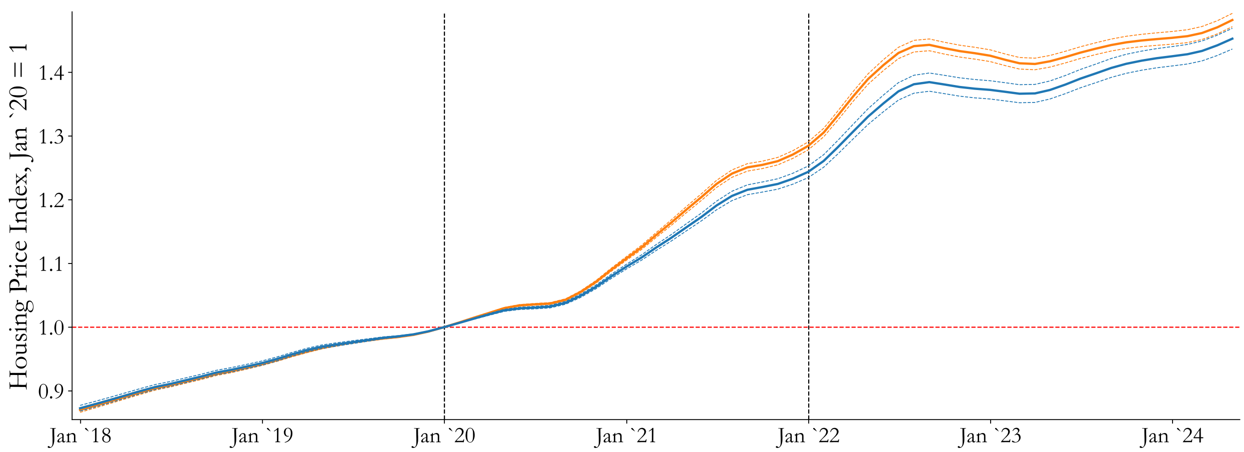


Figure IA.10. Synthetic Control, Alternative Measures
This figure replicates Figure 3 using various measures of suspicious lending. The dashed lines are 95% confidence intervals.

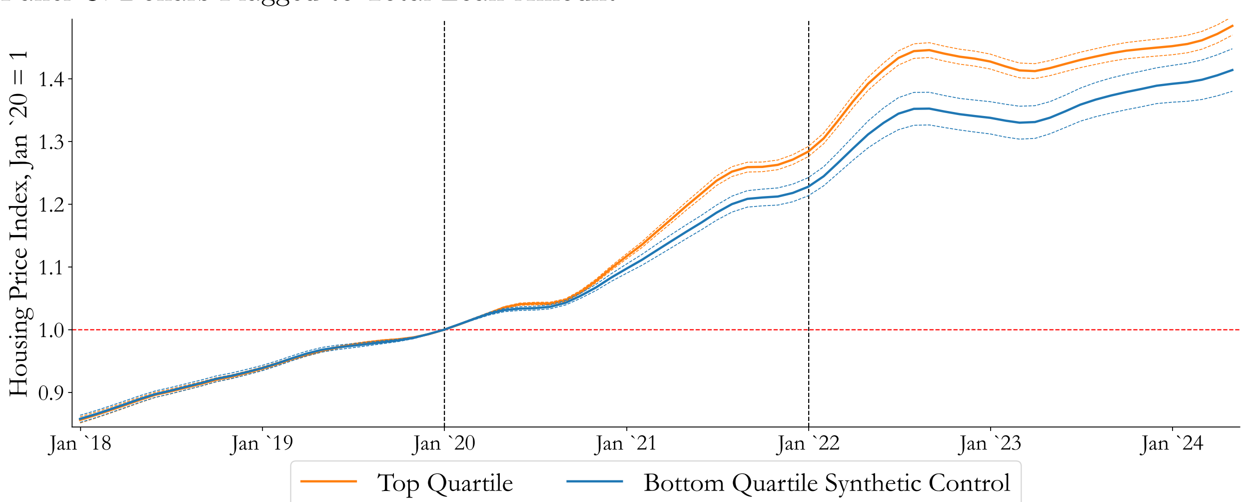
Panel A. Suspicious Composite Per Capita



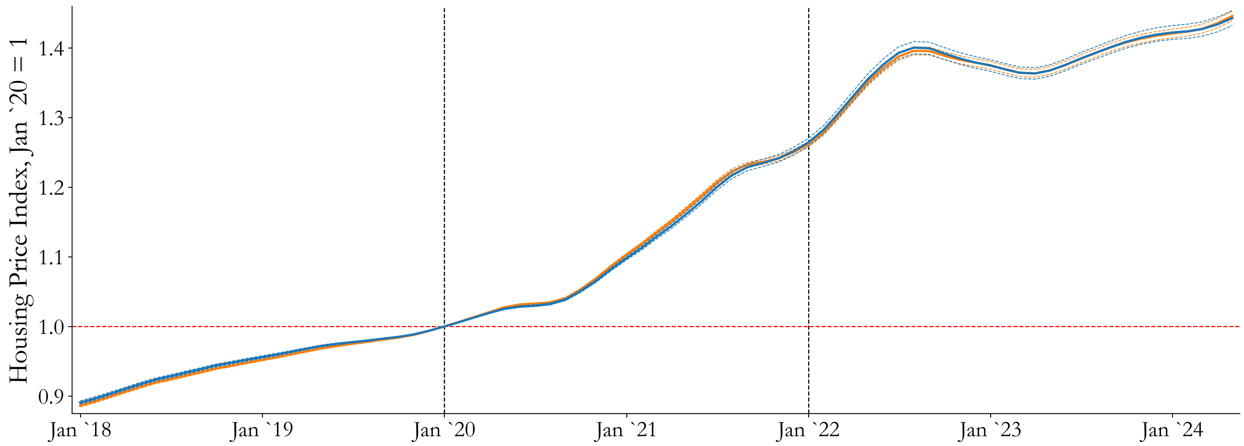
Panel B. Dollars Flagged to Total Income



Panel C. Dollars Flagged to Total Loan Amount



Panel D. Total Loan Amount to Total Income



Panel E. Loans Per Capita

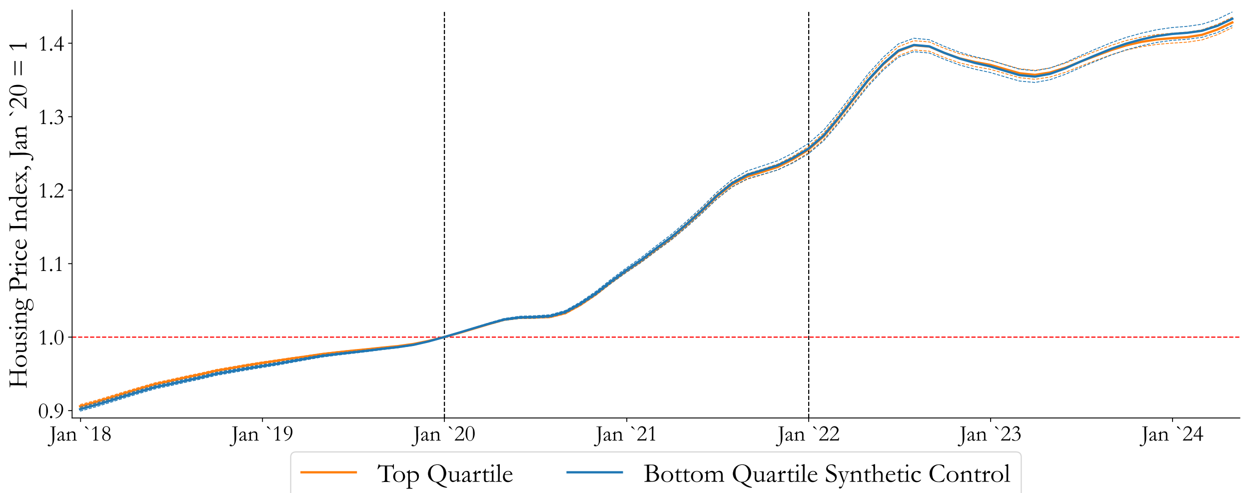
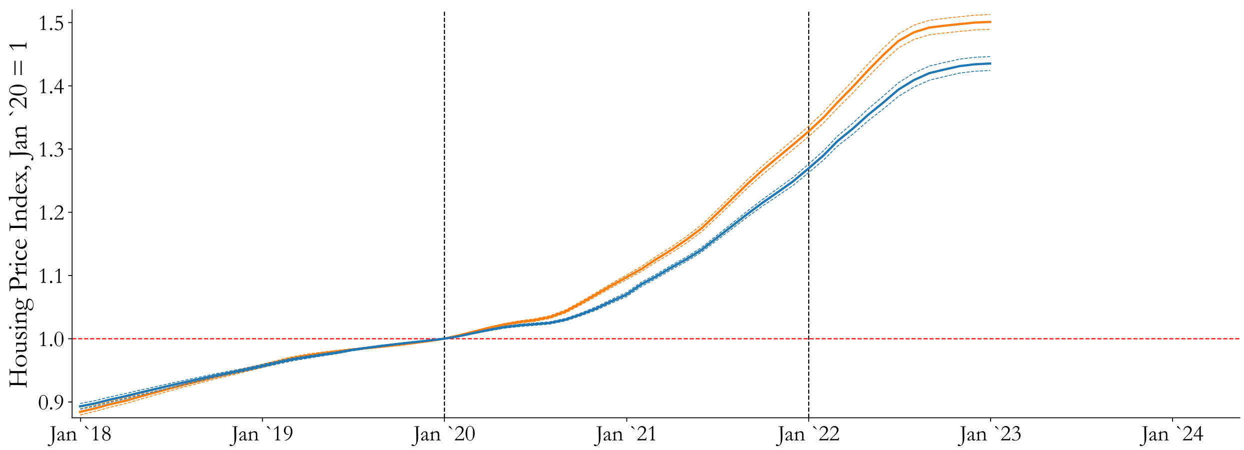


Figure IA.11. Matching
This figure shows the effect of suspicious lending on housing price growth using a matching methodology. In particular, each zip code in the top quartile of *Flagged Per Capita* is matched to a zip code in the same CBSA that is in the bottom quartile of *Flagged Per Capita*. Zip codes are split into quartiles within CBSA. CBSAs where the difference between the 75th and 25th percentile of Flagged Per Capita within the CBSA is at least half as large as the standard deviation across the entire nation and have at least 10 zip codes are included. The dashed lines are 95% confidence intervals based on standard errors clustered by zip code to account for the same zip code being used as a control multiple times. Panel A is based on the legacy Zillow Home Price Index ZHVI that Zillow generated pre-2023 and Panel B is based on the neural network ZHVI that Zillow currently generates.

Panel A. Pre-2023 Legacy ZHVI Data



Panel B. Current Neural Network ZHVI Data

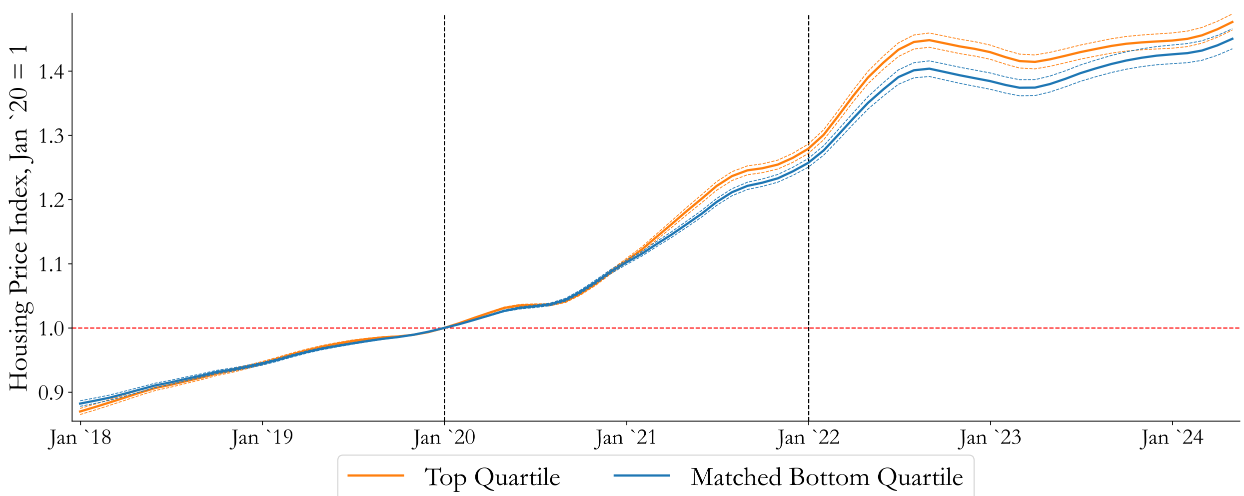


Figure IA.12. Effect of Other Proposed Variables on Housing Prices, Alternative Specifications

This figure replicates Panel A of Figure 4 using alternative specifications. Panel A shows the results when OLS is used instead of WLS. Panel B shows the results when each of the proposed variables is calculated as a weighted average (weighted by population) over a five-mile radius around each zip code. Panel C only includes county fixed effects. The error bars correspond to 95% confidence intervals based on standard errors that are clustered at the county level. Univariate (multivariate) coefficients are shown in blue (pink). Positive (negative) coefficients are shown as squares (triangles).

