Multi-establishment Firms, Pricing and the Propagation of Local Shocks: Evidence from US Retail*

Ezequiel García-Lembergman†

Pontifical Catholic University of Chile

Dec 10, 2021

[Link to the latest version]

Abstract

I study whether and how retail chains and their geographic distribution of stores contribute to the propagation of shocks across regions in the United States. Linking detailed store scanner micro-data to a county-level house price dataset for the period of the Great Recession, I investigate the spread of house-price induced local shocks through the networks of retail chains. My main empirical finding is that county-level prices are sensitive to shocks in distant counties that happen to be served by the same retail chains. A 10% drop in house prices in other counties that are served by the same retailers leads, on average, to a 1.4% decline in the local consumer retail price index. My results hold after conditioning on trade relationships due to geographic proximity. In fact, I document that once the retail chains' channel of propagation is taken into account, trade relationships due to geographic proximity play no role in propagating shocks to retail prices. I rationalize the reduced-form estimates in a model in which retail chains vary prices uniformly across their stores as a function of changes in market demand that they face at the (aggregate) chain level. I find that the calibrated model with uniform pricing can fully account for the reduced-form effects. Counterfactual analysis shows that uniform pricing and the geographic distribution of retail chains reduced cross-county dispersion of inflation by 40% during the Great Recession, benefiting consumers from low-income counties that were less exposed to drops in local house prices.

*I am extremely grateful to Benjamin Faber, Andrés Rodriguez-Clare, Cecile Gaubert and Stefano DellaVigna for their continuous support and guidance in this project. I also want to thank Facundo Albornoz, Matilde Bombardini, Chris Campos, Thibault Fally, Alex Favela, Andrés González-Lira, Yuriy Gorodnichenko, Pierre-Olivier Gourinchas, Matthias Hoelzlein, Leticia Juarez, Patrick Kennedy, Isabela Manelici, Ted Miguel, Emi Nakamura, Mathieu Pedemonte, Piyush Panigrahi, Todd Messer, Joseph Shapiro, Darío Tortarolo, Jose P. Vásquez and Roman Zarate for very helpful comments and discussions. Researcher own analyses calculated based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

†Department of Economics, UC Berkeley, 666 Evans Hall, Berkeley, CA. Email: eglembergman@berkeley.edu. Website: www.elembergman.com.

1 Introduction

The propagation of local shocks through the economy has been a central topic in macroeconomics. The most common way economists have been thinking of how shocks propagate is through trade linkages, both in inputs and outputs, as well as through flows of factors of production, such as labor or capital. How does our understanding change once we take into account the role of multi-establishment firms with presence in many regions?

If the various establishments of these firms operate as independent business units, these firms will play no additional role in propagating shocks across regions. However, some characteristics of these firms can create inter-dependencies between their establishments. For example, if firms have a common pricing strategy across their establishments, then shocks in one region will affect prices in all their locations.[1] There are reasons to think that this phenomenon occurs for an important sector of the US economy: retail trade -i.e. grocery, drugstore and mass merchandise stores.[2] Given that the distribution of retail chains across regions in the U.S is far from uniform, their networks of stores might affect how shocks are propagated across the economy. Yet, surprisingly, the existing literature has so far paid relatively little attention to how the spatial networks of retail chains stores shape the propagation of shocks across regions.

In this paper, I fill this gap. I study whether and how local demand shocks propagate to consumer prices in distant regions through retail chain networks. Exploiting regional variation in local consumer demand from the collapse in house prices during the Great Recession, I find that the local consumer price index depends not only on local demand conditions, but also on shocks in distant regions that happen to be served by the same retail chains.

My main analysis is based on store-level scanner data from the Nielsen-Kilts retail panel. The data includes sales and prices in more than 35,000 participating grocery, drug and mass merchandise stores in more than 2,000 counties. In addition, it provides the location of each store and a unique identifier for the retail chain that is the ultimate owner of the store. I combine this with data on county-level changes in house prices during the Great Recession (2007-2011) from the Federal Housing Finance Agency (FHFA).

In order to identify the role of retail chains in propagating shocks across counties in the U.S., I use the Nielsen data to construct a spatial network of the retail chains' stores. Given that retail chains are unevenly distributed in space, their spatial network naturally creates linkages between counties. At the center of my analysis is a new measure of connectedness that characterizes the exposure between each pair of counties. The bilateral exposure of county c to county k is a weighted average of the share of each retail chain's national sales that take place in county k, where the weights are given by the market share of each retail chain in county c.[3] Intuitively, a county c will be more exposed to county k if county k is an important market for retail chains that are dominant

[1] Another example is technology transfer from the parent company to their affiliates, which can generate co-movement in the sales growth of multinationals across countries (Cravino and Levchenko (2017)).

[2] DellaVigna and Gentzkow (2019) show that most US food, drugstore, and mass merchandise chains charge nearly-uniform prices across stores, despite wide variation in consumer demographics and competition. Nakamura (2008) shows that retail chains' fixed effects explain a substantial share of price variation.

[3] Formally, exposure of county c to county k is given by: $\omega_{ck} = \sum_{r} l_{rc} S_{rk}$. S_{rk} denotes the share of retail chain r's national sales that take place in market k. l_{rc} denotes the share of a county's sales that correspond to retail chain r (See Equation 4.4). Note that, unlike bilateral distance, this measure of connectedness is intrinsically asymmetric.

in county c.[4] I then use these bilateral measures in order to compute a county's exposure to house price changes in other counties. The exposure of county c to house price changes elsewhere is the network-weighted percentage change in house prices, where the weights are given by the bilateral exposure of county c to each county k.[5]

My main empirical finding is that county-level prices are sensitive to house price-induced local shocks in distant counties that are linked by retail chain networks. In order to deal with endogeneity of house price changes, I follow the identification strategy in Mian and Sufi (2011) and use the local housing supply elasticity to instrument for local house price changes. More importantly, I use the network-weighted housing supply elasticity to instrument for network-weighted house price changes in other counties.[6] Across a variety of empirical specifications, I find a positive elasticity of local prices to house price movements in other counties, linked by the retail chains' networks. Specifically, a ten percent drop in house prices in distant counties linked by the network of retail chains leads, on average, to a 1.4 percent decline in county-level consumer prices.

My identification strategy faces several challenges. Since retail chains are not placed randomly across space, the most important challenge is that it is hard to distinguish propagation of shocks through the network of retail chains from common shocks in the regions in which the retail chains operate. Notably, to minimize transportation costs, retail chains might locate their stores in nearby counties, where wages and prices co-move due to, for example, integrated labor markets or trade relationships. In a series of empirical exercises, I show that common shocks to regions linked by the network of retail chains do not explain my results.

First, I document that my results hold after conditioning on trade relationships due to geographic proximity. In fact, once the retail chains' channel of propagation is taken into account, trade relationships due to geographic proximity play no role in propagating shocks to retail prices. Furthermore, I exclude counties within the same state and show that the shocks also spread to distant regions located outside the state. Second, I do not observe co-movement in wages between counties that are linked by the retail chains' networks. Finally, and more importantly, in a complementary empirical strategy, I turn to more granular data at the store-by-county level. This allows me to include county-by-time fixed effects, which absorb any common variation within a county that is due to a regional shock, regardless of whether the shock is specific to that county or correlated with shocks in other counties. Using variation in price changes across stores within a given county, I find an elasticity of store-level prices with respect to house price movements in other counties of 0.12-0.20. That is, a ten percent larger reduction in weighted average house prices in counties where the same retail chain operates leads, on average, to a 1.2-2.0 percent reduction in the local prices of that chain's establishments relative to prices of stores belonging to other chains.[7]

After showing how the geographic distribution of retail chains shapes the propagation of local demand shocks to consumer prices, I explore whether it also affects other economic outcomes.

