BANK OF ENGLAND

Appendix to Staff Working Paper No. 921

Income inequality, mortgage debt and house prices

Sevim Kösem

May 2021

BANK OF ENGLAND

Appendix to Staff Working Paper No. 921

Income inequality, mortgage debt and house prices

Sevim Kösem[1]

Abstract

This paper studies housing and credit market implications of increasing income inequality and discusses how a low interest rate environment can alter its consequences. I develop an analytical general equilibrium model with a novel borrower risk composition channel of income inequality. Following a rise in income inequality house prices and mortgage debt decline, and aggregate default risk increases. I then show that low real rates mitigate the depressing effect of inequality on house prices at the cost of amplifying the aggregate default risk. Using a panel of US states and instrumental variables approach, I verify the model's predictions.

Key words: Income inequality, mortgage lending, mortgage default, house prices, real interest rates, risk taking, shift-share instruments.

JEL classification: D31, E44, E58, G21, R21.

(1) Bank of England, Centre for Macroeconomics and Systemic Risk Centre. Email: sevim.kosem@bankofengland.co.uk.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

In this online appendix I provide additional cross-sectional facts to those in the main text. First I present the findings using the county level data, then those with the state level data and finally report first-step estimation results for the 2SLS regressions in the main text. County level data allows me to analyse the relevance of different quantiles of the income distribution, which is not available in the state level data, for inequality, house prices and mortgage developments.

A County level data

County level data primarily comes from the U.S. Census and the American Community Survey (ACS) 5-year averages.[1] The Gini coefficient, population, mean household income, number of households are obtained from these source. I use the 1990, 2000, 2011 and 2016 releases. County level data gives rise to a larger number of cross-sections than state level data.

House price data is from the Federal Housing Finance Agency. This is a repeat-sales index, that measures average price changes in repeat sales or refinancing on the same properties since 1975. I deflate nominal quantities using the CPI-U-RS price index provided by the Bureau of Labor Statistics.[2]

County level debt data is from the Federal Reserve Bank of New York Consumer Credit Panel (FRBNY CCP). This is publicly available for the period 1999 to 2011.[3] I use the per capita balance of mortgage debt excluding home equity lines of credit as my measure of mortgage debt. My measure of delinquency is the percent of the mortgage debt balance that has been delinquent for more than ninety days. The share of subprime borrowers is also from this source. The data used for Figure 2 includes 2093 US counties that have data for both house prices and mortgage variables.

Income inequality and house prices

In this section I provide nonparametric evidence regarding the correlation between real house prices and income inequality growth. This complements the evidence in

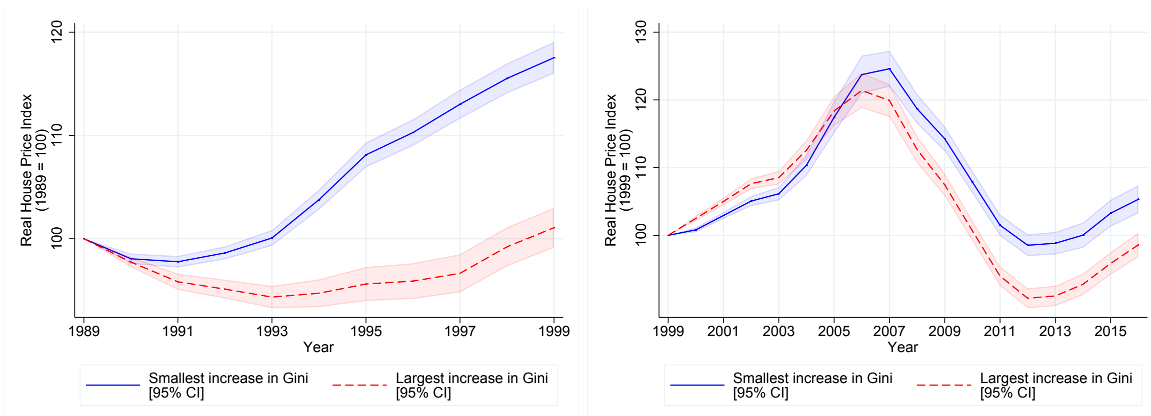
[1]For the ACS, sampling error from the survey decreases with the size of the county and the number of yearly surveys used, and some counties are not reported in 1 year surveys. In the decennial Census, income data is for the previous calender year. That is, the 1990 Census reports income data for the year 1989. In the ACS, income is for the year prior to the interview date, and the survey is conducted monthly. To avoid sampling error, income inequality data for 2016 thus includes incomes reported as early as year 2012 for some respondents. However, income levels are adjusted to 2016 current dollars. The Census Bureau advises the use of ACS 5 years estimates for areas with a population below 65000.

[2]This series is considered to be the most detailed and systematic estimate available of a consistent CPI series. This matters as there was an important methodological in the construction of CPI series before 2000.