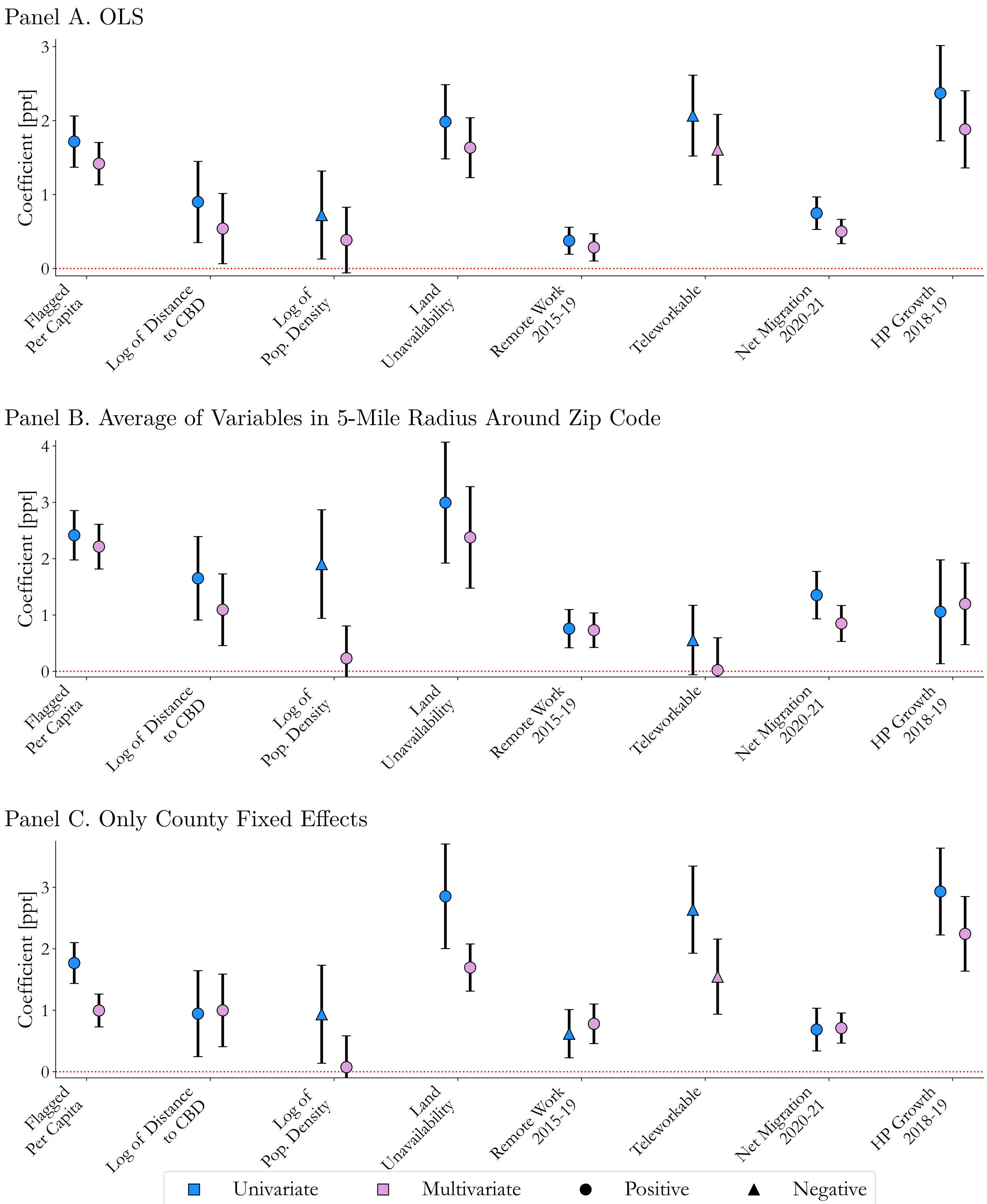
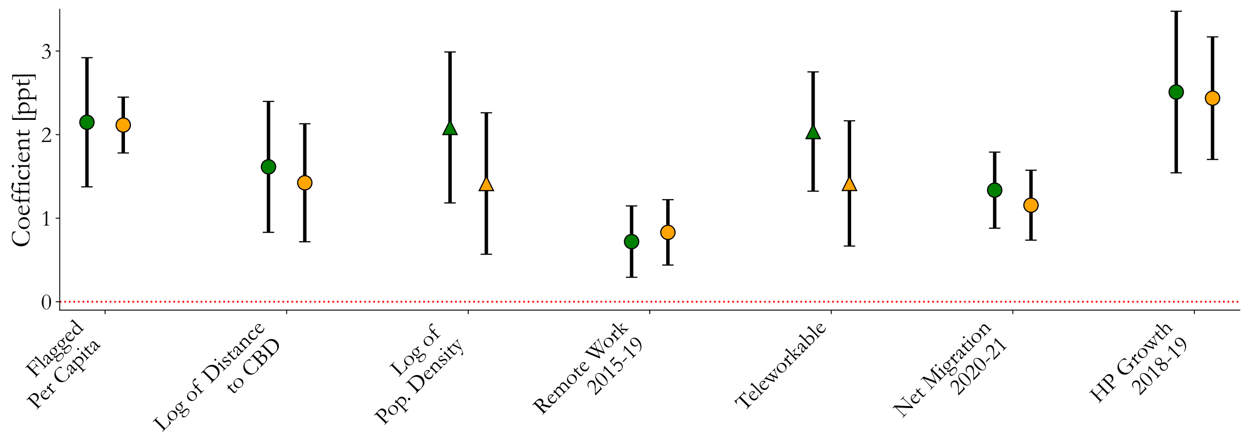
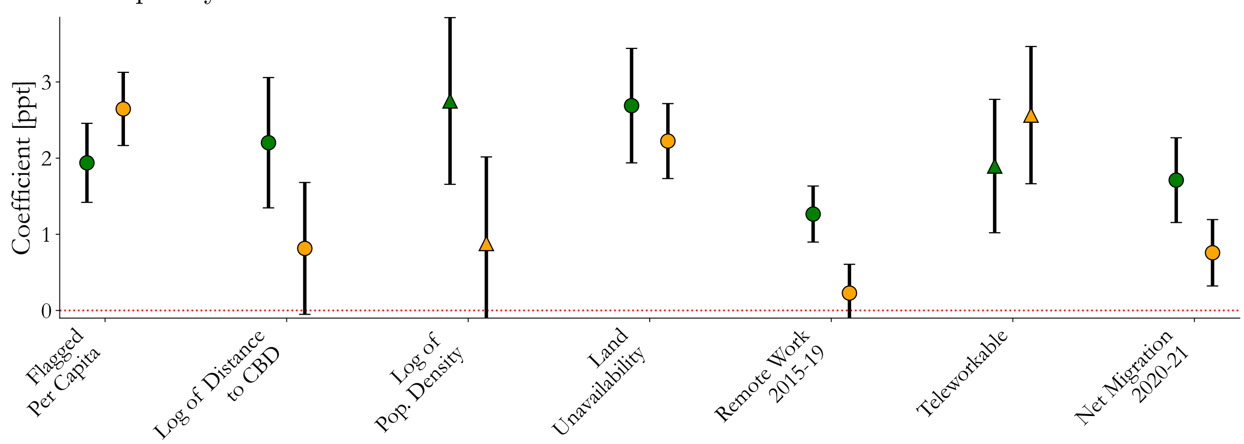


Figure IA.13. Effect of Other Proposed Variables on Housing Prices, Various Split
This figure replicates the univariate regressions shown in Panel A of Figure 4 for sample splits based on land unavailability (Panel A), house price growth in 2018-19 (Panel B), house price as of January 2020 (Panels C and D), COVID mortgage forbearance (Panel E), 2000-19 beta with national house prices (Panel F), and 2000-19 house price volatility. Splits are at the national median values of the variables for Panels A, B, C, E, F, and G and at the county median value for Panel D. The error bars correspond to 95% confidence intervals based on standard errors that are clustered at the county level. Coefficients for the below (above) median sample are shown in green (orange). Positive (negative) coefficients are shown as squares (triangles).

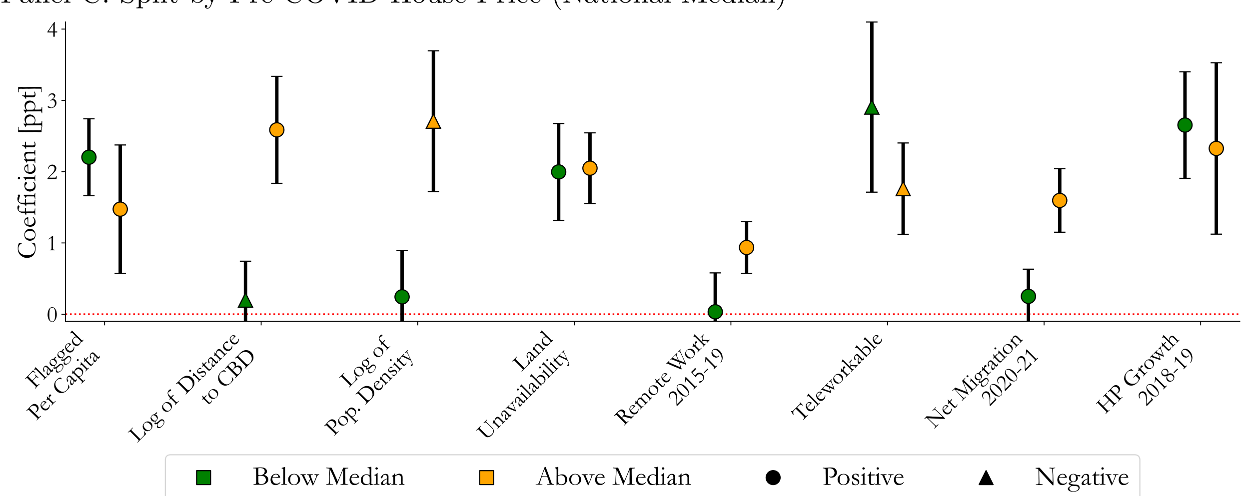
Panel A. Split by Land Unavailability



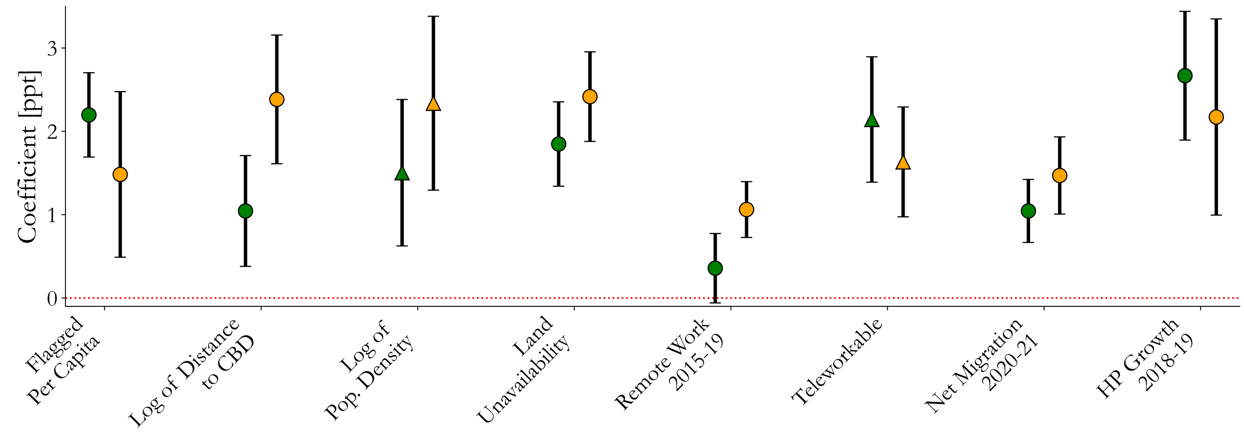
Panel B. Split by Previous House Price Growth



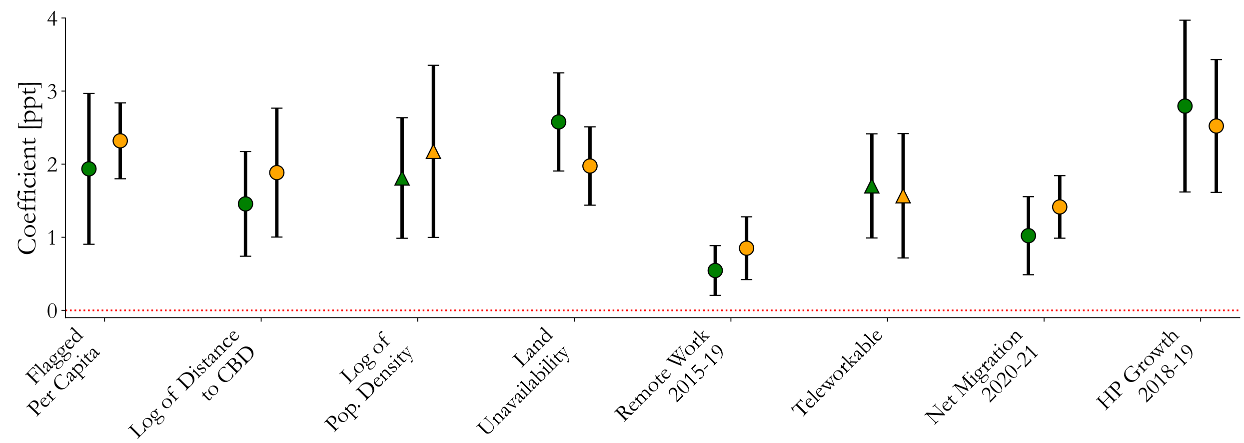
Panel C. Split by Pre-COVID House Price (National Median)