[4]This paper is the first to characterize the bilateral linkages between counties in the retail chains dimension. I make the matrix that contains the bilateral linkages accessible to members of the broader research community.

[5]County $\Delta \log HP(others)_c = \sum_{k \neq c} \omega_{ck} \Delta \log HP_k$, where HP denotes house price index (See equation 4.3).

[6]Several papers have used this instrument for changes in local house prices (Mian and Sufi (2011), Adelino et al. (2015), Stroebel and Vavra (2019), among others).

[7]This result also relates to recent work by Handbury and Moshary (2020) on the effects of the national school lunch program on retail prices. See discussion of related literature below.

I find no similar effects on local wages or employment, which implies that those are unaffected by house price shocks in other counties that are linked by retail chains networks. This, in turn, suggests that retail chains and their geographic distribution of stores affects not only consumers' prices in distant regions, but also consumers' purchasing power.

In the final part of the paper, I propose a model of retail chains' pricing decisions to quantitatively interpret my findings and evaluate the role of retail chains in connecting economically counties in the U.S. Retailers compete under monopolistic competition and charge markups that can vary as a function of local demand conditions. In particular, I allow demand elasticity (and markups) to vary with changes in local house prices. In the model, retail chains' headquarters can either set a uniform price or price-discriminate across their locations. If a retail chain sets uniform prices, its optimal price depends on a weighted average of the demand conditions in the different markets where the chain operates. Hence, when faced with a negative demand shock in a given region, the retail chain decreases its prices not only in that region, but also in other regions in which the retail chain operates. The extent to which the retailer contributes to the propagation of demand shocks depends on the geographic distribution of its sales.

The model presented here provides a novel, yet intuitive test for uniform pricing responses to shocks. If retail chains charge uniform prices, then in response to a local demand shock, the effect from shocks elsewhere should be equal to the local effect, once both are weighted properly.[8] According to the results of this test, I cannot reject the hypothesis of uniform pricing strategies. The model also highlights that the effect of local shocks on local retail price indices is heterogeneous: under uniform pricing, the pass-through of local shocks to local prices becomes a function of how important the local consumer market is in the national sales of the retailers that enter the local consumption basket. I find that this degree of local pass-through varies greatly across U.S. counties.

The theoretical framework naturally lends itself to quantifying the aggregate and distributive consequences of uniform pricing during the Great Recession. I conduct two counterfactual exercises that explore 1) different pricing strategies, and 2) different spatial networks of retail chains due to mergers or acquisitions. In the first counterfactual, I consider flexible pricing strategies. Compared to uniform pricing, under flexible pricing the counterfactual cross-county dispersion of inflation is 40% larger. This implies that uniform pricing smoothed-out the effect of the shocks across regions during the Great Recession. Then, I explore the distributive consequences of uniform pricing and I find that consumers in low-income counties benefited. Intuitively, these counties were not as affected by the house price slump directly, but still experienced a drop in their local consumer prices because their retail chains were affected by the shocks in other counties.

In the second counterfactual, I assume uniform pricing and explore changes in the spatial networks of retail chains due to mergers and acquisitions. I simulate a scenario in which there is a merger between the largest retail chain in each of the four census regions of the United States. This merger would have reduced the cross-county dispersion of inflation rates during the Great Recession by 12%. This result emphasizes a new aspect of mergers: they can intensify the linkages between regions and lead to more synchronized changes in consumer prices.

[8] The proper theoretical weight for the local shock is related to the importance of the local consumer market for the dominant retail chains operating in the county.

This paper contributes to several strands of the literature. First, it contributes to the international and intra-national trade literature that studies how shocks propagate across regions. Most of this literature emphasizes input-output linkages between firms and sectors as the main driver of the propagation of shocks (e.g. Acemoglu et al. (2012), Carvalho et al. (2016), and Caliendo et al. (2017), Stumpner (2019)). There has been less study of how shocks propagate within firms' internal networks of establishments. In the international context, a number of papers explore whether characteristics of parent companies and affiliates of multinationals generate co-movements between regions, including channels such as internal capital markets (Desai et al. (2009)), intermediate input linkages (Boehm et al. (2019)), and transmission of technology across establishments (Cravino and Levchenko (2017), Alviarez et al. (2020), and Bilir and Morales (2020)). My paper focuses on the propagation of shocks across firms' establishments in a given country and sheds light on a new channel: inter-dependencies in pricing strategies between the establishments of a firm. An advantage of studying propagation of shocks within the U.S. is that Nielsen scanner data allows me to observe establishment-level prices. As a consequence, unlike previous papers, I am able to directly study how shocks propagate to consumer prices.

Less has been done to understand the role of firms' internal networks in propagating shocks in the domestic context, with two notable exceptions: Hyun and Kim (2019) and Giroud and Mueller (2019). Hyun and Kim (2019) study U.S. manufacturers that are located in a given region who sell (*export*) multiple products to multiple regions (*multi-destination firms*). They find that a negative demand shock in a region can induce producers to substitute production of high-quality products for low-quality products, creating co-movement in sales across markets. In contrast, I study firms in the non-tradable sector that have establishments in multiple regions and sell locally in each one (*multi-establishment firms*).[9] My findings are compatible with multi-market producers changing their product mix, but I focus on a different mechanism. Holding constant the set of products (and quality) sold by the retailer, local demand shocks affect prices of continuing products in distant markets that are connected by the retail chains' network of stores.[10]

In a study more closely related to my paper, Giroud and Mueller (2019) analyze the role of firms' cashflow constraints in propagating local demand shocks. They find that financially constrained firms spread local demand shocks, affecting employment in distant regions where the parent firm operates. The channel I study is different: regardless of retail chains' financial constraints, their pricing strategies create inter-dependencies between their establishments. Note also that while the results of Giroud and Mueller (2019) have implications for the firms' workers, the results in my paper have implications for their consumers.

Second, this paper relates to the literature that studies the collapse in house prices during the Great Recession (Mian and Sufi (2011), Mian et al. (2013), Giroud and Mueller (2019), and Stroebel and Vavra (2019)). In particular, this paper is closely related to Stroebel and Vavra (2019), which analyzes how markups and prices respond to local demand shocks during the Great Recession. They find a large positive effect of local house price movements on local retail prices. However, if retail chains set prices nationally, markups not only depend on local demand, but also on demand

[9] The international trade literature also makes a sharp distinction between firms exporting to many markets (multi-destination exporters) and firms that have establishments in many markets (multi-nationals). In my context, retail chains occupy the second category.

[10] Results in Hyun and Kim (2019) are entirely driven by extensive margin adjustments. In contrast, my results are entirely driven by changes in prices of continuing products.

shocks in other regions. I complement their results by showing that house price-induced local demand shocks not only affect local prices, but also prices in distant regions that are linked by the retail chains' networks. This finding advances our understanding of the mechanisms through which the Great Recession propagated across the U.S economy.