[3]The data has not been publicly available for the period after 2011.

Figure 2. Each panel of Figure A.1 displays the real house price trends for the US counties that had the highest and the lowest increase in income inequality over a given time period.[4] The red dashed line shows house price growth for counties in the top quintile for income inequality growth. The blue solid line shows income growth for counties in the bottom quintile. In both subperiods, being in the bottom quintile corresponds to experiencing a decline in income inequality. Both subplots have the same message: house price growth is higher for counties in the bottom quintile for income inequality growth. The difference is as high as 15.3% in 1999 and 8.8% in 2012.

Figure A.1: House price growth and income inequality change for US counties



Source: US Census Bureau, Federal Housing and Finance Agency, own calculations.

Note: The red dashed line is real house prices for US counties in the top quintile for income inequality growth. The blue line is real house prices for counties in the bottom quintile. Growth in income inequality is measured by the change in Gini coefficient between the first and the last year of the subperiod. The counties in each group remain the same over time within a subperiod. To ensure comparability, only counties where data for the Gini coefficient and house prices is available for both subperiods are used. This corresponds to 1390 counties, which comprise about 90% of the total population in 1999.

High income inequality growth is associated with low house price growth in comparison not only to other counties but also to the initial time period. That is, for both subsamples, high growth in income inequality is associated with a real terms decline in house prices.

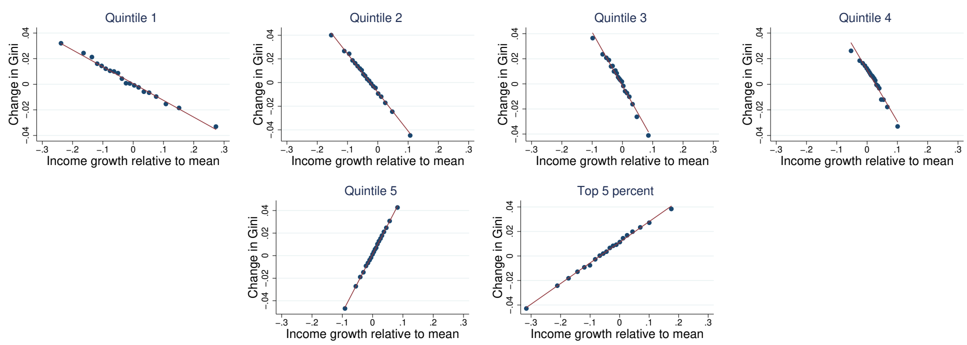
The second panel of Figure A.1 shows that, between 1999-2005, counties in the highest inequality growth quintile experienced slightly higher house price growth than other counties. House price growth in these counties is around 2% higher that of the lowest inequality growth quintile. Limiting my analysis to this specific time span would lead to the opposite conclusion to the rest of this paper. In fact, counties where

[4]House price data is available for 1390 counties from 1989 onwards. These counties comprise 89.5% of the total population in 1999.

income inequality growth was lowest experienced a larger boom and a smaller bust than counties with high income inequality growth that experienced high house price growth at the beginning of the cycle. Over the entire span of the data, the boom episode preceding the Great Recession is the exception, rather than the rule, in terms of the relationship between house prices and inequality.

Digging deeper: the relevance of the different quintiles of the income distribution for inequality, housing and credit market developments

Figure A.2: Income inequality change and quintile relative income growth between the years 1999 and 2011



Source: US Census Bureau.

Note: The binscatter command of Stata is used to produce this figure. The x-axis in each subplot is the income growth in each income quintile relative to the mean income of the county. Relative income gain is grouped into 20 equally sized bins. The position of each point in the graph is the mean value of the change in the Gini coefficient and mean value of relative income growth for one of these bins. All growth rates and changes are calculated between the years 1999 and 2011.

In this section I decompose the change in income distribution into changes in income at different quintiles. This enables me to further evaluate the potential explanations for the relationship between house prices, mortgage debt and income inequality.

Figure A.2 plots the change in income inequality against the relative income gains for each of the five income quintiles and the top 5%. The relative income gain for quintile j in county i is the growth of mean income in that quintile X_i^j relative to the mean income growth for a given county $\bar{X}_i$.

$$x_i^j = \Delta_t \ln \left(\frac{X_i^j}{\bar{X}_i} \right)$$

The figure suggests that a change rise in income inequality is associated with both low relative income growth at the bottom 80 percent population and high relative income growth at the top of the income distribution. Therefore, at the cross-section, counties that experienced high increase in income inequality saw declines for the lowest 4 income quintiles and increases for the top income quintile relative to the mean. The message is similar to the one from Figure A.7.