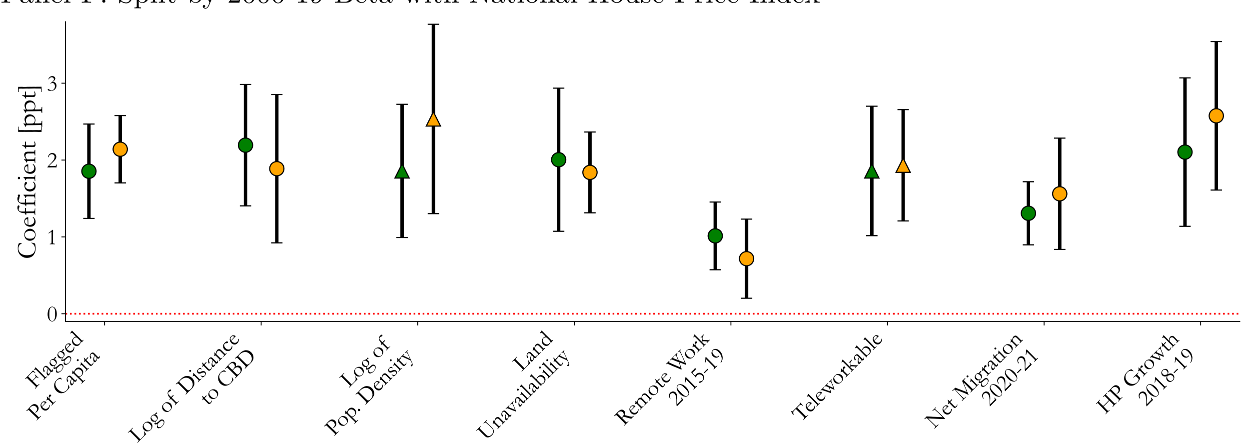
Panel D. Split by Pre-COVID House Price (County Median)



Panel E. Split by COVID Mortgage Forbearance



Panel F. Split by 2000-19 Beta with National House Price Index



Panel G. Split by 2000-19 House Price Volatility

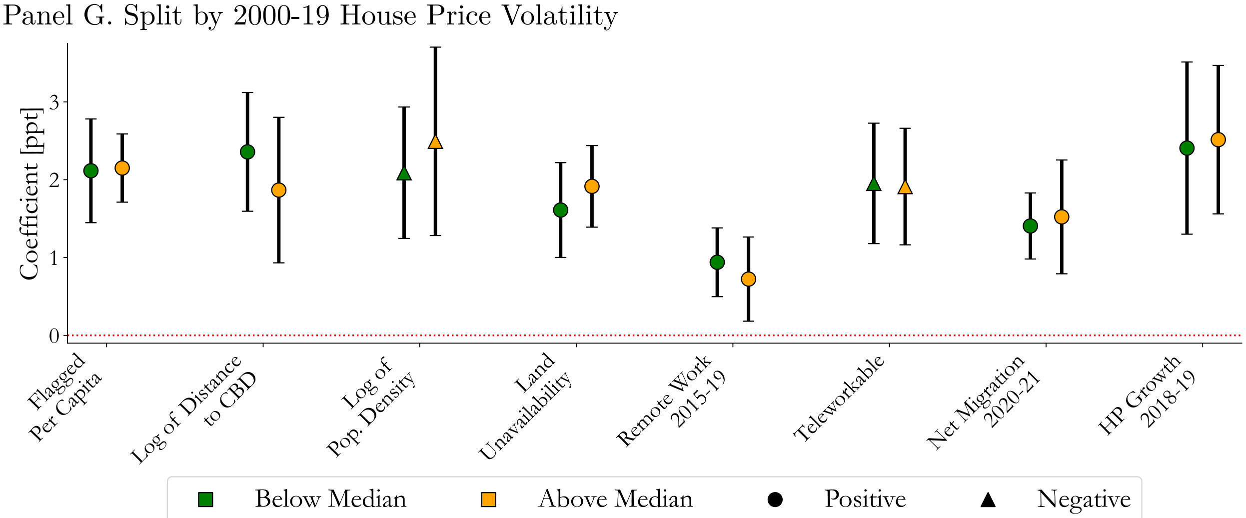


Figure IA.14. Effects of Other Proposed Variables on House Price, Over Time
This figure shows the effect of each proposed variable over time. For each proposed variable, we estimate the following regression for each month plot β_t:

$$CumulativeHPGrowth_{i,t} = \beta_t \times Variable_i + County_k + LoansPerCapitaPercentile_i + HPGrowth2018\text{-}19Percentile_i + Controls_i$$

where i is a zip code, k is the county that zip code i is in, and t is a month. We control for overall PPP loans per capita and house price growth in 2018-19 using percentile indicator variables to allow for nonlinearity and include county fixed effects. The controls included are vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. All proposed variables are standardized to have a mean of 0 and standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being population of the zip code in 2019. Only zip codes for which all proposed variables can be determined are used. The error bars represent 95% confidence intervals based on standard errors clustered by county.

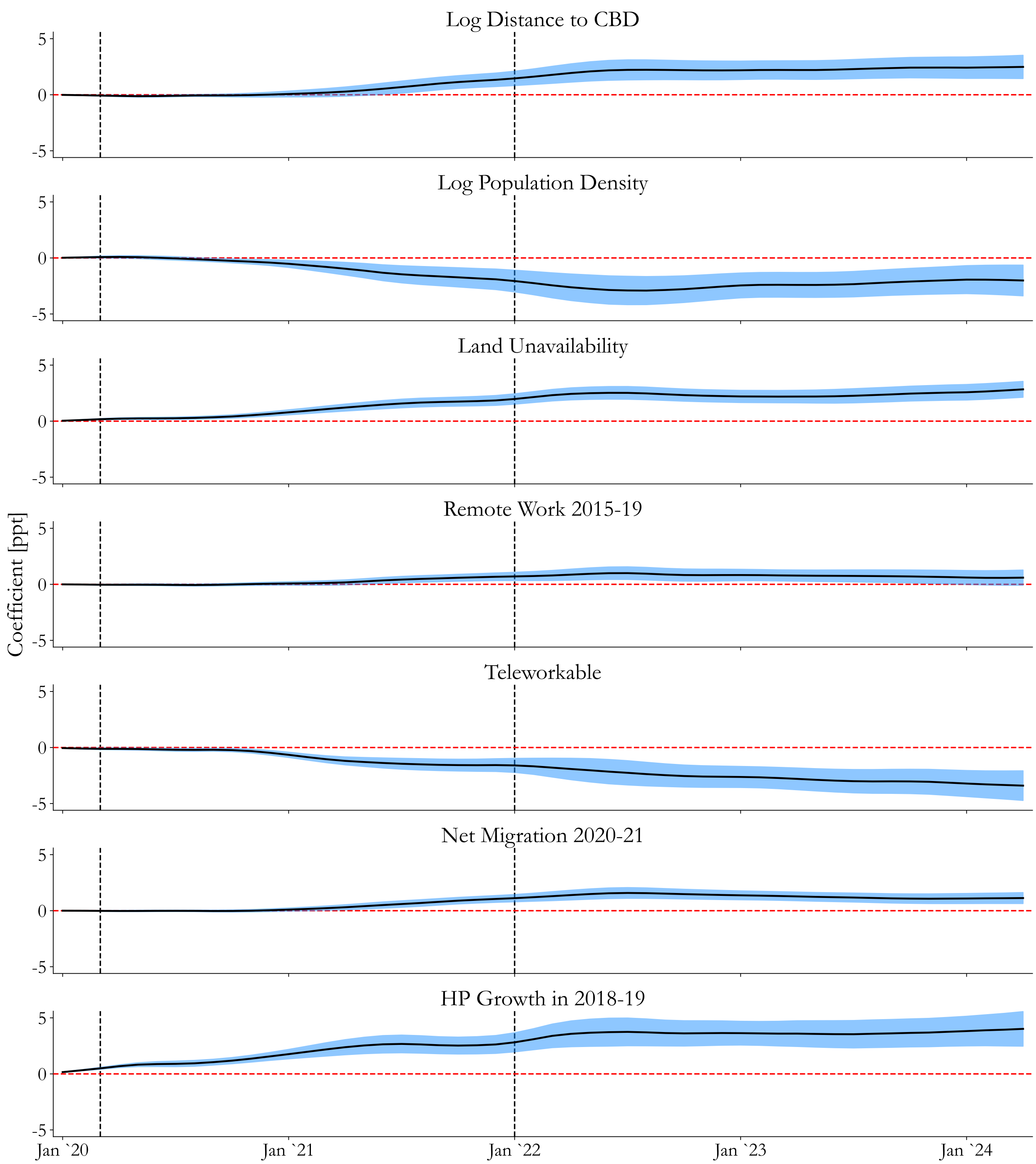


Figure IA.15. Effects of Other Proposed Variable Over Time, Monthly Non-Cumulative Effect

This figure shows the effect of a one standard deviation change in each proposed variable on monthly house price growth. For each proposed variable, we estimate the following regression for each month plot β_t:

$$HPGrowth_{i,t} = \beta_t \times Variable_i + County_k + LoansPerCapitaPercentile_i + HPGrowth2018\text{-}19Percentile_i + Controls_i$$

where i is a zip code, k is the county that zip code i is in, and t is a month. We control for overall PPP loans per capita and house price growth in 2018-19 using percentile indicator variables to allow for nonlinearity and include county fixed effects. The controls included are vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. All proposed variables are standardized to have a mean of 0 and standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being population of the zip code in 2019. Only zip codes for which all proposed variables can be determined are used. The error bars represent 95% confidence intervals based on standard errors clustered by county.

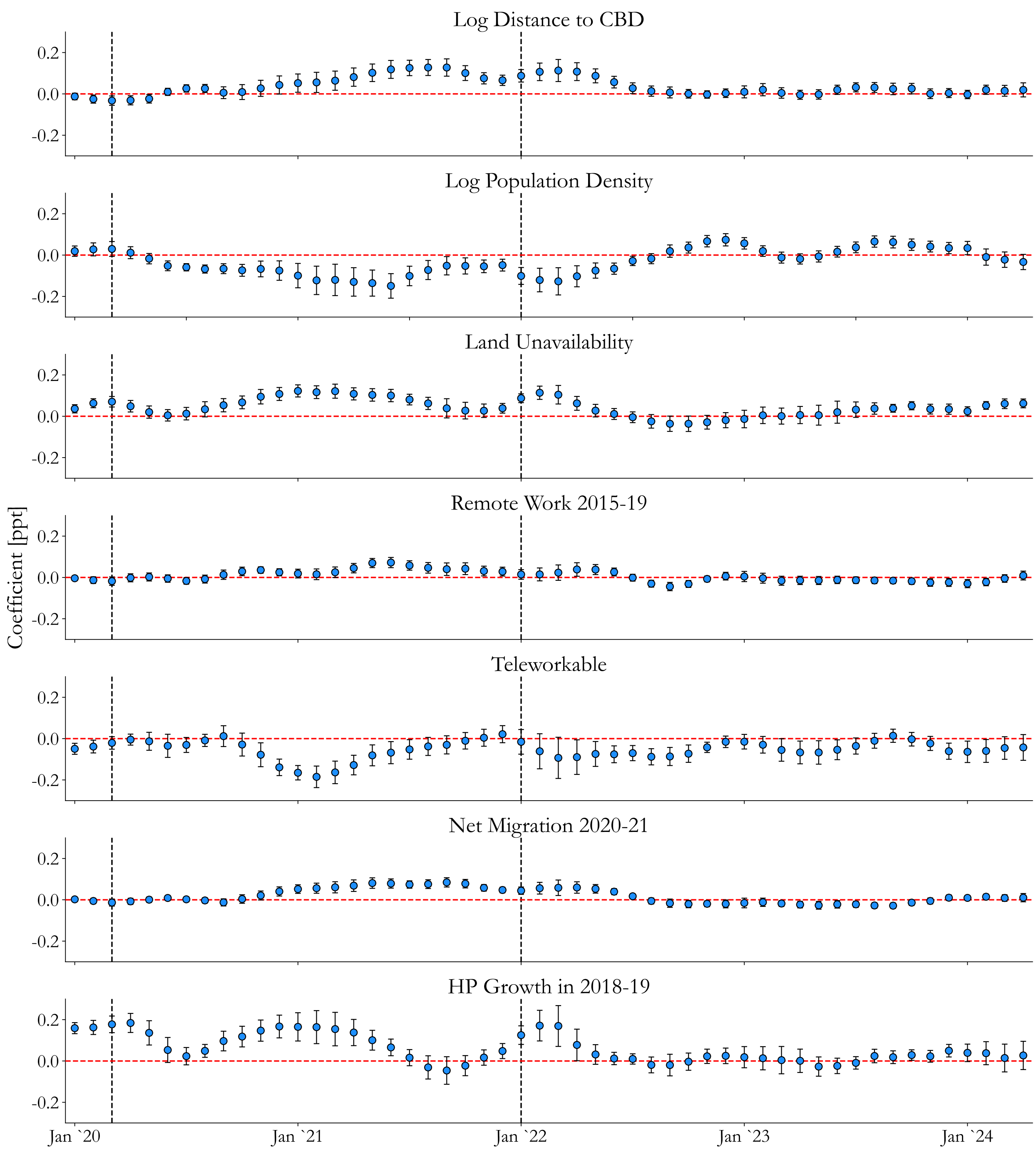
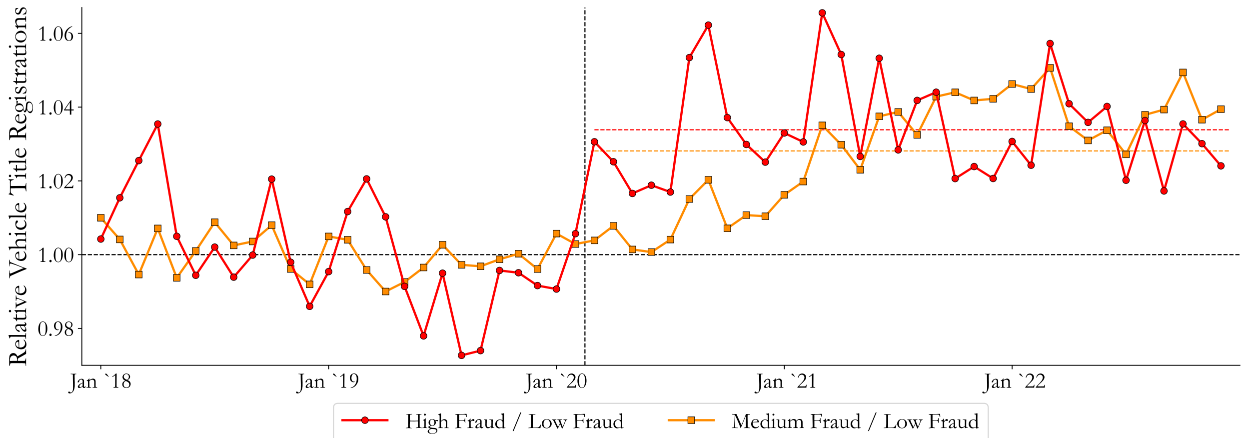