Third, this paper helps reconcile the conflicting findings in DellaVigna and Gentzkow (2019) and in Stroebel and Vavra (2019). While DellaVigna and Gentzkow (2019) show that retail chains charge similar prices in all their stores, Stroebel and Vavra (2019) find large effects of local demand shocks on retail prices. I show that once I control for shocks from regions linked by the same retail chains, the elasticity of local retail prices with respect to house price changes decreases by 35%. In addition, the predictions from my theoretical model indicate that uniform pricing can be consistent with large local responses to demand shocks, as long as the regions experiencing the shocks are an important market for their retail chains. On this ground, this paper highlights that to understand how local markups respond to changes in local economic conditions, one should consider the decisions of national retail chains and their spatial network of stores.

Fourth, my paper relates to a literature that studies supply-side effects of demand side policies. In a related paper, Handbury and Moshary (2020) study the effects of the National School Lunch Program (NSLP) on retail prices. Consistent with zone pricing, they find that retail chains highly exposed to the NSLP reduced prices on all their outlets. While their study focuses on the policy implications of the NSLP, this paper studies the geography of multi-establishment firms to quantify a new mechanism through which local economic shocks propagate across regions of the U.S. I derive measures of bilateral market exposure directly from a theory of firm pricing decisions, and find quantitatively that the reduced-form empirical findings can be accounted for by uniform pricing. This also allows me to quantify the distributive consequences of alternative pricing strategies.

Finally, this paper also contributes to the literature in industrial organization and macroeconomics that studies firms' pricing dynamics, and, in particular, a recent literature that has documented the existence of uniform pricing in different sectors. Nakamura (2008) shows that retail chains' effects explain a substantial share of price variation in the U.S. In the same context, DellaVigna and Gentzkow (2019) show that a given retail chain charges similar prices across different locations. Daruich and Kozlowski (2020) document uniform pricing in supermarket chains in Argentina and analyze its implications for inferring aggregate elasticities from local estimates. Cavallo (2017) finds that retailers charge similar prices online and offline. Adams and Williams (2019) find similar patterns in the retail home improvement industry. I take the insights of these papers to empirically assess the role the geographic distribution of retail chains in shaping the propagation of local demand shocks across regions in the United States.

The remainder of the paper proceeds as follows. Section 2 describes the data. Section 3 documents stylized patterns regarding the geographic distribution of retail chains. Section 4 presents the empirical analysis. Section 5 presents the model. Section 6 presents the quantitative analysis and explores counterfactual scenarios. Section 7 concludes.

2 Data

I combine county-level data on changes in house prices during the Great Recession (2007-2011) with Nielsen Retail Scanner data.

2.1 Retailer Scanner Data:

I use the Retail Scanner Database collected by AC Nielsen and made available through the Kilts Center at The University of Chicago Booth School of Business. The retail scanner data consist of information on weekly price and quantity sold generated by point-of-sale systems for more than 100 participating retail chains across all its markets between 2006 and 2016. When a retail chain agrees to share their data, all of their stores enter the database. As a result, the database includes more than 35,000 participating grocery, drug and mass merchandiser stores located in more than 2000 counties. These stores cover more than half the total sales volume of US grocery and drug stores and more than 30 percent of all US mass merchandiser sales volume.

I define a retail chain to be a unique combination of two identifiers in the Nielsen data: code of parent and code of retailer. "Parent code" indicates the company that owns a chain and "retailer code" indicates the chain itself (e.g: Albertson's LLC which owns Safeway). I introduce sample restrictions at the chain level. First, I require that the chains are present in the sample in every year from 2007 to 2011. Second, I require that chains are present in two or more counties. This leaves me with 91 retail chains. Finally, I require chains to be present in the counties for which there is information for all the variables in the analysis, leaving us with 84 retail chains. I also impose restriction on the sample of products. In particular, I focus on products that are frequently sold and available across markets. For each retail chain, I focus on products that are available in 80% or more of its markets in a given half-year.[11]

While the raw data are sampled weekly, I aggregate to construct half-year prices, since this reduces high-frequency noise. I construct the half-year price of an item by dividing its total value of sales by the total quantity of units sold in the half-year. As the main focus of the paper is on a) prices of existing products that are similar within chain across stores, and b) variation of price indices across time, we include an item only if it has positive sales in 2007 and 2011. We track the price of identical items (UPC-store combinations) across time, so that changes in quality or issues with comparing nonidentical products are less relevant for our results. I next describe the construction of the county-level price indices from these individual price observations.

County-level Price Index
The construction of the county-level price index necessarily entails various measurement choices. In the body of the paper I concentrate on a single benchmark price index, but in Appendix D, I show that the main results hold for price indices constructed under various alternative assumptions, including, for example, adjusting the price index for entry and exit of barcodes and stores (Appendix D.2).

I assume that consumer behavior features multi-stage budgeting in two stages. In the first stage,

[11] I also consider different sets of products for additional robustness. Results are robust to not-restricting the sample of products, restrict it to 50% or more markets and only considering products in the top 40 main product modules.

consumers in a county decide which of 1000 product modules to buy from based on the product module price index. In the second stage, conditional on the product module, consumers decide which variety to purchase; where a variety is defined as a store-barcode combination (eg. 12 oz. Coke in 7-eleven). Specifically, I first construct a product-module level price index as in Sato (1976) and Vartia (1976):

$$P_{mct} = \Pi_{u \in I_{mc}} \left(\frac{p_{umct}}{p_{umct-1}} \right)^{w_{umct}},$$

where p_{umct} is the price of variety u, in product module m, sold in county c at year t, and

$$w_{ucmt} = \frac{(s_{umct} - s_{umct-1}) / (ln(s_{umct}) - ln(s_{umct-1}))}{\sum_{v \in I_{mc}} (s_{vmct} - s_{vmct-1}) / (ln(s_{vmct}) - ln(s_{vmct-1}))} \ ; \ s_{umct} = \frac{p_{umct} q_{umct}}{\sum_{v \in I_{mc}} p_{vmct} q_{vmct}},$$

where q denotes quantity. I_{mc} is the set of varieties u in product module m in county c that are consumed in t and $t-1$ (continuing varieties). The weights w are ideal log-change weights and they are county specific to allow for spatial variation in the relative weight of an item.[12]

I then construct the overall county-specific price index by weighting the product module price indices by the revenue share of a particular product module in the initial year (α_{mct-1}) ,[13]

$$P_{ct} = \Pi_m \left(\frac{P_{mct}}{P_{mct-1}} \right)^{\alpha_{mct-1}},$$

where

$$\alpha_{mct-1} = \frac{\sum_{u \in m} Sales_{umct-1}}{\sum_u Sales_{uct-1}}$$

In Figure A2 of the Appendix, I plot the histogram of county-level inflation rate between 2007 and 2011. The population-weighted average inflation is 11.09%, which contrasts well with the inflation in the food at home official CPI from BLS. Figure A2 also shows substantial cross-county dispersion in inflation rates.

2.2 House price data

House price data is obtained from from the Federal Housing Finance agency (FHFA). The agency construct a House Price Index (HPI), which is a broad measure of the movement of single-family house prices. The HPI is a weighted, repeat-sales index, meaning that it measures average price

[12]Note that they are always bounded between the shares of spending in period t and period t-1.