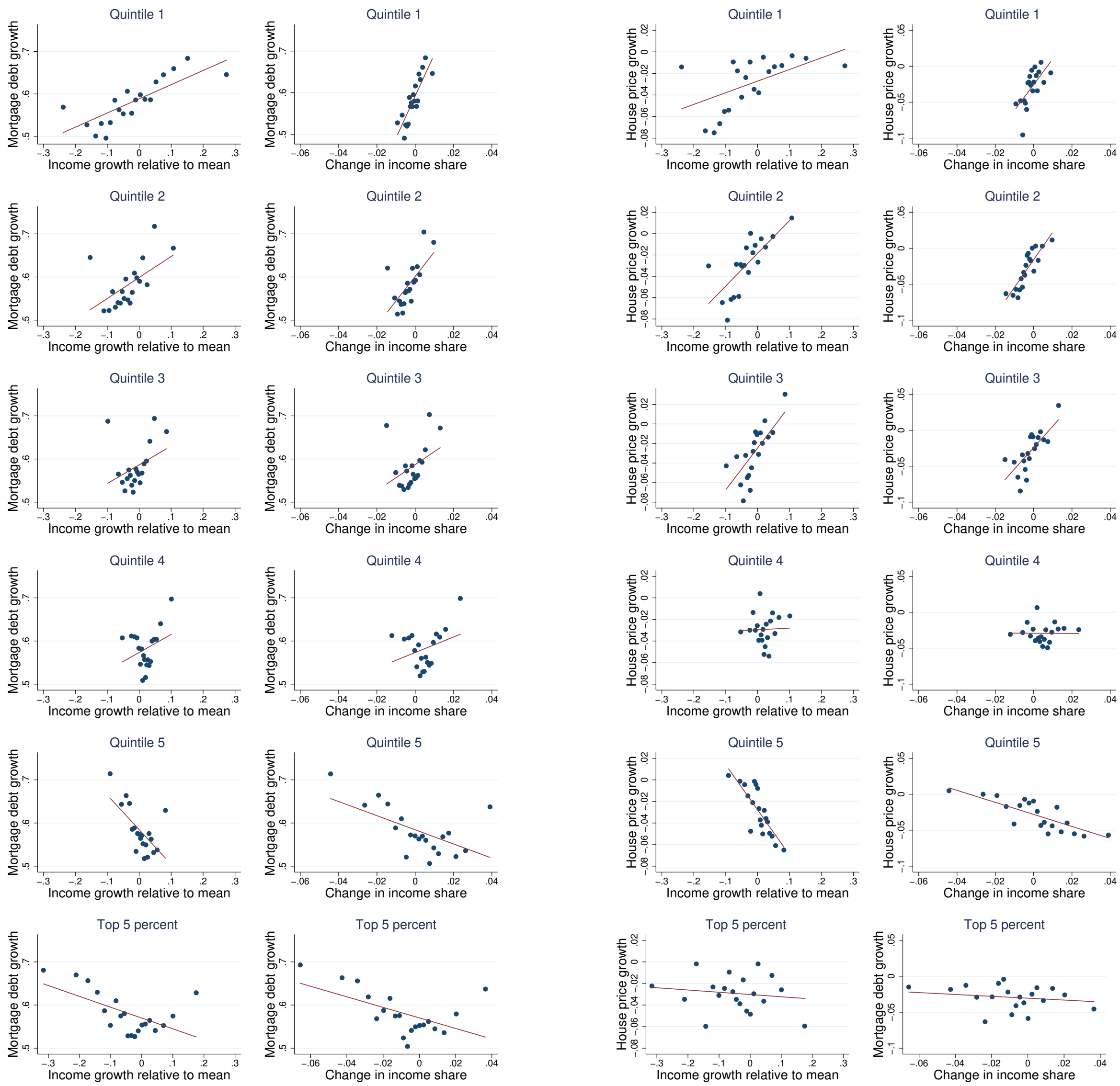
The first column of Figure A.3 shows the relationship between debt growth and relative income gains for different income quintiles. Consistent with the mechanism proposed in this paper, as long as incomes for the low quintiles fare well, mortgage debt increases. That is, relative income gains for the bottom 60% of the population are positively associated with mortgage debt growth. On the other hand, the cross-sectional data suggests that large income gains at the higher end of the income distribution, i.e. of the top income quintile or the top 5 percent, are negatively correlated with mortgage debt growth. The second columns of Figure A.3 implies similar dynamics by displaying the relationship between mortgage debt growth and change in income share of different income quintiles. At the cross-section an increase in income shares of the top earners, i.e. top 20 % or top 5%, are negatively correlated with debt growth. This finding contradicts explanations based around higher income gains at the top of the distribution leading to an increase in debt.