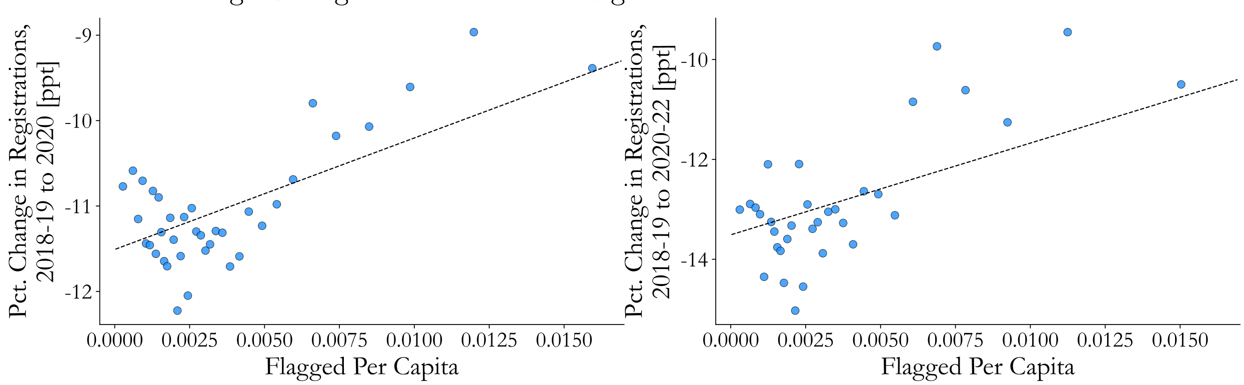
Figure IA.16. Effect on Vehicle Purchases

This figure expands on Panel A of Figure 5. Panel A examines the differences in vehicle title registrations across terciles of *Flagged Per Capita* over time. Zip codes are split into terciles within each county and each zip code's monthly number of vehicle title registrations is normalized by the average number of registrations in the zip code from January 2018 to February 2020. The series in red (orange) shows the ratio of average normalized registrations in the top (middle) tercile of zip codes to average normalized registrations in the bottom tercile of zip codes. Panel B replicates the right subpanel of Figure 5, Panel A over different periods. Panel C examines heterogeneity in the relationship between vehicle registrations and *Flagged Per Capita* across splits on various demographics, in particular columns (1) and (4) of Table 6 are re-estimated for zip codes with below and above the median value of the demographic. To have a nationally representative estimate, all panels are weighted by the zip code's population as of 2019.

Panel A. Vehicle Title Registrations Across Terciles of PPP Fraud



Panel B. Percentage Change in Vehicle Title Registration



Panel C. Heterogeneity Across Demographics

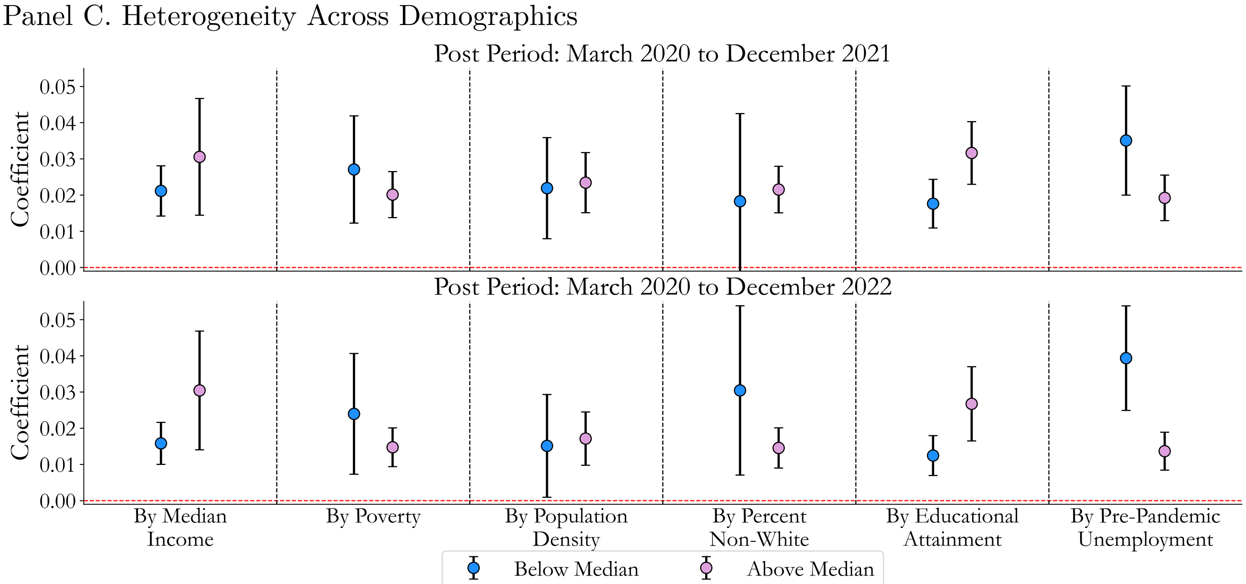


Figure IA.17. Effect on Vehicle Ownership Based on American Community Survey
THis figure shows the effects of suspicious lending on vehicle ownership using data from the American Community Survey (ACS). Data from 2015 to 2022 at the cenus tract-year level is used. The left subpanel examine the effects of a one standard deviation change in the number of flagged PPP loans per capita and includes census tract and year × county fixed effects. The error bars in the left subpanel represent 95% confidence intervals based on standard errors clustered by county. The right subpanel examine the within-county effects of flagged PPP loans per capita. To have a nationally representative estimate, both subpanels use weighted least squares (WLS) regressions with the weight being the census tract's population as of 2019.

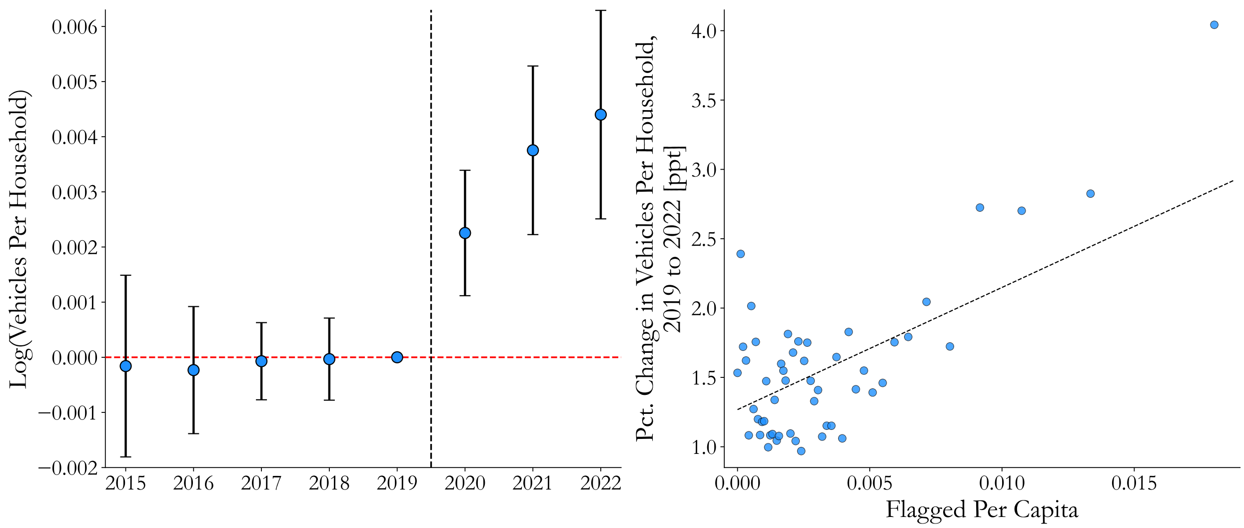


Figure IA.18. Effect on Auto Debt
This figure shows the effect of a one standard deviation change in *Flagged Per Capita* on the percentage of individuals with auto debt. Data on the percentage of individuals with auto debt is from the Federal Reserve Bank of Philadelphia's Consumer Credit Explorer. The data is provided for each MSA at a quarterly frequency. MSA and quarter fixed effect are included. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the MSA's population as of 2019. The error bars correspond to 95% confidence intervals based on standard errors that are double clustered by MSA and quarter.

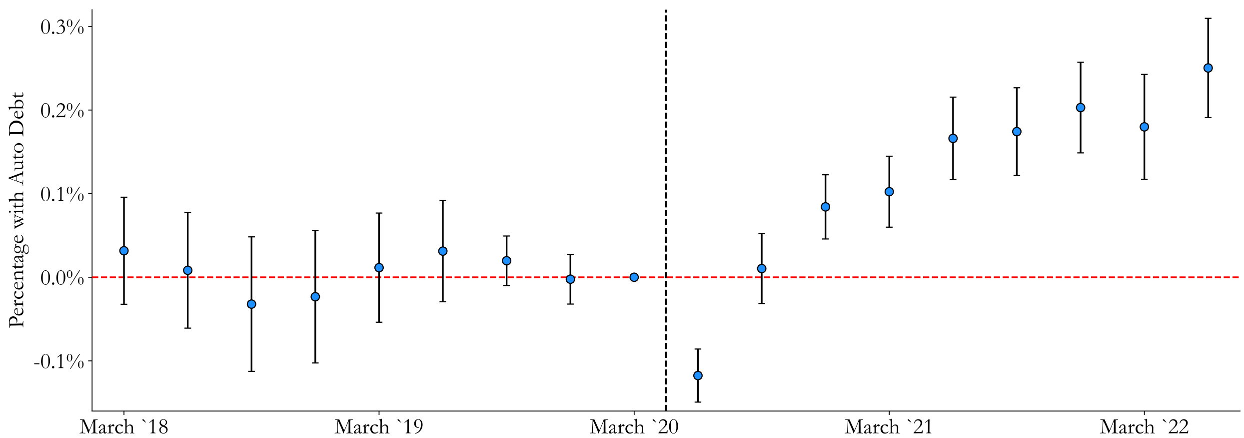


Figure IA.19. Heterogeneity in Effect on Consumer Spending by Demographics
This figure shows heterogeneity in the effects of suspicious lending on consumer spending across demographics. Annual data for 2017 to 2022 at the census tract level from Mastercard's Center for Inclusive Growth is used. Mastercard ranks each census tract's consumer spending per capita each year in the national distribution and releases the percentile rank of the tract. The demographic splits are made at the median value of the demographic across all census tracts. Census tract and year × county fixed effects are included. The error bars represent 95% confidence intervals based on standard errors clustered by county. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the census tract's population as of 2019.