[13]This price index is consistent with the following utility function:

$$U_c = \prod_{m \in I_c} \left[\sum_{u \in I_{mc}} q_{umc}^{\frac{\sigma_m - 1}{\sigma_m}} \right]^{\alpha_{mc} \frac{\sigma_m}{\sigma_m - 1}}$$

where σ_{mc} is the elasticity of substitution between varieties within product module, and α_{mc} is the fraction of expenditures spent on product module m in county c.

changes in repeat sales on the same properties.[14] The HPI serves as a timely, accurate indicator of house price trends at various geographic levels, including county, zip code, state and Metropolitan Statiscal Areas (MSAs) levels.[15] For the main analysis, I focus on long differences between the first semester of 2007 and the first semester of 2011.[16]

The other major data set related to house prices used in the paper is obtained from the 2005 Wharton Regulation Survey. Gyourko et al. (2008a) use the survey to produce a number of indexes that provides information on various aspects of local land use control environments, including the general characteristics of the regulatory process, statutory limits on development, density restrictions, open space requirements, infrastructure cost sharing and approval delay. It is supplemented by information on local ballot initiatives and state involvement in land use controls. Lower values in the Wharton Land Regulation Index (WLRI, henceforth) indicates the adoption of less restrictive policies toward real estate development. In contrast, high values of the WLRI are associated with municipalities that have zoning regulations or project approval practices that constrain new residential real estate development. I process the original municipal-based data to create average regulation indexes at the county level. A limitation of the WLRI is that it is only available for 910 counties, out of the 2300 counties for which I have retail price information from Nielsen Data. These 910 counties represent more than 70% of total sales in the Nielsen Data and cover more than 70% of the U.S. population (See Table A1 in the appendix).

2.3 County-level macroeconomic variables

I complement the main datasets with county-level data from different sources. I use county-level data on wages, employment and number of establishments from the BLS. Data on education levels, age, and population density come from the American Community Survey (ACS).

3 Facts on the geographic distribution of retail chains

I motivate my analysis by documenting two stylized facts about the geographic distribution of retail chains in the United States.

Fact 1. There is substantial geographic dispersion in retail chains' sales

Define the Geographic Dispersion Index (GDI) for each retail chain r:

$$GDI_r = 1 - \sum_c S_{rc}^2,$$

where $S_{rc} \equiv \frac{Sales_{rc}}{\sum_c Sales_{rc}}$ represents the share of retail chain r's national sales that take place in county

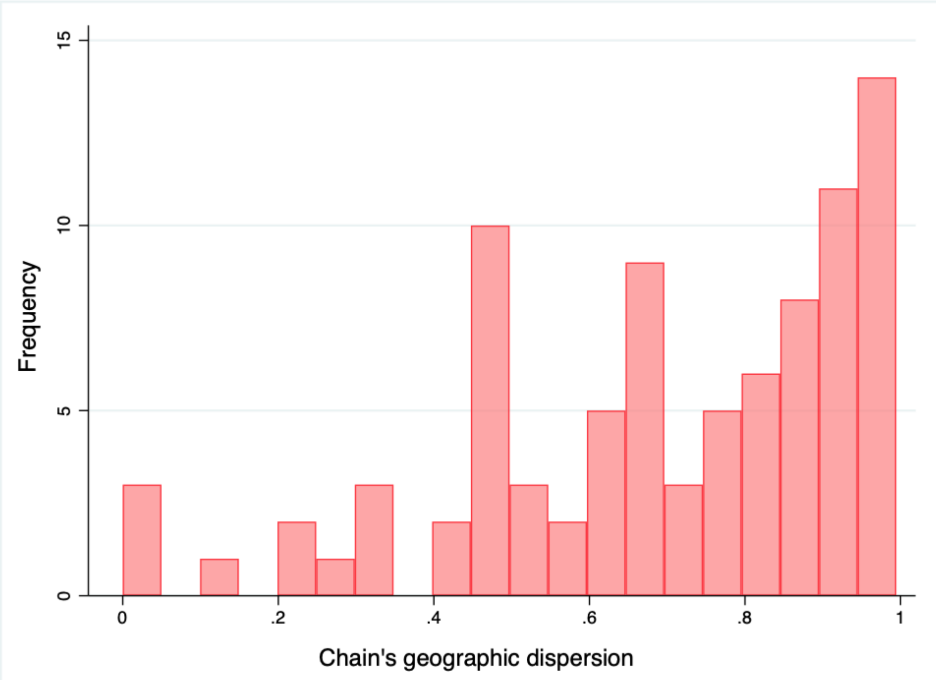
[14]This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975.

[15]These data is highly correlated with data from Zillow. The correlation in house prices changes from 2007 to 2011 between these two alternative datasets is 96%.

[16]This timing convention follows Stroebel and Vavra (2019) to facilitate comparisons with their results. Additionally, the house-price collapse started at the end of 2006/beginning of 2007 and from the second half of 2011, house prices stopped declining (See Figure A1 in the appendix). All results are robust to the alternative timing 2007-2009; following, for example, Mian et al. (2013) and Giroud and Mueller (2019).

c. Low values of the GDI indicate that sales are concentrated in a few locations. Figure 1 plots the distribution of the GDI. We observe substantial geographic dispersion. For example, fewer than 10% of the retail chains in the sample have a GDI below 0.5.