Moving on to the third column of Figure A.3 I show that income gains at the lower end of the income distribution are positively related to house price growth. This finding is consistent with the mechanism proposed in this paper, and inconsistent with explanations that predict an increase in house prices together with large relative gains in top incomes. The last column of Figure A.3 confirms this prediction by showing the relationship between house price growth and change in income share of different income quintiles. An increase in the income shares of top 5% and the top 20% of the income distribution is negatively associated with house price growth. In the main text I show that this is the case also when state level data is used.

Taking stock

The fact that higher income inequality is associated with lower debt, higher delinquencies and lower house prices is consistent with the following explanation. An increase in income inequality worsens the pool of borrowers, in the sense that they are more likely to default. Mortgage debt falls as lenders price in the increased risk from the change in the pool of borrowers. This leads to lower housing demand and prices. The theoretical model described in this paper formalizes this intuition.

Figure A.3: Real mortgage debt growth and quintile relative income growth between the years 1999 and 2011



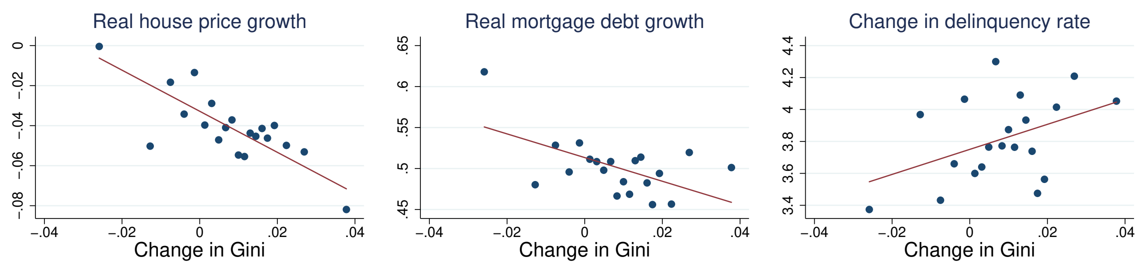
Source: US Census Bureau, New York Fed Consumer Credit Panel.

Note: Binned scatter plots with state fixed effects, mean income growth, population growth, the share of subprime borrowers in 2000, median income in 1999, and the number of households in 1999 as controls. First two columns presents the partial correlations on different income quintiles and mortgage debt growth and the last two columns displays the same for house price growth.

A.1 Controlling for housing supply elasticity

In this section, I first show that stylized facts presented in Figure 2 are robust to inclusion of housing supply elasticity as a control variable. If housing supply elasticity is a common driver of house prices and income inequality, then it is essential to control for it to study whether income inequality is an independent vector affecting house prices. Mian and Sufi (2009) uses housing supply elasticity as an instrument for expected house price growth which might confound with the effect of income inequality in affecting demand for mortgage debt and expected risk of default.

Figure A.4: Changes in income inequality, real house price growth, mortgage debt growth and change in mortgage delinquency rate over US counties between the years 1999 and 2011



Source: US Census Bureau, New York Fed Consumer Credit Panel, Federal Housing and Finance Agency, Bureau of Labor Statistics, Saiz (2010).

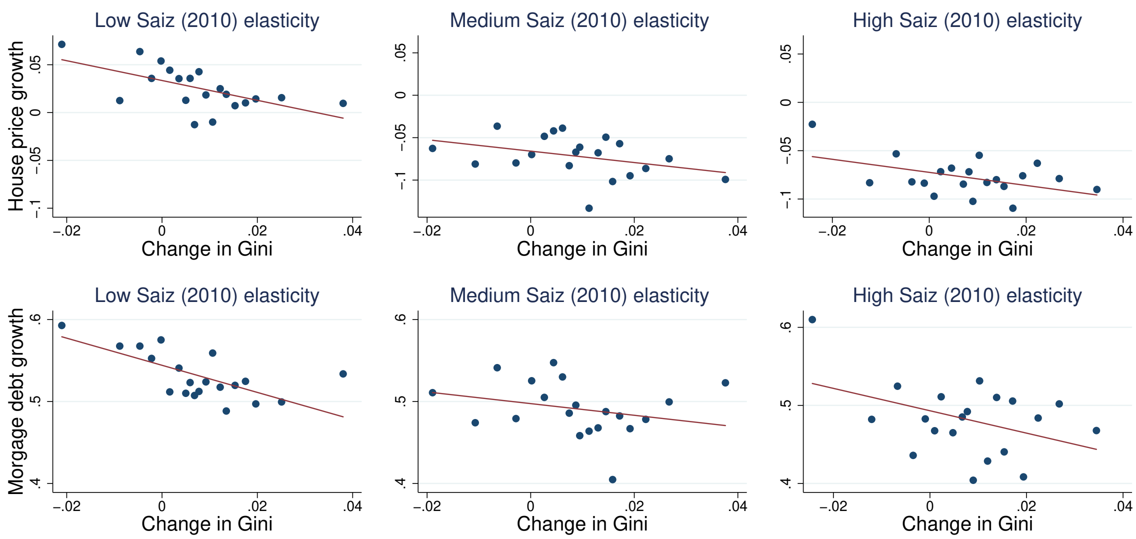
Note: To construct this figure I use the binscatter command in Stata, which regresses the three title variables on the change in Gini coefficient, Saiz (2010) housing supply elasticity, state fixed effects, mean income growth, population growth, the share of subprime borrowers in 2000, median income in 1999, and the number of households in 1999. The slope of the line of fit is the coefficient for the change in Gini coefficient in this regression. For the data points, it first obtains the residuals from regressions of the title variable and the change in Gini coefficient on the other control variables. These are then grouped in twenty equally sized bins for the Gini coefficient residual. The position of each point is the mean value of the title variable residual and Gini coefficient residual for one of these bins. All growth rates and changes are calculated between the years 1999 and 2011.