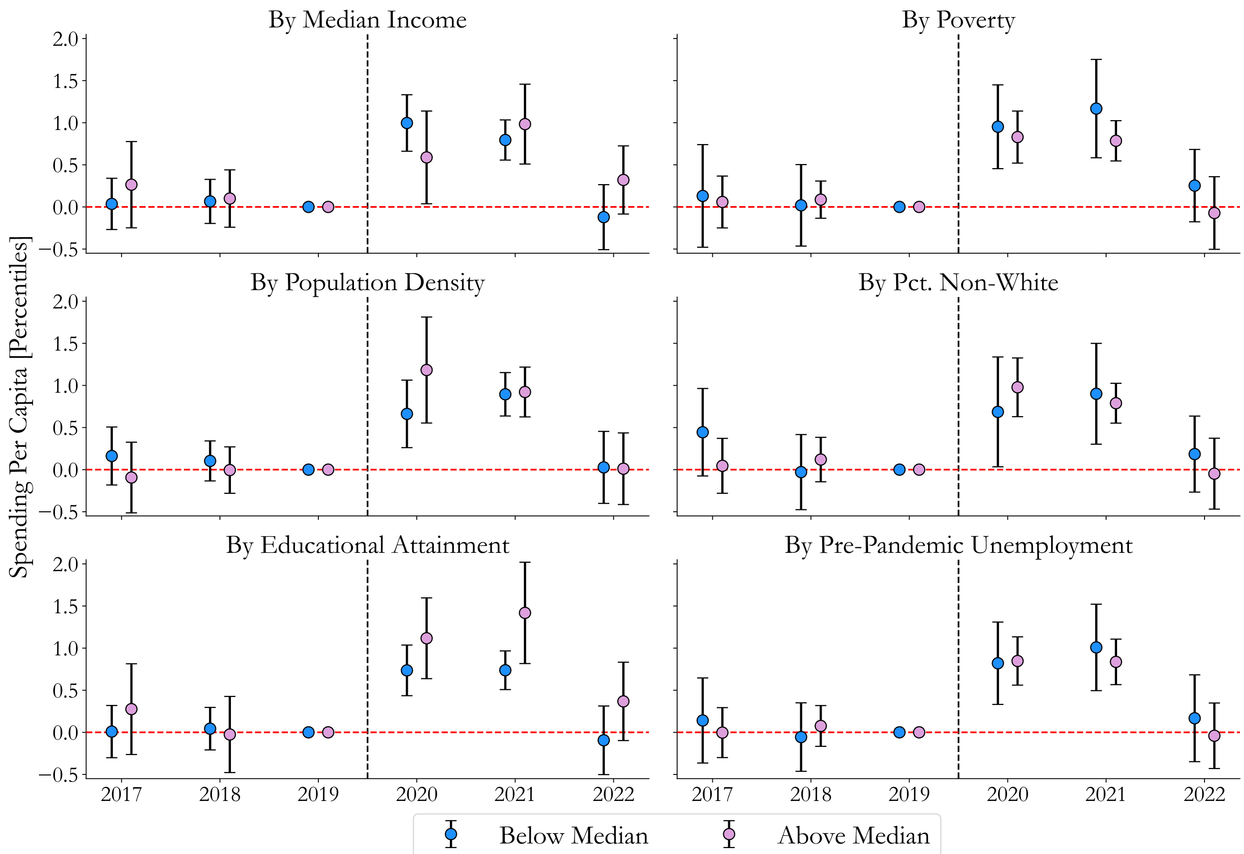


Figure IA.20. Effect on Regional Inflation Based on Regional CPI, Scatterplot
This figure shows the effect of suspicious lending on regional inflation from March 2020 to June 2022 using the all items regional consumer price indices (CPI) from the BLS. The dashed line is a line of best fit and the correlation is noted in the bottom left corner.

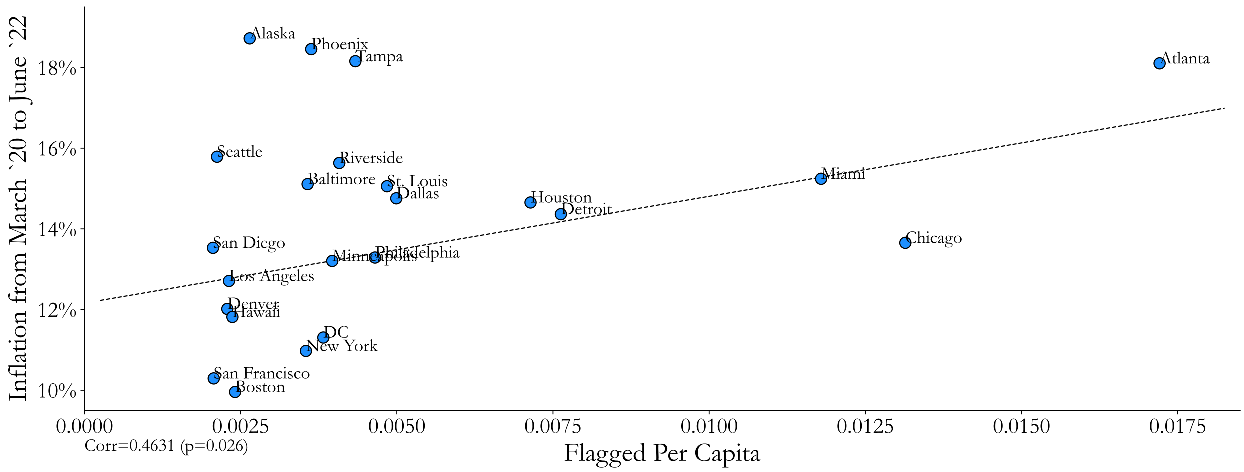
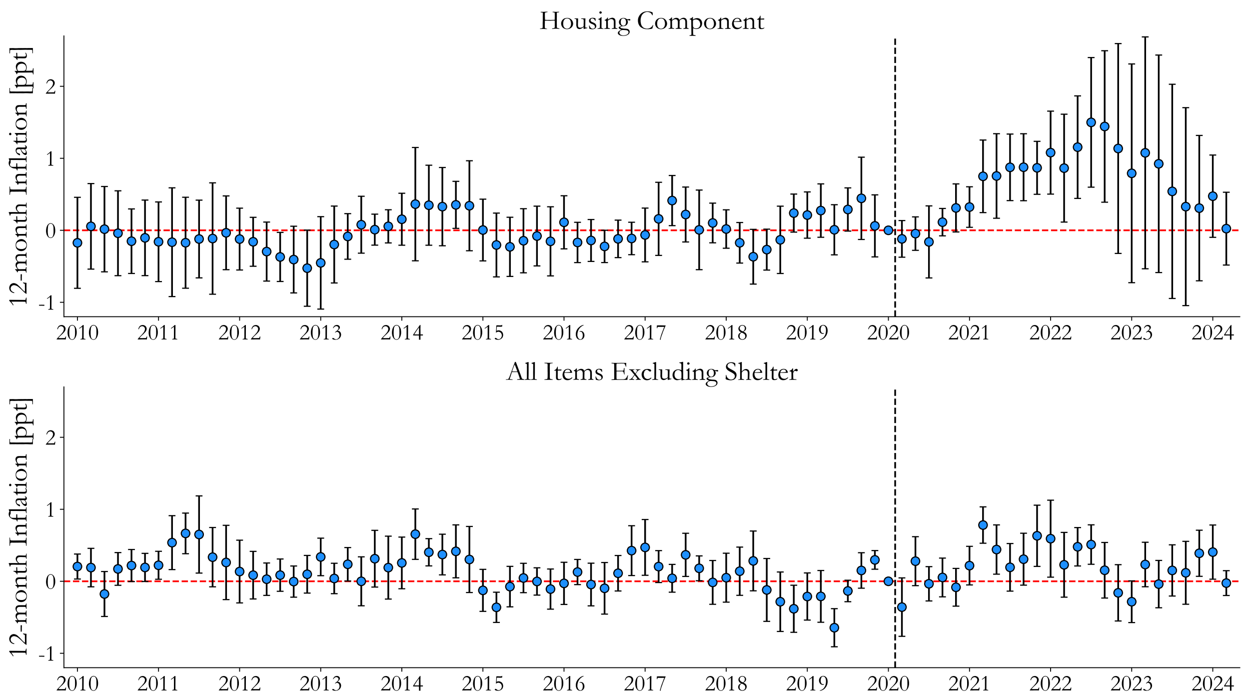


Figure IA.21. Effect on Regional Inflation Based on Regional CPI, Components
This figure replicates Figure 6 for the various component of regional CPI. Panel A examines the housing and non-housing components of CPI and Panel B examines the vehicle component of CPI. Data for 23 CBSAs is released bi-monthly. 12-month inflation is determined by dividing the given month's CPI by the CPI 12 months earlier. *Flagged Per Capita* is standardized, so the coefficients represent the inflation effect of a one standard deviation change in suspicious lending. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the CBSA's population as of 2019. The error bars represent 95% confidence intervals based on standard errors double clustered by CBSA and bi-month.

Panel A. Housing and Non-Housing Components



Panel B. Vehicle Component

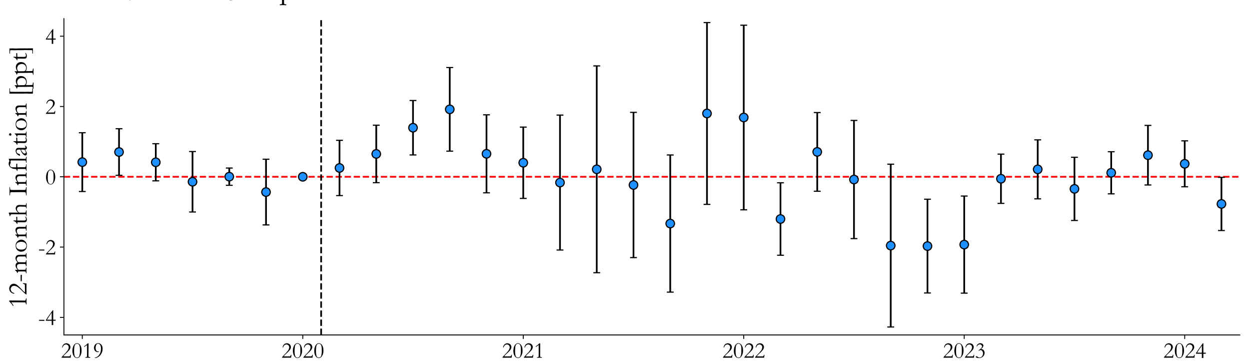


Figure IA.22. Effect of Suspicious Lending on Housing Price, IV
This figure replicates Panel B of Figure 2 using social connections outside each zip code's CBSA as an instrument. The error bars correspond to 95% confidence intervals based on standard errors that are clustered at the county level.

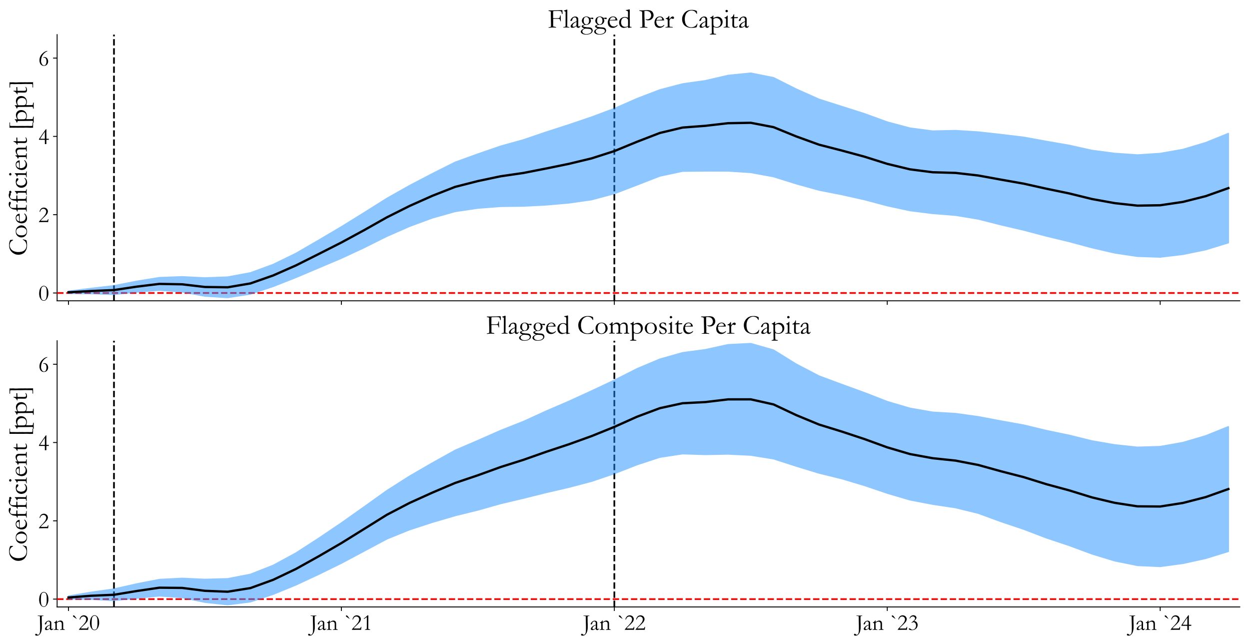


Table IA.1. Housing Purchases, FinTech Versus Traditional

This table examines whether heterogeneity in individuals purchased homes after receiving a flagged PPP loan by whether they received their PPPP loan from a FinTech or traditional lender. Data on home purchases for a sample of 250,000 loans from PropertyRadar is used. For each individual, we include monthly observations for the five years before they received their PPP loan to 18 months after. $1(HousingPurchase)$ takes a value of 12 (multiplied by 12 to annualize) if the individual bought a house during the given month. $1(Flagged)$ takes a value of 1 if the individual received a PPP loan that is flagged by at least one of the primary measures from Griffin, Kruger, and Mahajan (2023a). $1(Post)$ takes a value of 1 if the month is after the individual received their PPP loan. $1(FinTech)$ and $1(Traditional)$ take a value of 1 if the individual received their PPP loan from a FinTech and traditional lender, respectively. Fixed effects are indicated at the bottom of each column. Robust standard errors are double clustered by PPP loan and month.