Figure 1: Geographic Dispersion of retail chains



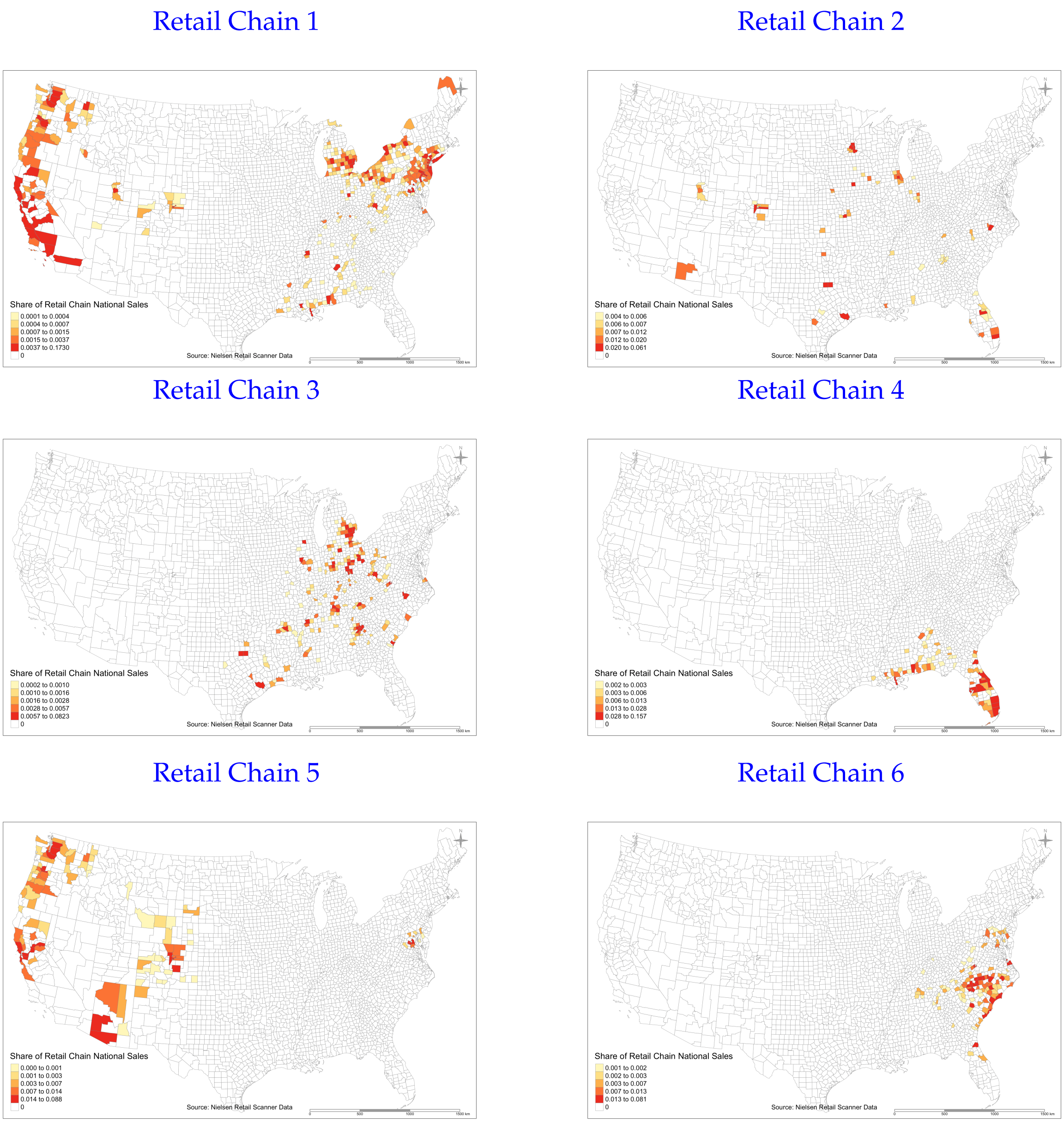
Fact 2. There is substantial heterogeneity in the spatial network of retail chains

Retail chains are distributed differently across the U.S. In Figure 2, I show the spatial network of six anonymous chains in my data.[17] The maps use a red-to-yellow color scale to display the sales of the retail chain in each county, where red counties are those in which the retail chain has the highest sales volume, and yellow are those in which the retail chain has the lowest sales volume. Figure A3 shows that retail chains' networks have very different geographic layouts. For example, there are big and small retail chains, such as Chain 1 and Chain 2, which are spread across several regions in the U.S. and are active in more than 15 states. There are also other chains, such as Chain 3 and Chain 5, that cover an extensive area but are spread across counties and states in those areas. Finally, there are other retail chains that are concentrated in a specific geographic area, such as Chain 4 and Chain 6.

Notice that even though proximity seems to play a role in determining the location of retail chain stores, the spatial networks of retail chains go beyond neighboring counties. That is, firms are not only located in groups of nearby counties. Furthermore, they are widely spread throughout different states. For example, some chains have only one or two stores per state.

[17]Nielsen does not disclose the names of the chains in the data. So, in order to provide a broader view of the geographic distribution of retail chains, in Figure A3 of Appendix C, I present a graph from AggData to illustrate the geographic distribution of some popular retail chains.

Figure 2: Examples of Retail Chains in the Data



Note: Each panel plots the distribution of sales of a retail chain in the data. The maps use a yellow-to-red color scale to display the sales of the retail chain in each county, where red represent counties where the chain has the highest amount of sales, and yellow counties where the chain has the lowest amount of sales.

The main takeaway from these empirical patterns is that retail chains are multi-regional firms that are unevenly distributed in space. This heterogeneity in the spatial networks of retail chains will creates differences in how each pair of counties is connected to each other. For instance, two counties will be very exposed to each other if they are served by the exact same retail chains. In the next section, I characterize these linkages and study their role in propagating local shocks across the geography.

4 Empirical analysis

I begin the analysis by constructing bilateral weights that characterize how connected each pair of counties is, given the geographic distribution of retail chains' stores. Given these weights, I construct the average exposure of each county to house price shocks in other counties and study the sensitivity of county-level retail prices to house price-induced shocks in other counties. I then discuss the identification assumptions and address potential threats to the validity of those assumptions.

4.1 Bilateral linkages and exposure variables

I use Nielsen Scanner data to characterize the bilateral linkages between each pair of counties and the exposure of stores and counties to shocks in other counties.

Define the retail chain's network weights S_{rc} as the share of retail chain r's national sales that takes place in county c in 2007:

$$S_{rc} \equiv \frac{Sales_{rc}}{\sum_c Sales_{rc}} = \frac{\text{Sales of retail chain r in county c}}{\text{National sales of retail chain r}}.$$

These weights measure how important market c is for retail chain r. Intuitively, a store located in county c belonging to a retail chain r will be more exposed to shocks in county k if county k is an important market for the retail chain r. Given these weights, I define the exposure of retail chain r in county c to house price shocks in other counties k as a weighted average of house price changes across counties in the U.S:

$$\text{Store}\Delta \log HP(others)_{rc} \equiv \sum_{k \neq c} S_{rk} \Delta \log(HP_k)^{07-11}, \tag{4.1}$$

where $\Delta \log(HP_k)^{07-11}$ is the percentage change in house prices between 2007 and 2011 in county k.

The network of retail chains create linkages between each pair of counties. These linkages will depend on which retail chains are important in each county. Define l_{rc} as the share of total retail sales in county c that correspond to retail chain r in 2007,

$$l_{rc} \equiv \frac{Sales_{rc}}{\sum_r Sales_{rc}} = \frac{\text{Sales of retail chain r in county c}}{\text{Total sales in county c}}.$$

l_{rc} measures how important the retail r is in county c. With this, I can define a county's exposure to house price changes in other counties as the weighted average exposure of retail chains that operate in the county:

$$\text{County } \Delta \log HP(others)_c \equiv \sum_r l_{rc} \text{Store}\Delta \log HP(others)_{rc}. \tag{4.2}$$

Exchanging the order of summation across r and k, I can re-write the county-level exposure (Equation 4.2) as,

$$\text{County}\Delta\log HP(others)_c \equiv \sum_{k\neq c}\underbrace{\sum_r l_{rc}S_{rk}}_{\omega_{ck}}\Delta\log(HP_k)^{07-11}, \tag{4.3}$$