Figure A.4 plots the partial correlation with the change in Gini coefficient between 1999 and 2011 for three variables using data from US counties. These plots include county level controls. The first panel shows the relationship between the change in Gini coefficient and real house price growth, the second the relationship with real mortgage debt growth, and the third the relationship with the change in the delinquency rate. In constructing this figure I control for a variety of county characteristics including the housing supply elasticity measured by Saiz (2010). This measure is available for a subset of counties and reduces the sample size from 2093 to 746.[5] Therefore, even when controlled for housing supply elasticity, cross-sectional correlations qualitatively remain

[5]Saiz (2010) housing supply elasticity measure is available at the metropolitan statistical area (MSA) level, I assume that the counties in the same MSA have the same elasticity.

the same. This is not surprising since counties with low housing supply elasticity are on average densely populated and I control for household size in Figure 2.

Figure A.5: Changes in income inequality and real house price growth in US counties with different housing supply elasticity between the years 1999 and 2011



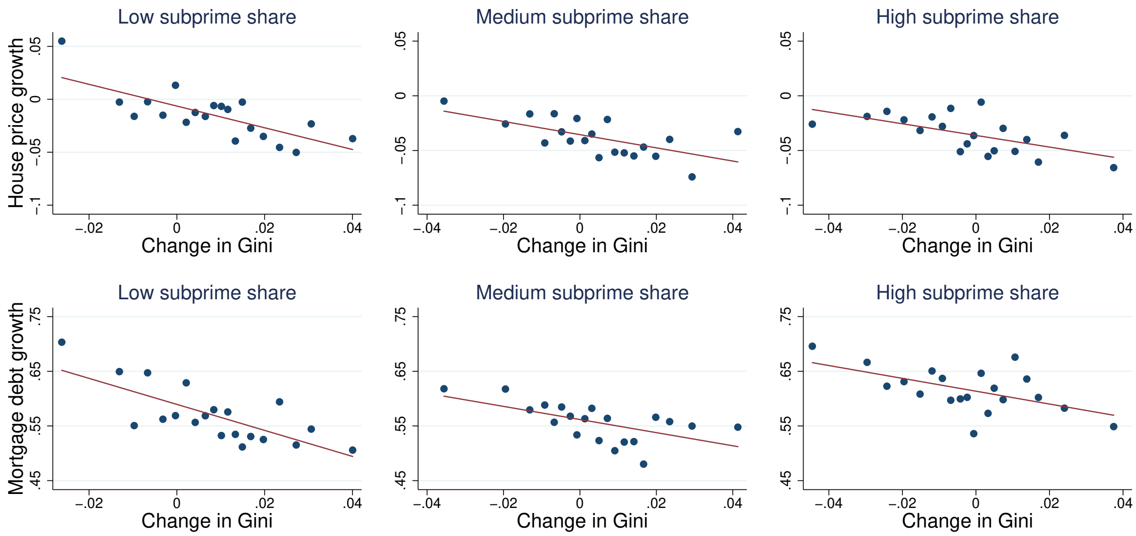
Source: US Census Bureau, New York Fed Consumer Credit Panel, Federal Housing and Finance Agency, Bureau of Labor Statistics, Saiz (2010).

Note: To construct this figure I use the binscatter command in Stata, which regresses the real mortgage debt growth on the change in Gini coefficient, state fixed effects, mean income growth, population growth, the share of subprime borrowers in 2000, median income in 1999, and the number of households in 1999 for each elasticity group. The slope of the line of fit is the coefficient for the change in Gini coefficient in this regression. For the data points, it first obtains the residuals from regressions of real mortgage debt growth and the change in Gini coefficient on the other control variables. These are then grouped in 20 equally sized bins for the Gini coefficient residual. The position of each point is the mean value of real mortgage debt growth residual and Gini coefficient residual for one of these bins. All growth rates and changes are calculated between the years 1999 and 2011. Counties are grouped into three according to housing supply elasticity. Low and high correspond to the lowest and the highest Saiz (2010) housing supply elasticity terciles, respectively.