Dep. Variable: 1(Housing Purchase) × 12

	(1)	(2)
1(FinTech) × 1(Post)	0.00825***	0.00815***
	(4.95)	(4.90)
1(Flagged) × 1(FinTech) × 1(Post)		0.00574***
		(3.53)
1(Flagged) × 1(Traditional) × 1(Post)		0.0160***
		(6.58)
1(Post)	0.00231	0.00135
	(1.02)	(0.59)
Loan FE	Yes	Yes
Month of Year FE	Yes	Yes
Observations	19,500,000	19,500,000
R^2	0.0279	0.0279
Mean of Dep. Variable	0.0501	0.0501

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.2. Moving

This table examines whether individuals are more likely to move in the two years after a receiving flagged PPP loan. Whether the individual moved is determined based on Melissa Data's Multisource Change of Address (mCOA) database. We collect this data on moving for the same sample of 150,000 individuals that Griffin, Kruger, and Mahajan (2023a) collected LexisNexis data for. Fixed effects are indicated at the bottom of each column. Robust standard errors are clustered at the zip code level.

Dep. Variable: 1(Moved in 2 Years After PPP Approval)	(1)	(2)	(3)	(4)
1(Flagged)	0.00950***	0.0126***	0.00636***	0.00499**
	(5.20)	(6.84)	(3.12)	(2.40)
ln(Loan Amount)	No	Yes	Yes	Yes
CBSA FE	No	No	Yes	Yes
Business Type FE	No	No	No	Yes
Week Approved FE	No	No	No	Yes
Observations	150,000	149,979	140,435	140,430
R^2	0.000171	0.000617	0.00932	0.0104
Mean of Dep. Variable	0.0487	0.0487	0.0507	0.0508

Table IA.3. Moving, Difference in Demographics
This table examines whether there is a difference in demographics between where individuals receiving suspicious PPP loans originally lived and where they moved to. Whether the individual moved is determined based on Melissa Data's Multisource Change of Address (mCOA) database. We collect this data of moving for the same sample of 150,000 individuals that Griffin, Kruger, and Mahajan (2023a) collected LexisNexis data for. Only data for individuals who moved during the two years after receiving their PPP loan are included. Demographics are at the zip code level. Fixed effects and controls are indicated at the bottom of each column. Robust standard errors are double clustered at the original and moved to zip code levels.

Dep. Var.: Difference in	(1) Median Income	(2) Poverty	(3) Unemployment	(4) Educ. Attainment	(5) House Prices
1(Flagged)	2508.9**	-0.00882**	-0.00378**	0.0327***	32320.0***
	(2.23)	(-2.15)	(-2.38)	(4.55)	(3.36)
ln(Loan Amount)	Yes	Yes	Yes	Yes	Yes
Observations	7,029	7,042	7,041	7,042	6,949
R^2	0.00105	0.00263	0.00323	0.00398	0.00128
Mean of Dep. Var.	1662.6	-0.00730	-0.00184	-0.00291	-18077.1

Table IA.4. Housing Price Growth, OLS
This table replicates Table 2 using OLS instead of WLS. Zip codes with populations of at least than 1000 individuals as of 2019 are included. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flaggedp Per Capita	0.0165***			
	(8.42)			
High Loan-to-Est. Per Capita		0.0148***		
		(4.35)		
High Similarity Per Capita			0.0146***	
			(6.73)	
Flagged Composite Per Capita				0.0172***
				(6.02)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,491	18,491	18,491	18,491
Num. Counties	2,204	2,204	2,204	2,204
R^2	0.800	0.797	0.799	0.799
Mean of Dep. Var.	0.263	0.263	0.263	0.263

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.5. Housing Price Growth, Previous House Price Growth

This table examines robustness of columns (1) and (5) Table 2 to alternative functional forms of the previous house price growth control. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Flagged Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0212***	0.0211***	0.0211***	0.0212***
Per Capita	(10.20)	(9.59)	(9.53)	(9.65)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Polynomial	1st Deg.	2nd Deg.	4th Deg.	6th Deg.
Loans Per Capita Percentile	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. County	2,215	2,215	2,215	2,215
R^2	0.825	0.825	0.825	0.826
Mean of Dep. Variable	0.259	0.259	0.259	0.259

Panel B. Flagged Composite Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Composite	0.0270***	0.0268***	0.0269***	0.0270***
Per Capita	(13.68)	(12.81)	(12.77)	(13.12)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Polynomial	1st Deg.	2nd Deg.	4th Deg.	6th Deg.
Loans Per Capita Percentile	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. County	2,215	2,215	2,215	2,215
R^2	0.825	0.825	0.825	0.825
Mean of Dep. Variable	0.259	0.259	0.259	0.259

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.6. Housing Price Growth, Loans Per Capita

This table examines robustness of columns (1) and (5) Table 2 to alternative functional forms of the loans per capita control. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Flagged Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flaggedp	0.0195***	0.0204***	0.0206***	0.0213***
Per Capita	(7.13)	(7.10)	(7.87)	(8.75)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Percentile	Yes	Yes	Yes	Yes
Loans Per Capita Polynomial	1st Deg.	2nd Deg.	4th Deg.	6th Deg.
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. County	2,215	2,215	2,215	2,215
R^2	0.825	0.826	0.826	0.826
Mean of Dep. Variable	0.259	0.259	0.259	0.259

Panel B. Flagged Composite Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Composite	0.0243***	0.0259***	0.0262***	0.0272***
Per Capita	(10.46)	(10.26)	(11.41)	(12.60)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Percentile	Yes	Yes	Yes	Yes
Loans Per Capita Polynomial	1st Deg.	2nd Deg.	4th Deg.	6th Deg.
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. County	2,215	2,215	2,215	2,215
R^2	0.825	0.825	0.826	0.826
Mean of Dep. Variable	0.259	0.259	0.259	0.259

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.7. Housing Price Growth, Controlling for Non-Flagged Loans Per Capita
This table replicates Table 2 while controlling for non-flagged loans per capita instead of loans per capita. Non-flagegd loans per capita is based on the baseline flagged per capita measure. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by CBSA.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0158***			
Per Capita	(6.79)			
High Loan-to-Est.		0.0173***		
Per Capita		(8.95)		
High Similarity			0.0184***	
Per Capita			(7.27)	
Flagged Composite				0.0202***
Per Capita				(8.52)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Non-Flagged Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215
R^2	0.827	0.825	0.828	0.827
Mean of Dep. Variable	0.259	0.259	0.259	0.259

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.8. Housing Price Growth, Effect of FinTech versus Traditional Flagged Loans This table replicates Table 2 using *Flagged Per Capita* split by whether the loan was originated by a FinTech or traditional lender. *FinTech Flagged Per Capita* and *Traditional Flagged Per Capita* are based on only flagged FinTech and traditional flagged loans, respectively. Column (1) shows a weighted least squares (WLS) regression with the weight being the zip code's population as of 2019. Column (2) shows a ordinary least squares (OLS) regression based on zip codes with populations of at least 1000 individuals as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1) WLS	(2) OLS
FinTech Flagged Per Capita	0.0194***	0.0179***
	(7.84)	(11.66)
Traditional Flagged Per Capita	-0.00207*	-0.00183***
	(-1.67)	(-2.74)
County FE	Yes	Yes
Past HP Growth Perc.	Yes	Yes
Loans Per Capita Perc.	Yes	Yes
Controls	Yes	Yes
Observations	18,761	18,491
Num. Counties	2,215	2,204
R^2	0.829	0.802
Mean of Dep. Variable	0.259	0.263

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.9. House Price, Propensity Score Weighted
This table uses propensity score weighting to control for previous house price growth. Propensity scores are estimated using WLS regressions of the measures of suspicious lending on 10th order polynomials of 2018-19 house price growth. Columns (1) and (2) replicate the results of Columns (1) and (5), respectively, of Table 2 with weighting based on inverse propensity scores. Columns (3) and (4) replicate Columns (1) and (2), respectively, with the dependent variable being house price growth in 2018-19. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

	(1)	(2)	(3)	(4)
Dep. Var.:	Housing Price Growth 2020-21		Housing Price Growth 2018-19	
Flagged	0.0206***		-0.000122	
Per Capita	(7.83)		(-0.06)	
Flagged Composite		0.0231***		0.00527
Per Capita		(5.36)		(1.49)
County FE	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	17,555	17,555	17,555	17,555
Num. Counties	2,188	2,188	2,188	2,188
R^2	0.907	0.841	0.838	0.783

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.10. Housing Price Growth, Alternative Measures

This table examines alternative measures of suspicious lending. *Pct. Flagged* is the percentage of PPP to the zip code that are flagged by at least one primary flag. *Dollars Flagged to Total Income* is the ratio of the dollar value of flagged PPP loans to the zip code to the total income of the zip code (per IRS SOI data). *Dollars Flagged to Total Loan Amount* is the ratio of the dollar value of flagged PPP loans to the zip code to the dollar value of all PPP loans in the zip code. *Total Loan Amount to Total Income* is the ratio of the dollar value of all PPP loans in the zip code to the total income of the zip code (per IRS SOI data). *Loans Per Capita* is the ratio of the number of PPP loans in the zip code to the 2019 population of the zip code. *Non-flagged Loans Per Capita* is the ratio of the number of PPP loans not flagged by a primary flag in the zip code to the 2019 population of the zip code. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)	(6)
Pct. Flagged	0.0216***					
	(11.53)					
Dollars Flagged to Total Income		0.00993***				
		(4.20)				
Dollars Flagged to Total Loan Amount			0.00829***			
			(4.90)			
Total Loan Amount to Total Income				-0.00501***		
				(-6.54)		
Loans Per Capita					-0.00365*	
					(-1.69)	
Non-flagged Loans Per Capita						-0.00648***
						(-6.02)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	No	No	No	No	No
Loan Amount to Income Perc.	No	Yes	Yes	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215	2,215	2,215
R^2	0.830	0.824	0.824	0.817	0.816	0.817
Mean of Dep. Variable	0.259	0.259	0.259	0.259	0.259	0.259

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.11. Housing Price Growth, Robustness of WLS to Different Controls and Fixed Effects

This table shows the robustness of columns (1) and (5) of Table 2 to varying the fixed effects and controls included. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Flagged Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0177***	0.0215***	0.0211***	0.0175***
Per Capita	(10.42)	(8.50)	(8.49)	(7.31)
County FE	Yes	Yes	Yes	Yes
Loans Per Capita Percentiles	No	Yes	Yes	Yes
Past HP Growth Percentiles	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Other Proposed House Price Drivers	No	No	No	Yes
Observations	12,305	12,305	12,305	12,305
Num. Counties	1,013	1,013	1,013	1,013
R^2	0.763	0.808	0.819	0.831
Mean Dep. Variable	0.257	0.257	0.257	0.257

Panel B. Flagged Composite Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Composite	0.0201***	0.0271***	0.0274***	0.0224***
Per Capita	(8.66)	(10.38)	(11.36)	(9.38)
County FE	Yes	Yes	Yes	Yes
Loans Per Capita Percentiles	No	Yes	Yes	Yes
Past HP Growth Percentiles	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Other Proposed House Price Drivers	No	No	No	Yes
Observations	12,305	12,305	12,305	12,305
Num. Counties	1,013	1,013	1,013	1,013
R^2	0.762	0.807	0.818	0.831
Mean Dep. Variable	0.257	0.257	0.257	0.257

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.12. Housing Price Growth, Within-CBSA

This table replicates Table 2 within-CBSA instead of within-county. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by CBSA.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0222*** (9.41)			
High Loan-to-Est. Per Capita		0.0154*** (3.51)		
High Similarity Per Capita			0.0156*** (4.09)	
Flagged Composite Per Capita				0.0191*** (4.17)
CBSA FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	17,279	17,279	17,279	17,279
Num. CBSA	894	894	894	894
R^2	0.739	0.731	0.732	0.732
Mean of Dep. Variable	0.257	0.257	0.257	0.257