where

$$\omega_{ck} = \sum_r l_{rc}S_{rk}. \tag{4.4}$$

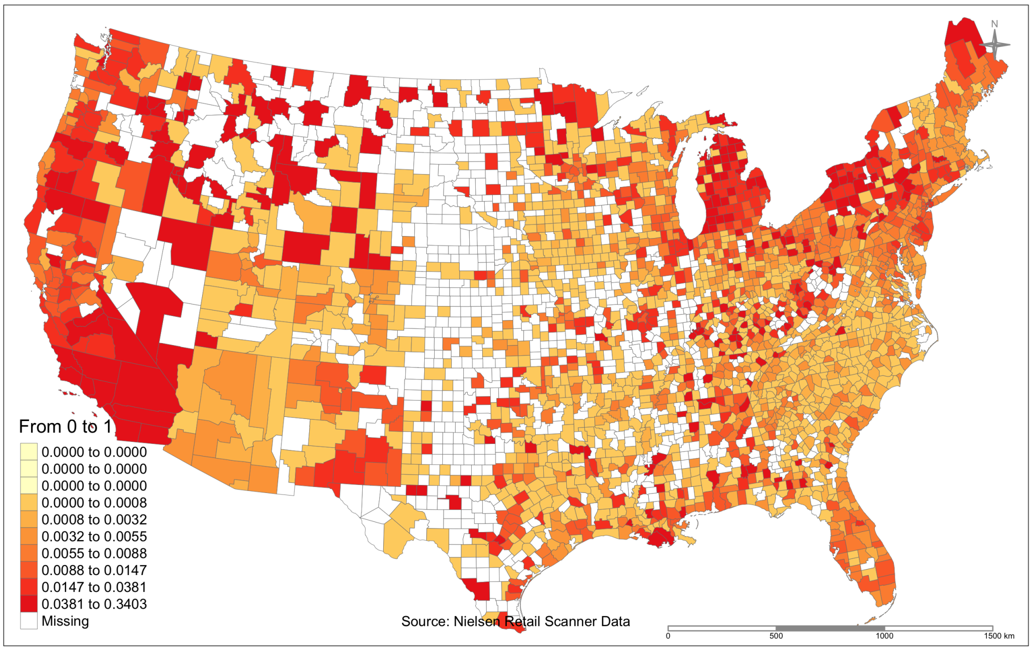
defines the bilateral exposure of county c to county k. Equation 4.3 presents the main variable in my analysis. The exposure of a county to shocks in other counties is the network-weighted average change in house prices in other counties, where the weights are given by the bilateral exposure of county c to each county k (ω_{ck}).

The weights ω_{ck} are at the center of my analysis. Intuitively, a county c will be more exposed to shocks in county k if market k is an important market (S_{rk}) for the dominant chains that operate in c (l_{rc}). The matrix that contains $\omega_{ck}\ \forall\ c,k$ characterizes all the bilateral linkages between each pair of counties in the retail chains' dimension.[18] Note that, unlike bilateral distance, the bilateral linkages created by the network of retail chains are intrinsically asymmetric. As an illustrative example, in Figure 3 I show how counties in the sample are exposed to shocks in Los Angeles County (L.A., from now on). The map plots $\omega_{c,LA}$ for every county c in the sample. I assign colors ranging from light yellow to red, with red counties c being those that are more exposed to shocks in Los Angeles (high $\omega_{c,LA}$). We can observe three patterns. First, not surprisingly, counties located near L.A. tend to be connected to L.A. Second, we also observe strong linkages with counties that are far away. For example, Santa Barbara (located 100 miles away from L.A.) is less exposed to shocks in L.A than some counties located in the states of Michigan or Maine.[19] Third, there is substantial heterogeneity in the exposure to shocks in L.A, even for counties that are next to each other.

[18]A complementary contribution of this paper is constructing the matrix that summarizes these linkages and making it publicly available for other researchers. The matrix can be useful to i) study other topics related to propagation through retail chains' networks, and ii) control for retail chain networks if the aim of the study is analyzing the effect of local shocks on retail prices. Researchers interested in using the matrix can contact the author at eglembergman@berkeley.edu.

[19]In section 4.3, I discuss in detail the role of proximity in explaining my results.

Figure 3: $\omega_{c,LA}$: Exposure of counties c to shocks in Los Angeles



Note: The yellow-to-red color scale represents the degree to which counties are exposed to shocks in L.A., based on $\omega_{c,LA}$ (See Equation 4.4). Red indicates higher $\omega_{c,LA}$. Source: Nielsen Retail Scanner Data.

4.2 Main analysis: OLS and Instrumental Variables

After constructing the exposure variables, I empirically show to what extent house price-induced local demand shocks propagate to distant regions through the network of retail chains.

I start by examining the sensitivity of county-level prices to changes in house prices in other counties that are linked through the network of retail chains.

I estimate the following equation in differences from 2007 to 2011:

$$\Delta \log(P_c)^{07-11} = \beta_0 + \beta_1 \Delta \log(HP_c)^{07-11} + \beta_2 \sum_{k \neq c} \omega_{c,k} \Delta \log(HP_k)^{07-11} + X_c + \Delta\epsilon_c, \tag{4.5}$$

$\Delta \log(P_c)^{07-11}$ denotes the percentage change in county-level retail price index for continuing varieties in county c.[20] $\Delta \log(HP_c)^{07-11}$ denotes the local change in house prices in county c. This variable allows me to control for the direct effect of house price changes on retail prices.[21] My main variable of interest is the county-level exposure to shocks in other counties that are linked by the networks of retail chains: $\sum_{k \neq c} \omega_{c,k} \Delta \log(HP_k)^{07-11}$, as defined in Equation 4.3. X_c is a vector of time-varying controls at the county-level. It includes changes in local wages, changes in employ-

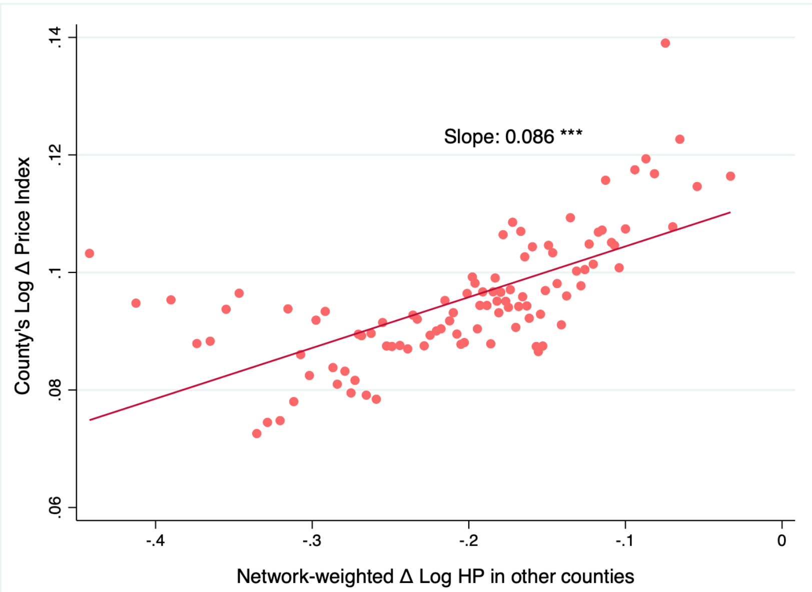
[20] The price index considers only varieties that existed both in 2007 and 2011. Varieties are defined as a combination of store-barcodes. I describe in detail the construction of the price index in Appendix B. In Appendix D.2, I explore the effects on the extensive margin (e.g. retail chains closing stores or discontinuing products). I do not find any effect of the network of retail chains on the extensive margin.