Next, I consider whether the association of mortgage debt and house price growth with income inequality is qualitatively different across high and low housing supply elasticity areas. I group counties into three categories depending on their Saiz (2010) elasticity measure. The first column in Figure A.5 represents the counties at the lowest tercile of supply elasticity. The first row charts of the figure shows that house price growth and change in income inequality is negatively correlated at the cross-section, independent of the level of supply elasticity. Therefore, the results in Figure 2 is not reflecting the dynamics of low housing supply elasticity areas that would on average be expected to have the largest house price changes. Finally, the second row of Figure A.5 depicts that real mortgage debt growth and change in income inequality are negatively associated in each housing supply elasticity group controlling for county characteristics.

A.2 Controlling for the share of subprime borrowers

Figure A.6: Changes in income inequality and real mortgage debt growth in US counties with different initial subprime credit population share between the years 1999 and 2011



Source: US Census Bureau, New York Fed Consumer Credit Panel, Federal Housing and Finance Agency, Bureau of Labor Statistics.

Note: To construct this figure I use the binscatter command in Stata, which regresses the house price growth on the change in Gini coefficient, state fixed effects, mean income growth, population growth, the share of subprime borrowers in 2000, median income in 1999, and the number of households in 1999 for each elasticity group. The slope of the line of fit is the coefficient for the change in Gini coefficient in this regression. For the data points, it first obtains the residuals from regressions of the title variable and the change in Gini coefficient on the other control variables. These are then grouped in 20 equally sized bins for the Gini coefficient residual. The position of each point is the mean value of the house price growth residual and Gini coefficient residual for one of these bins. All growth rates and changes are calculated between the years 1999 and 2011. Counties are grouped into three according to share of subprime credit population share as of 2000. Low and high correspond to the lowest and the highest share of subprime credit population, respectively.

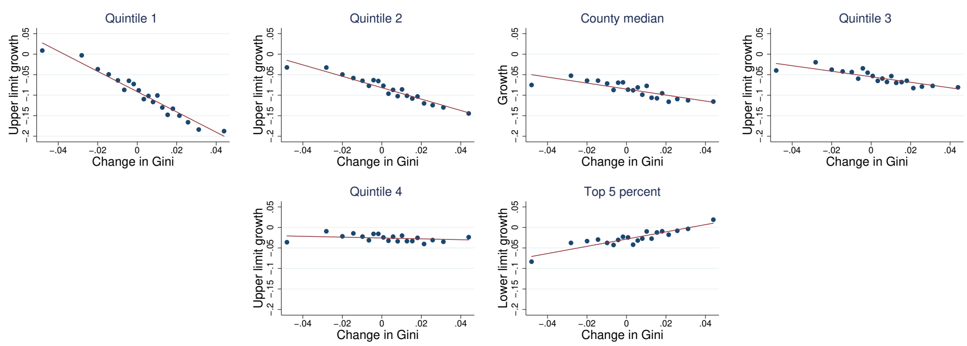
In this section I show that the negative association of income inequality with both house prices and mortgage debt holds for subsamples of counties with different share of subprime credit population share as of 2000.[6] Mian and Sufi (2009) shows that ZIP codes with high share of subprime borrowers observed high debt and house price growth in the run-up to the financial crises and proposes a supply side view of the credit boom. In order to control for this credit supply effect, I study whether negative association of income inequality with house prices and with mortgage debt growth is present at a closer investigation that can to some extent control for location specific lending practices.

[6]The data includes a larger fraction of counties if I consider the share of subprime credit population in year 2000 instead of year 1999.

Figure A.6 displays the partial correlations of house price and mortgage debt growth with the Gini coefficient having controlled for county characteristics. I group counties into three categories depending on their subprime population share. The first column in Figure A.6 corresponds to the counties at the lowest tercile of subprime borrower share. The figure shows that both debt and house price growth are negatively associated with the change in income inequality independent of the subprime population share.

A.3 Change in the Gini coefficient and income quantile limits

Figure A.7: The relationship between income inequality and the upper income limit of different income quintiles



Source: US Census Bureau.

Note: The y-axis of each subplot is the real growth rate of the upper limit of a particular income quintile, median or lower limit of top 5 percent. The binscatter command of Stata is used to produce this figure. Change in the Gini coefficient is grouped into 20 equally sized bins. The position of each point in the graph is the mean value of the change in the Gini coefficient and mean value of y-axis variable for one of these bins. All growth rates and changes are calculated between the years 1999 and 2011.