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.13. Housing Price Growth, Controlling for Pre-COVID House Prices
This table replicates Table 2 while controlling for house prices as of January 2020. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0194*** (16.19)			
High Loan-to-Est. Per Capita		0.0206*** (9.45)		
High Similarity Per Capita			0.0201*** (14.65)	
Flagged Composite Per Capita				0.0249*** (14.67)
County FE	Yes	Yes	Yes	Yes
Pre-COVID House Price Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	19,013	19,013	19,013	19,013
Num. Counties	2,250	2,250	2,250	2,250
R^2	0.842	0.839	0.842	0.842
Mean of Dep. Var.	0.259	0.259	0.259	0.259

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.14. Housing Price Growth, Controlling for COVID Mortgage Forbearance
This table replicates Table 2 while controlling for COVID mortgage forbearance. Forbearance data is as of the June 2020 reporting period (the peak of COVID mortgage forbearance) and is from the Mortgage Analytics and Performance Dashboard created by the Federal Reserve Bank of Atlanta based on Black Knight's McDash Flash daily mortgage performance data. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0212*** (9.33)			
High Loan-to-Est. Per Capita		0.0226*** (10.92)		
High Similarity Per Capita			0.0228*** (9.80)	
Flagged Composite Per Capita				0.0271*** (11.56)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Forbearance Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	14,251	14,251	14,251	14,251
Num. Counties	1,759	1,759	1,759	1,759
R^2	0.840	0.837	0.840	0.839
Mean of Dep. Var.	0.261	0.261	0.261	0.261

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.15. Housing Price Growth, Robustness of WLS to Clustering at State Level This table replicates Table 2 with clustering at the state level. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by state.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0211*** (7.05)			
High Loan-to-Est. Per Capita		0.0222*** (8.44)		
High Similarity Per Capita			0.0224*** (7.47)	
Flagged Composite Per Capita				0.0268*** (8.84)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
R^2	0.828	0.824	0.828	0.827
Mean of Dep. Var.	0.259	0.259	0.259	0.259

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.16. Housing Price Growth, Measures Based on Different Loan Samples
This table replicates Table 2 using measures based on subsets of the PPP loans. Panel A uses only loans to residential addresses and Panel B uses loans to independent contractors, self-employed, and sole-proprietors. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Loans to Residential Addresses
Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0224***			
Per Capita	(10.59)			
High Loan-to-Est.		0.0249***		
Per Capita		(13.83)		
High Similarity			0.0251***	
Per Capita			(13.75)	
Flagged Composite				0.0269***
Per Capita				(15.42)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	17,381	17,381	17,381	17,381
Num. Counties	2,146	2,146	2,146	2,146
R^2	0.836	0.835	0.835	0.837
Mean of Dep. Variable	0.259	0.259	0.259	0.259

Panel B. Loans to Independent Contractors, Self-Employed, and Sole Proprietors
Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0194***			
Per Capita	(7.41)			
High Loan-to-Est.		0.0194***		
Per Capita		(10.46)		
High Similarity			0.0217***	
Per Capita			(9.48)	
Flagged Composite				0.0229***
Per Capita				(10.59)
Same Fixed Effects and Controls as Panel A.				
Observations	16,150	16,150	16,150	16,150
Num. Counties	2,077	2,077	2,077	2,077
R^2	0.830	0.828	0.830	0.830
Mean of Dep. Variable	0.258	0.258	0.258	0.258

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.17. Housing Price Growth, County Level

This table replicates Table 2 using data at the county-level. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the county's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by CBSA.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.00938*** (2.95)			
High Loan-to-Est. Per Capita		0.0109*** (4.18)		
High Similarity Per Capita			0.0127*** (3.93)	
Flagged Composite Per Capita				0.0148*** (4.24)
CBSA FE	Yes	Yes	Yes	Yes
Past HP Growth Decile	Yes	Yes	Yes	Yes
Loans Per Capita Decile	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,204	1,204	1,204	1,204
Num. CBSAs	316	316	316	316
R^2	0.848	0.850	0.850	0.850
Mean of Dep. Var.	0.248	0.248	0.248	0.248

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.18. Housing Price Growth, Pre-2023 Legacy ZHVI

This table replicates Tables 2, 3, and 4 using the pre-2023 legacy version of the Zillow Home Value Index (i.e., the version of the ZHVI that Zillow generated before January 2023). To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Replicating Table 2

Dep. Variable: House Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Flagged	0.0200***	0.0218***			
Per Capita	(11.02)	(10.23)			
High Loan-to-Est			0.0222***		
Per Capita			(11.90)		
High Similarity				0.0218***	
Per Capita				(14.47)	
Flagged Composite					0.0268***
Per Capita					(14.53)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	No	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	No	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
Observations	18,733	18,733	18,733	18,733	18,733
Num. County	2,198	2,198	2,198	2,198	2,198
R^2	0.802	0.855	0.852	0.854	0.854
Mean of Dep. Var.	0.288	0.288	0.288	0.288	0.288

Panel B. Replicating Table 3

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Instrument:	Outside CBSA	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi	Concentric Rings
Flagged	0.0345***	0.0354***	0.0341***	0.0338***	0.0356***
Per Capita	(6.99)	(7.28)	(6.83)	(6.44)	(7.28)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth	Yes	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,792	18,733	18,733	18,733	18,697
Num. Counties	1,774	2,198	2,198	2,198	2,196
R^2	0.291	0.288	0.288	0.288	0.288
Mean of Dep. Var.	0.317	0.309	0.312	0.313	0.308
First Stage F-stat	36.73	40.17	36.67	30.29	14.32
Hansen's J-stat					1.640
(p-value)					(0.440)

Panel C. Replicating Table 4

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flagge	0.0222***								0.0182***
Per Capita	(9.38)								(6.87)
Log of		0.0123***							0.00397
Dist. to CBD		(3.68)							(1.42)
Log of			-0.0160***						-0.00134
Pop. Density			(-3.93)						(-0.65)
Land				0.0243***					0.0203***
Unavailability				(9.82)					(10.47)
Remote Work					0.00566***				0.00432***
2015-19					(3.39)				(2.75)
Teleworkable						-0.0195***			-0.0145***
						(-7.28)			(-6.63)
Net Migration							0.0103***		0.00475***
2020-21							(6.04)		(4.62)
HP Growth								0.0191***	0.0137***
2018-19								(6.66)	(5.45)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Loans Per Capita	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	12,302	12,302	12,302	12,302	12,302	12,302	12,302	12,302	12,302
Num. Counties	1,017	1,017	1,017	1,017	1,017	1,017	1,017	1,017	1,017
R^2	0.843	0.838	0.837	0.843	0.834	0.837	0.838	0.828	0.852
Mean of Dep, Var.	0.297	0.297	0.297	0.297	0.297	0.297	0.297	0.297	0.297

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.19. Housing Price Growth, Realtor.com Data

This table replicates Table 2 using data from Realtor.com. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Median List Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged	0.0474***			
Per Capita	(10.75)			
High Loan-to-Est		0.0452***		
Per Capita		(6.00)		
High Similarity			0.0504***	
Per Capita			(10.25)	
Flagged Composite				0.0582***
Per Capita				(8.94)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,187	18,187	18,187	18,187
Num. County	2,235	2,235	2,235	2,235
R^2	0.281	0.280	0.281	0.281
Mean of Dep. Var.	0.283	0.283	0.283	0.283

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.20. Housing Price Growth, Economic Impact Payments

This table examines the robustness of columns (1) and (5) of Table 2 to controlling for economic impact payments (also known as stimulus checks). Data on economic impact payments is from 2020 IRS ZIP Code SOI data. *Pct. Receiving EIP* is the higher of the number of tax returns with first round EIP or second round EIP divided by number of returns. *Dollars of EIP Per Capita* is the sum of first round EIP and second round EIP divided by the total number of individuals. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Panel A. Flagged Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capit	0.0211***	0.0199***	0.0203***	0.0190***
	(9.84)	(10.60)	(9.62)	(10.43)
Pct. Receiving EIP	-0.00260			
	(-1.13)			
Dollars of EIP Per Capita			-0.00917***	
			(-3.64)	
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Pct. Receiving EIP Perc.	No	Yes	No	No
Dollars of EIP Per Capita Perc.	No	No	No	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215
R^2	0.828	0.831	0.829	0.833
Mean of Dep. Var.	0.259	0.259	0.259	0.259

Panel B. Flagged Composite Per Capita

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Composite Per Capit	0.0270***	0.0255***	0.0262***	0.0245***
	(13.38)	(14.04)	(12.77)	(13.34)
Pct. Receiving EIP	-0.00395*			
	(-1.71)			
Dollars of EIP Per Capita			-0.0106***	
			(-4.24)	
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Pct. Receiving EIP Perc.	No	Yes	No	No
Dollars of EIP Per Capita Perc.	No	No	No	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215
R^2	0.828	0.831	0.829	0.833
Mean of Dep. Var.	0.259	0.259	0.259	0.259

Table IA.21. Rent Growth

This table replicates Table 2 using rent growth instead of housing price growth. Rent growth is based on the Zillow Observed Rent Index (ZORI). To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Rent Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.00640*** (2.83)			
High Loan-to-Est. Per Capita		0.00714* (1.88)		
High Similarity Per Capita			0.00956*** (3.61)	
Flagged Composite Per Capita				0.00876** (2.52)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	1,397	1,397	1,397	1,397
Num. County	178	178	178	178
R^2	0.905	0.905	0.906	0.905
Mean of Dep. Var.	0.204	0.204	0.204	0.204

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.22. House Price Growth, Heterogeneity by Political Lean and COVID
This table examines heterogeneity in the results shown in column 1 of Table 2. *Red State* (*Red County*) is based on the 2020 presidential election and takes a value of 1 if more voters voted for Trump than Biden in the state (county). All other interaction variables are based on splits at the median value of the variables. SVI is the Social Vulnerability Index from the CDC. First Dose and Complete Dose are from the CDC as of 12/31/21. COVID Cases are cumulative cases as of 12/31/21 per the Johns Hopkins Coronavirus Resource Center. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)	(6)
Flagged Per Capita	0.0215***	0.0210***	0.0230***	0.0214***	0.0274***	0.0208***
	(8.41)	(9.80)	(7.07)	(4.84)	(5.44)	(10.21)
× 1(Red State)	-0.00170					
	(-0.45)					
× 1(Red County)		0.00339				
		(0.74)				
× 1(High SVI)			-0.00199			
			(-0.57)			
× 1(High First Dose)				-0.000354		
				(-0.08)		
× 1(High Complete Dose)					-0.00729	
					(-1.47)	
× 1(High COVID Cases)						0.00209
						(0.49)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth Percentile	Yes	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Percentile	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,761	18,746	18,761	18,761	18,761	18,761
Num. County	2,215	2,211	2,215	2,215	2,215	2,215
R^2	0.828	0.828	0.828	0.828	0.828	0.828
Mean of Dep. Var.	0.259	0.259	0.259	0.259	0.259	0.259

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.23. Housing Price Growth, Heterogeneity by Race and Ethnicity and Robustness to Excluding Racially or Ethically Homogeneous Zip Codes

This table examines heterogeneity in columns (1) of Table 2 by race and ethnicity (Panel A) and its robustness to excluding zip codes that are racially or ethnically homogeneous. The race or ethnicity that the splits are performed by are noted at the top of each column. AIANNH is American Indian, Alaska Native, or Native Hawaiian. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. The F-stat and p-value at the bottom of the table test for equality in the effect above and below the median. Robust standard errors are clustered by county.

Panel A. Heterogeneity by Race and Ethnicity

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

Split by:	(1) White	(2) Black	(3) Asian	(4) AIANNH	(5) Other	(6) Hispanic/ Latino
Flagged Per Capita						
× 1(Above Median)	0.0304***	0.0213***	0.0162***	0.0209***	0.0216***	0.0194***
	(6.95)	(10.55)	(5.77)	(5.62)	(14.47)	(12.52)
× 1(Below Median)	0.0210***	0.0358***	0.0223***	0.0212***	0.0211***	0.0238***
	(10.54)	(6.49)	(8.92)	(11.81)	(5.77)	(7.11)
1(Above Median)	-0.00598**	-0.00265	0.000987	0.00104	0.00516***	0.00846***
	(-2.35)	(-0.94)	(0.43)	(0.90)	(3.43)	(4.41)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18,761	18,761	18,761	18,761	18,761	18,761
Num. Counties	2,215	2,215	2,215	2,215	2,215	2,215
R^2	0.828	0.828	0.828	0.828	0.828	0.828
Mean Dep. Variable	0.259	0.259	0.259	0.259	0.259	0.259
F-stat for Equality	5.169	8.640	10.078	0.0186	0.0328	1.829
(p-value)	0.023	0.003	0.002	0.892	0.856	0.176

Panel B. Robustness to Excluding Racially or Ethically Homogeneous Zip Codes

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

Sample:	(1) $\leq$ 90% White	(2) $\leq$ 75% White	(3) $\leq$ 90% Black	(4) $\leq$ 75% Black	(5) $\leq$ 90% Hispanic/ Latino	(6) $\leq$75% Hispanic/ Latino
Flagged	0.0211***	0.0213***	0.0207***	0.0201***	0.0210***	0.0209***
Per Capita	(10.34)	(8.69)	(11.85)	(9.59)	(9.88)	(10.20)
Same Fixed Effects and Controls as Panel A.						
Observations	8,804	4,336	18,667	18,474	18,662	18,479
Num. Counties	1,080	599	2,215	2,208	2,213	2,210
R^2	0.817	0.803	0.834	0.842	0.829	0.829
Mean Dep. Variable	0.256	0.250	0.258	0.257	0.259	0.259

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.24. Migration by Social Connectedness

This table examines whether individuals are more likely to move between counties that have higher social connectedness. Migration data is from the 2020 and 2021 IRS County-to-County Migration Data Files, which are based on changes in address between tax returns filled in the 2019 (2020) calendar year and 2020 (2021) calendar year. Only pairs of counties with at least 10 returns moving between them during each year are included in the IRS data. In Panel A, the dependent variable is the number of individuals moving between each pair of counties. In Panel B, the dependent variable is the ratio of the number of individuals moving between each pair of counties and the population of the origin county in 2019. Column (1) shows the results for all pairs of counties. Columns (2), (3), and (4) are based on counties that are at least 100, 250, and 500 miles apart, respectively. *Social Connectedness* is standardized to have a mean of 0 and a standard deviation of 1 based on data for counties that meet the distance threshold for the given column. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the origin county's population as of 2019. Robust standard errors are double clustered by origin and destination counties.