[21] The coefficient β_1 is also useful to compare with previous papers' estimates of the effect house price-induced local shocks on local prices (e.g. Stroebel and Vavra (2019)).

ment, and changes in number of retail establishments.[22] Standard errors are clustered at the state level.[23]

Figure 4 provides a visual impression of the OLS relationship between changes in the county-level retail price index and changes in house prices in other counties, linked through the spatial networks of the retail chains. The pattern is clear. Controlling for the local change in house prices, the elasticity of county-level prices with respect to changes in house prices in other counties is around 9%. In Columns (1) to (3) of table 1, I report the OLS estimates in table format. The first row reports the elasticity of county-level retail prices with respect to local house prices. The second row reports the elasticity of county-level prices with respect to house price changes in other counties that are linked by the network of retail chains. In Column (1), I report the direct effect of house price changes on county-level prices. Similar to previous papers, I find that there is a positive relationship between house price changes and local retail prices. In Column (2), I add network-weighted house price changes in other counties. Conditional on the local changes in house prices, a drop of 10% in network-weighted house prices in other counties is associated with a 0.88% reduction in county-level prices. This result remains almost unchanged in Column (3), after including county-level controls such as changes in employment, wages and number of establishments.

Figure 4: Relationship between network-weighted changes in house prices and county-level change in prices



Note: The binscatter plots the OLS relationship between the network weighted house price changes in other counties ($\sum_{k \neq c} \omega_{ck} \Delta \log(HP_k)$) and county-level retail prices ($\Delta \log(P_c)$). Counties are sorted into percentile bins based on their value on $\Delta \log(P_c)$. The sample includes 910 counties, so each percentile bin therefore represents 9 counties. To filter out any confounding effects, I control for the local change in house prices ($\Delta \log(HP_c)$). Hence, the dot indicates the average value of (residual) $\Delta \log(P_c)$.

Although the OLS estimates constitute suggestive evidence of the propagation of shocks across the network of retail chains, both the local change in house prices and the exposure to house price

[22] Results are robust to other combinations of controls at the county-level. See A3 of Appendix D.1.

[23] As a robustness check, in Table A11 of Appendix D.1, I compute standard errors allowing for arbitrary cross-regional correlation in the regression residuals as suggested by Adao et al. (2019).

changes in other counties can suffer from endogeneity. For instance, contemporaneous negative productivity shocks could lead to an increase in retail prices and, at the same time, a decrease in house prices. This would imply a downward bias in our estimate for the direct effect. More importantly, it is hard to argue that the location of retail chains' stores is random. For instance, to reduce transportation costs, retail chains' might be clustered geographically. If this were the case, regional productivity shocks that affect regional house prices could generate co-movement in prices between counties that are connected by the retail chains.

My first strategy to deal with potential endogeneity issues is to use instruments for the two main variables. Regarding local house prices changes, I follow an extensive literature that exploits across county variation in housing supply elasticity (e.g. Mian and Sufi (2011), Adelino et al. (2015), Stroebel and Vavra (2019), among others). The idea is that in response to a national level negative housing demand shock, areas with lower housing supply elasticity (e.g. tighter land regulations, WLRI) will experience a larger drop in house prices. In particular, in my baseline specification, I use the county-level Wharton Land Regulation Index (WLRI) from Gyourko et al. (2008b) and its interaction with the spatial network of retail chains: [24]

$$WLRI_c \to \Delta \log(HP)_c$$

More importantly, I combine the WLRI with data on the location and sales of the retail chain stores to construct an instrument for retail chains' exposure to house price shocks in other counties. I instrument network-weighted changes in house prices with network-weighted change in WLRI.

$$\sum_{k \neq c} \omega_{ck} WLRI_k \to \sum_{k \neq c} \omega_{ck} \Delta \log(HP)_k$$

Intuitively, a county will be more exposed to house prices drops in other counties if the dominant chains in the county have stores concentrated in counties with tighter land regulations. While house price changes in counties linked by the retail chains might correlate with regional productivity changes that generate co-movement in prices, the WLRI isolates variability in house prices that is not correlated with those productivity changes.[25] Hence, my main identifying assumption is that in the absence of linkages between counties through the retail chains' network of stores, changes in county-level retail prices would be uncorrelated with WLRI in regions that are linked by the network of retail chains.[26] It is important to note that in my shift-share design, identification

[24]Gyourko et al. (2008b) conduct a nationwide survey to construct a measure of local regulatory environments pertaining to land use or housing. The idea is that in areas with a tighter regulatory environment, is harder to expand (contract) the housing supply in response to demand shocks. In a complementary analysis, I repeat the estimations using the Saiz (2010) Housing supply elasticity (Saiz HSE). The Saiz HSE uses geographic information of the metropolitan area to measure how easy is constructing new houses (e.g: areas with a flat topology are assigned with a higher elasticity). The coefficient across different IV strategies is remarkably similar and I cannot reject the hypothesis that they are equal (See Table A4 of Appendix D.1).

[25]Note that these assumptions are milder than those in studies that use local housing supply elasticity to instrument house price changes. Those studies rely on the assumption that housing supply elasticity only affects the outcomes through its impact on house prices. That assumption is a sufficient condition (though not necessary) for exogeneity of the network-weighted WLRI.

[26]Note that this assumption is milder than the OLS assumptions; namely that in the absence of linkages between counties through the retail chains' network of stores, changes in county-level retail prices are uncorrelated with changes in house prices in regions that are linked by the network of retail chains.

relies on exogeneous variability in the shocks (WLRI, in this case). As shown by Borusyak et al. (2018), if the shocks are as good as random, then it is possible to identify the causal effect of interest even when, as in most applications, the exposure shares are not random.[27]

There are important remaining challenges to my identification assumptions. In particular, even when this instrument has been widely used in the literature as an exogeneous shock to house prices, it might not be as good as random. If it is not and, for example, retail chains locate their stores in neighboring counties, then unobserved regional shocks that correlate with housing supply elasticity could generate co-movement in prices, even in the absence of retail chains' networks. In this section, I present the main empirical results of the paper and I address this and other concerns in Section 4.3.

I begin by providing a visual impression of the first stage and the reduced-form coefficient of the indirect effect. In Panel a of Figure 5, I plot the relationship between the network-weighted WLRI and the network-weighted change in house prices (first-stage). As expected, the network-weighted WLRI is negatively correlated with the network-weighted change in house prices. That is, counties exposed to counties with tighter land regulations are counties exposed to counties that experienced higher drops in house prices. In Panel b of Figure 5, I plot the relationship between network-weighted WLRI and local retail prices (reduced-form). The pattern is clear: counties more exposed to a higher network-weighted WLRI in other counties experienced a higher drop in local retail prices.