In my model a mean preserving increase in income inequality gives rise to a rise in the share of population below $\bar{y}$, or equivalently, $\bar{y}$ in the more unequal income distribution corresponds to a higher income percentile. Figure A.7 plots the cross-sectional correlation between the change in the Gini coefficient and upper limits of different income quintiles, median income and the lower limit of top 5 percentile between the years 1999 and 2011. The Figure shows that an increase in income inequality is associated with an increase in the lower limit of the top 5 percentile only. That is, areas that experienced large increases in income inequality tended to experience large declines in income limits in the lowest three quintiles and the median county income. A rise in income inequality is associated with a lower decline in the 80^{th} income percentile. This implies that a rise income inequality at the cross-section corresponds to a more than half of the population that has 2011 real incomes below the median real income of 1999.

B State level data

B.1 A further reality check: a cross-section of US states

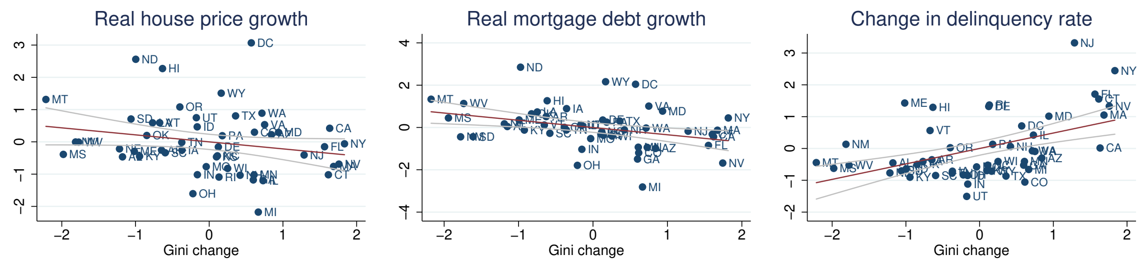


Figure B.8: Income inequality, real house prices, mortgage debt and mortgage delinquency rate in US States between 2003 and 2015

Source: New York Fed Consumer Credit Panel, Federal Housing and Finance Agency, Bureau of Labor Statistics.

Note: This figure uses the normalized average annual change of each variable. For the Gini coefficient, for instance, it is calculated as follows. I first compute the annual change in the Gini coefficient for each year between 2003 to 2015. I then calculate the average change for each state and the across state mean and standard deviation of average changes. The value for each state is its average change in Gini coefficient net of the across state mean and divided by the across state standard deviation. A state that takes value 2 in the x-axis of each panel experienced an increase in income inequality 2 standard deviations above that of the across state mean. The slope of the regression line is the estimated coefficient of a between regression estimated of each variable on the change in the Gini coefficient. Both the dependent variable and the Gini coefficient are normalized annual changes.

Figure B.8 displays the results of between regressions of annual real house price growth, real mortgage debt growth and the change in mortgage delinquency on the change in income inequality. The purpose of this exercise to display that there is sufficient variation across states and the result presented in the previous section is not coming from time variation that is not absorbed by year fixed effects. All variables are normalised average annual changes between 2003-2015. That is, each value is computed by subtracting the mean value for a state from the average annual change of the variable, and then dividing by the standard deviation. For example, Nevada and New York experienced increases in income inequality about two standard deviations greater than the mean increase across states, while the District of Columbia saw an average annual real house price growth that is about three standard deviations above the mean over states for this time period. Qualitatively, Figure B.8 displays results in line with those for US counties depicted in Figure 2, despite the source of the Gini coefficient being different and averages being calculated over a larger number of time periods.[7]

C First-step regression results

Here I present the first step regressions underlying the 2SLS estimations in the main text. Table 1 presents the estimates for the period 1992-2013 and Table 2 presents those for the sample between 2003 and 2015. As I discuss in the main text, for the first sample using all five of the Gini instruments imply rejection of under- and -weak- identifications tests for both of the specification with Gini, and Gini and its real rate interaction as the endogenous regressors. While the former is the regression specification (23) and the latter is (24) in the main text.

Table 1: First-step regressions with different digit NAICS industry instruments
Panel: US states from 1992 to 2015

	Gini	Gini	Gini	Gini	Gini	Gini	Gini	Gini × Real rate	Gini	Gini × Real rate
2-digit instrument	0.24***					0.47***	0.48**	-0.94		
	(0.08)					(0.12)	(0.20)	(0.83)		
3-digit instrument		0.14				-0.81***	-1.01***	0.33		
		(0.08)				(0.16)	(0.35)	(1.48)		
4-digit instrument			0.23**			-0.86*	-0.33	5.03**		
			(0.09)			(0.46)	(0.83)	(2.43)		
5-digit instrument				0.27***		1.09***	0.83	-6.26***		
				(0.09)		(0.39)	(0.77)	(2.28)		
6-digit instrument					0.32***	0.48***	0.47	1.73*	0.33***	-0.45
					(0.08)	(0.13)	(0.34)	(0.98)	(0.11)	(0.46)
2-digit × Real rate							0.01	0.63*		
							(0.07)	(0.32)		
3-digit × Real rate							0.08	-0.41		
							(0.11)	(0.52)		
4-digit × Real rate							-0.31	-1.94**		
							(0.29)	(0.95)		
5-digit × Real rate							0.18	2.36***		
							(0.28)	(0.83)		
6-digit × Real rate							0.00	-0.51	-0.01	0.20
							(0.09)	(0.32)	(0.03)	(0.15)
Observations	1200	1200	1200	1200	1200	1200	1200	1200	1200	1200
R-squared within	0.64	0.64	0.64	0.65	0.65	0.67	0.67	1.00	0.65	1.00
F-test	8.63	2.77	6.49	9.14	14.21	10.42	7.12	2.26	7.33	0.91