Panel A. Individuals Moving

	(1)	(2)	(3)	(4)
Counties:	All	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi
Social Connectedness	101.1***	121.2	36.35	21.79
	(6.77)	(1.60)	(1.20)	(1.00)
Constant	724.6***	433.6***	341.4***	318.4***
	(9.63)	(5.12)	(8.11)	(9.94)
Observations	46,007	26,703	19,736	15,409
R^2	0.000665	0.000720	0.00157	0.00216
Mean of Dep. Var.	731.7	437.4	344.3	322.3

Panel B. Individuals Moving Divided by Population of Origin County

	(1)	(2)	(3)	(4)
Counties:	All	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi
Social Connectedness	0.000913***	0.000241***	0.0000778	0.0000346
	(27.05)	(4.23)	(1.47)	(1.16)
Constant	0.000534***	0.000210***	0.000167***	0.000152***
	(8.50)	(8.55)	(6.61)	(5.89)
Observations	46,007	26,703	19,736	15,409
R^2	0.111	0.0513	0.0323	0.0206
Mean of Dep. Var.	0.000599	0.000217	0.000173	0.000158

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.25. Housing Price Growth, IV Controlling for Social Proximity to House Price Growth

This table shows the robustness of Table 3 to controlling for social proximity to house price growth during 2020-21. Column (5) includes three instruments based on social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles) at the same time. The J-stat and p-value for an overidentification test are provided at the bottom of column (5). All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions in both stages with the weight being population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

Instrument:	(1) Outside CBSA	(2) $\geq$ 100 Mi	(3) $\geq$ 250 Mi	(4) $\geq$ 500 Mi	(5) Non-overlapping
Flagged Per Capita	0.0339*** (6.38)	0.0353*** (6.43)	0.0350*** (6.38)	0.0367*** (5.91)	0.0400*** (6.69)
SP to HP Growth 2020-21 Outside CBSA	0.0393*** (8.28)				
SP to HP Growth 2020-21 $\geq$ 100 Mi		0.0295*** (6.72)			
SP to HP Growth 2020-2 $\geq$ 250 Mi			0.0242*** (7.14)		
SP to HP Growth 2020-21 $\geq$ 500 Mi				0.0221*** (5.59)	0.0199*** (5.19)
SP to HP Growth 2020-21 $[100, 250)$ Mi					0.0166*** (3.02)
SP to HP Growth 2020-21 $[250, 500)$ Mi					0.0179** (2.14)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,722
Num. Counties	1,789	2,215	2,215	2,215	2,212
R^2	0.283	0.273	0.273	0.259	0.256
Mean of Dep. Variable	0.257	0.259	0.259	0.259	0.259
First Stage F-stat	34.22	38.03	34.33	28.79	18.17
Hansen's J-stat					2.386
(p-value)					(0.303)

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.26. Housing Price Growth, Additional Measures of Migration
This table examines the effects of additional measures of migration during 2020-21. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions in both stages with the weight being population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Flagged Per Capita	0.0200***	0.0198***	0.0194***	0.0189***
	(9.60)	(9.21)	(9.01)	(8.64)
Net Migration 2020-21	0.00845***			
	(7.32)			
Inflow 2020-21		0.0167***		
		(6.67)		
Outflow 2020-21		-0.0186***		
		(-7.51)		
Net Migration 2020			0.0133***	
			(5.33)	
Net Migration 2021			-0.00427**	
			(-2.37)	
Inflow 2020				0.0335***
				(7.45)
Inflow 2021				-0.0174***
				(-4.81)
Outflow 2020				-0.0205***
				(-3.92)
Outflow 2021				0.00251
				(0.54)
County FE	Yes	Yes	Yes	Yes
Past HP Growth	Yes	Yes	Yes	Yes
Loans Per Capita	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observation	18,653	18,653	18,653	18,653
Num. Counties	2,209	2,209	2,209	2,209
R^2	0.831	0.831	0.832	0.833
Mean of Dep, Var.	0.259	0.259	0.259	0.259

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.27. Housing Price Growth, IV with Non-Overlapping Rings

This table replicates Table 3 using instruments based on zip codes in three mutually exclusive distance ranges from each zip code. Column (1) uses zip codes that are between 100 and 250 miles away, column (2) between 250 and 500 miles away, and column (3) more than 500 miles away. Column (4) includes all three instruments at the same time. The J-stat and p-value for an overidentification test are provided at the bottom of column (4). To have a nationally representative estimate, we use weighted least squares (WLS) regressions for both stages with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)
Instrument:	$[100, 250)$ Miles	$[250, 500)$ Mi	≥ 500 Mi	All Three
Flagged	0.0448***	0.0336***	0.0339***	0.0367***
Per Capita	(5.51)	(5.54)	(5.43)	(6.41)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	18,725	18,725	18,725	18,725
Num. Counties	2,213	2,213	2,213	2,213
R^2	0.212	0.251	0.251	0.243
Mean of Dep. Var.	0.259	0.259	0.259	0.259
First Stage F-stat	29.69	21.46	28.84	14.38
Hansen's J-stat				2.465
(p-value)				(0.292)

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.28. Housing Price Growth, IV With Geopolitical Threshold Instruments
This table replicates Table 3 using instruments based on zip codes based on different geopolitical thresholds. Column (1) uses zip codes that are outside the given zip code's county, column (2) outside the given zip code's CBSA, and column (3) outside the given zip code's state. Column (4) includes all three instruments at the same time. The J-stat and p-value for an overidentification test are provided at the bottom of column (4). To have a nationally representative estimate, we use weighted least squares (WLS) regressions for both stages with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

Instrument:	(1) Outside County	(2) Outside CBSA	(3) Outside State	(4) All Three
Flagged Per Capita	0.0296***	0.0347***	0.0316***	0.0299***
	(9.08)	(6.35)	(6.92)	(8.74)
County FE	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	16,869	16,869	16,869	16,869
Num. Counties	1,789	1,789	1,789	1,789
R^2	0.267	0.256	0.263	0.266
Mean of Dep. Var.	0.257	0.257	0.257	0.257
First Stage F-stat	36.39	34.33	27.35	20.91
Hansen's J-stat				7.895
(p-value)				(0.019)

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.29. Housing Price Growth, IV Based on Percentage of Loans Flagged
This table replicates Table 3 using percentage of loans that are flagged instead of per capita rates. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions in both stages with the weight being population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Instrument:	Outside CBSA	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi	Non-overlapping
Pct. Flagged	0.0304***	0.0258***	0.0249***	0.0260***	0.0278***
	(12.01)	(8.88)	(8.49)	(6.95)	(9.49)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,725
Num. Counties	1,789	2,215	2,215	2,215	2,213
R^2	0.274	0.274	0.275	0.274	0.272
Mean of Dep. Var.	0.257	0.259	0.259	0.259	0.259
First Stage F-stat	183.9	307.7	272.0	193.2	105.6
Hansen's J-stat					2.182
(p-value)					(0.336)

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.30. Housing Price Growth, First Stage of IV

This table reports the first stage estimates that correspond to Tables 3, IA.27, and IA.28. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Flagged Per Capita

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Proximity Outside CBSA	0.791***							
	(5.86)							
Social Proximity ≥ 100 Mi		0.634***						
		(6.17)						
Social Proximity ≥ 250 Mi			0.591***					
			(5.85)					
Social Proximity ≥ 500 Mi				0.579***				
				(5.34)				
Social Proximity $[100, 250)$ Mi					0.585***			
					(5.42)			
Social Proximity $[250, 500)$ Mi						0.583***		
						(4.63)		
Social Proximity Outside County							0.914***	
							(6.03)	
Social Proximity Outside State								0.706***
								(5.23)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Per Capita Perc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,761	18,725	16,869	16,869
Num. Counties	1,789	2,215	2,215	2,215	2,215	2,213	1,789	1,789
R^2	0.827	0.816	0.812	0.808	0.791	0.790	0.847	0.830

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.31. Housing Price Growth, Reduced Form

This table reports the reduced form estimates that correspond to Tables 3, IA.27, and IA.28. To have a nationally representative estimate, we use weighted least squares (WLS) regressions with the weight being the zip code's population as of 2019. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Social Proximity Outside CBSA	0.0275***							
	(8.69)							
Social Proximity $\geq$ 100 Mi		0.0224***						
		(8.18)						
Social Proximity $\geq$ 250 Mi			0.0199***					
			(7.72)					
Social Proximity $\geq$ 500 Mi				0.0197***				
				(6.48)				
Social Proximity $[100, 250)$ M					0.0262***			
					(6.68)			
Social Proximity $[250, 500)$ Mi						0.0196***		
						(5.16)		
Social Proximity Outside County							0.0270***	
							(6.50)	
Social Proximity Outside State								0.0223***
								(6.72)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Per Capita Perc.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,761	18,725	16,869	16,869
Num. Counties	1,789	2,215	2,215	2,215	2,215	2,213	1,789	1,789
R^2	0.824	0.825	0.824	0.824	0.825	0.822	0.823	0.823
Mean of Dep. Var.	0.257	0.259	0.259	0.259	0.259	0.259	0.257	0.257

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010

Table IA.32. Housing Price Growth, IV Without Weighting by Population
This table replicates Table 3 without weighting by population in each zip code. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Instrument:	Outside CBSA	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi	Non-overlapping
Flagged Per Capita	0.0368*** (6.33)	0.0349*** (6.12)	0.0341*** (6.09)	0.0347*** (5.29)	0.0360*** (6.02)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,725
Num. Counties	1,789	2,215	2,215	2,215	2,213
R^2	0.168	0.154	0.157	0.154	0.151
Mean of Dep. Var.	0.259	0.263	0.263	0.263	0.263
First Stage F-stat	30.49	33.97	29.86	23.94	21.83
Hansen's J-stat (p-value)					1.206 (0.547)

t-statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Table IA.33. Housing Price Growth, IV: Flagged Composite Per Capita

This table replicates Table 3 using the *Flagged Composite Per Capita* measure. Column (1) is based on social connectedness between each zip codes and zip codes that are outside the given zip code's CBSA. Columns (2), (3), and (4) are based on social connectedness between each zip code and zip codes that are at least 100, 250, and 500 miles away, respectively. Column (5) includes three instruments based on social connections to zip codes in non-overlapping rings (between 100 and 250 miles, between 250 and 500 miles, and over 500 miles) at the same time. The J-stat and p-value for an overidentification test are provided at the bottom of column (5). *Past HP Growth Perc.* and *Loans Per Capita Perc.* control for house price growth in 2018-19 and PPP lending intensity, respectively, using percentile fixed effects. The controls included are log population density, vacancy rate, log housing units, log average household income, and the share of Facebook friends within 50 and 150 miles. All independent variables are standardized to have a mean of 0 and a standard deviation of 1. To have a nationally representative estimate, we use weighted least squares (WLS) regressions in both stages with the weight being population of the zip code in 2019. Fixed effects are as indicated at the bottom of each column. Robust standard errors are clustered by county.

Dep. Variable: Housing Price Growth from January 1, 2020 to December 31, 2021

	(1)	(2)	(3)	(4)	(5)
Instrument:	Outside CBSA	$\geq$ 100 Mi	$\geq$ 250 Mi	$\geq$ 500 Mi	Concentric Rings
Flagged Composite Per Capita	0.0420*** (7.06)	0.0432*** (6.76)	0.0396*** (6.50)	0.0405*** (5.58)	0.0437*** (6.72)
County FE	Yes	Yes	Yes	Yes	Yes
Past HP Growth Perc.	Yes	Yes	Yes	Yes	Yes
Loans Per Capita Perc.	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Observations	16,869	18,761	18,761	18,761	18,725
Num. Counties	1,789	2,215	2,215	2,215	2,213
R^2	0.261	0.250	0.256	0.255	0.249
Mean of Dep. Var.	0.257	0.259	0.259	0.259	0.259
First Stage F-stat	50.59	52.07	39.88	29.78	81.95
Hansen's J-stat					6.057
(p-value)					(0.048)

t-statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.010