*p<0.1, **p<0.05, ***p<0.01.
All regressions include year and state fixed effects, and control variables log real mean state income, log population, log new housing permits and homeownership rate. Errors are clustered at the state level. Gini is instrumented with shift-share instruments which are derived using predetermined (2-year lagged) industrial employment shares in a given state and national wage growth in each industry. Columns (1) to (6) present first stage regression results when Gini coefficient is the only endogenous regressor, columns (7) to (10) presents when both the Gini coefficient and its interaction with 10-year real rate are both endogenous regressors. Columns (1) to (5) include one digit industry at a time, from NAICS 2- to 6-digit industry shares and national wage growth, respectively. Column (6) includes all instruments. In columns (7) and (8) all five instruments' real rate interactions are included, while columns (9) and (10) uses the highest digit industry instrument and its real rate interaction. F-test displays the F-statistic for the null hypothesis that coefficients of the instruments are zero. The F-statistics are robust to heteroskedasticity and clustering on states.

[7] County level data gives the growth between the years 1999 and 2011, whereas state level data is an average of 13 annual changes. County level inequality data is calculated from Census Surveys, whereas state level inequality data is calculated from IRS tax returns.

Table 2: First-step regressions with different digit NAICS industry instruments
Panel: US states from 2003 to 2015

	Gini	Gini	Gini	Gini	Gini	Gini	Gini	Gini × Real rate	Gini	Gini × Real rate
2-digit instrument	0.48*					-0.44	0.21	0.39		
	(0.07)					(0.31)	(0.53)	(0.63)		
3-digit instrument		0.84***				-0.28	-0.91	-1.42		
		(0.00)				(0.66)	(0.18)	(0.34)		
4-digit instrument			0.93***			0.79	-0.52	1.84		
			(0.00)			(0.47)	(0.69)	(0.27)		
5-digit instrument				0.95***		0.02	1.47	-1.91		
				(0.00)		(0.98)	(0.20)	(0.23)		
6-digit instrument					1.04***	0.82**	0.55	1.23	1.09***	0.28
					(0.00)	(0.05)	(0.31)	(0.24)	(0.00)	(0.49)
2-digit × Real rate							-0.62***	-0.31		
							(0.00)	(0.59)		
3-digit × Real rate							0.24	-0.05		
							(0.57)	(0.96)		
4-digit × Real rate							1.99***	1.79		
							(0.01)	(0.26)		
5-digit × Real rate							-1.81***	-0.81		
							(0.00)	(0.55)		
6-digit × Real rate							-0.02	-1.00	-0.31***	-0.46**
							(0.93)	(0.10)	(0.00)	(0.03)
Observations	650	650	650	650	650	650	650	650	650	650
R-squared within	0.37	0.39	0.42	0.42	0.43	0.44	0.58	1.00	0.51	1.00
F-test	3.31	9.29	12.55	14.69	22.14	4.72	11.92	2.43	17.65	2.42

*p<0.1, **p<0.05, ***p<0.01, p-values in paranthesis.
All regressions include year and state fixed effects, and control variables log real mean state income, log population, log new housing permits and homeownership rate. Errors are clustered at the state level. Gini is instrumented with shift-share instruments which are derived using predetermined (2-year lagged) industrial employment shares in a given state and national wage growth in each industry. Columns (1) to (6) present first stage regression results when Gini coefficient is the only endogenous regressor, columns (7) to (10) presents when both the Gini coefficient and its interaction with 10-year real rate are both endogenous regressors. Columns (1) to (5) include one digit industry at a time, from NAICS 2- to 6-digit industry shares and national wage growth, respectively. Column (6) includes all instruments. In columns (7) and (8) all five instruments' real rate interactions are included, while columns (9) and (10) uses the highest digit industry instrument and its real rate interaction. F-test displays the F-statistic for the null hypothesis that coefficients of the instruments are zero. The F-statistics are robust to heteroskedasticity and clustering on states.

References

Mian, Atif and Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, 2009, *124* (4), 1449–1496.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.