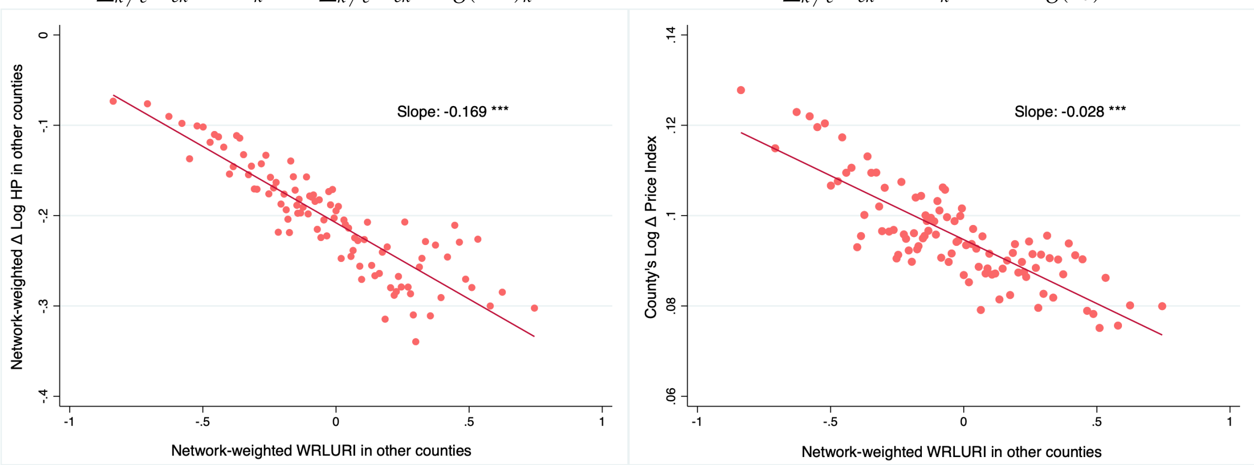
Figure 5: Propagation of house price-induced local shocks: first stage and reduced-form

Panel a: First Stage

$\sum_{k \neq c} \omega_{ck} WLRI_k \longrightarrow \sum_{k \neq c} \omega_{ck} \Delta \log(HP)_k$

Panel b: Reduced Form

$\sum_{k \neq c} \omega_{ck} WLRI_k \longrightarrow \Delta \log(P_c)$



Note: The binscatter plots the first stage and reduced-form relationship. Panel A plots the first stage: the relationship between the network-weighted WLRI (x-axis) and the network weighted house price changes (y-axis). Panel B reports the reduced-form coefficient: the relationship between network-weighted WLRI (x-axis) and county-level change in consumer prices (y-axis). Counties are sorted into percentile bins based on their value on network-weighted change in house prices (Panel A) and county-level changes in consumer prices (Panel B), respectively. My sample includes 910 counties, so each percentile bin therefore represents 9 counties. To filter out any confounding effects, I control in both cases for the local change in house prices (so the dot indicates the average value of the (residual) outcome variable.

[27] In a recent paper, Goldsmith-Pinkham et al. (2020) provides conditions under which the shift-share instrument is exogeneous when the shares are exogeneous. However, in most of the interesting economic settings (including the one studies here), the shares are not necessarily random. Borusyak et al. (2018) shows that as long as the shocks are as good as random, then identification is granted, even when the exposure shares are endogeneous. In section 4.3 I perform a series of robustness exercises and discuss extensively the plausibility of my identification assumptions.

After inspecting the data visually, I report the main findings of the paper in Table 1. Columns (4) to (6) present the IV estimates. In Column (4), I report the local elasticity of county-level retail prices with respect to house prices. The elasticity is 0.125. This estimate is in line with findings by Stroebel and Vavra (2019) that range between 0.15 and 0.25. In Column (5), I add the network-weighted change in house prices in other counties. First, note that once I control for the propagation from other counties, the direct effect declines 40%. This suggests that previous papers that did not account for the network of retail chains overestimated the elasticity of retail prices with respect to house prices. More importantly, in the second row, I report the elasticity of local retail prices with respect to shocks in other counties that are linked by the retail chain networks. Conditional on the local changes in house prices, a 10% drop in house prices in other counties linked by the network of retail chains leads, on average, to a decrease of 1.35% in county-level consumer price index. Reassuringly, the coefficient remains stable and almost unchanged after including county level controls in Column (6). To provide a benchmark of the quantitative relevance of these results, a county exposed to an average drop in house prices in other counties of 26% (25th percentile) has a predicted inflation of around 9.5%. In contrast, a county exposed to a 14% (75th percentile) drop in house prices in other counties has a predicted inflation of 11.5%.[28]

Table 1: Propagation of house price-induced local shocks through the network of retailer chains

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.056***	0.031**	0.029*	0.125***	0.074**	0.076**
	(0.014)	(0.013)	(0.015)	(0.027)	(0.034)	(0.037)
Chain-linked Δ Log HP (other counties)		0.088***	0.086***		0.129***	0.135***
		(0.026)	(0.026)		(0.039)	(0.040)
Panel B: First Stage						
F-stat				21.358	18.642	14.756
County controls	no	no	yes	no	no	yes
Observations	910	910	910	910	910	910
R-squared	0.154	0.257	0.264	0.084	0.071	0.060

A unit of observation is a county. The dependent variable is the percentage change in the county price index between 2007 and 2011 ($\Delta \log(P_c)$). County's $\Delta \log HP$ is the percentage change in house prices between 2007 and 2011. Chain-linked Δ Log HP (other counties) denotes the network-weighted percentage change in house prices in other counties ($\sum_{k\neq c} \omega_{ck} \Delta \log(HP_k)$), as defined in Equation 4.3. Columns (1) to (3) report results for OLS estimations. Columns (4) to (6) report results, after instrumenting the two main endogeneous variables. I instrument local percentage change in house prices with local WLRI. I instrument network-weighted percentage change in house prices ($\sum_{k\neq c} \omega_{ck} \Delta \log(HP_k)$) with network-weighted WLRI ($\sum_{k\neq c} \omega_{ck} WLRI_k$). County-level controls in columns (3) and (6) include changes in log wages, changes in log number of retail establishments and changes in log employment. Panel A reports the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Retail Scanner Data. Housing price data is obtained from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a). County-level macroeconomic variables are obtained from the Bureau of Labor Statistics.

[28] These predictions are calculated including the constant and assigning the median change in local house prices in both cases.

In Appendix D.1, I explore the sensitivity of my results to alternative specifications. First, I show that the main coefficient remains stable after adding different county-level controls (Table A3). Second, I repeat the main analysis, but now instrumenting the endogeneous variables with the Saiz (2010) geography-based housing supply elasticity (Table A4). The main coefficients remains almost unchanged under the alternative instruments. Third, I show that my results are not sensitive to: a) constructing price indices under different assumptions (Table A6), b) replicating the analysis for the period 2007-2009 (Table A7), c) including different combinations of regional fixed effects (Table A8), d) considering three digits zip code as the relevant market, and e) excluding California (Table A10). Fourth, in Table A11, I follow Adao et al. (2019) methodology to conduct valid inference for shift-share designs under arbitrary cross-regional correlation in the regression residuals. Although the standard errors increase slightly, the coefficients remain statistically significant at the usual levels.[29]

In the next section, I discuss the main challenges to my identification assumptions and address each of these challenges.

4.3 Validity of identification assumptions

The key identification assumption is that, in the absence of the retail chains' network of stores, changes in county-level prices are uncorrelated with housing supply elasticity in the regions that are linked by the retail chains networks. In regards to this assumption, the biggest challenge is to separate propagation of shocks through retail chains' networks from common shocks in the regions in which the retail chains operate. For example, to minimize transportation costs, retail chains might locate their stores in neighboring counties where wages and prices co-move because of other reasons, such as integrated labor markets or trade linkages. In this section, I address these concerns.

First, I document that my results hold after controlling for trade relationships due to proximity between counties. In fact, once the retail chain's channel of propagation of shocks is taken into account, the effect of proximity disappears. Second, I show that the network of retail chains does not affect other economic outcomes in distant locations, which suggests the absence of other factors creating co-movements. Third, I filter out any remaining concerns about common regional shocks by turning to more granular data at the retail chain-by-county level, which allows me to include county by time fixed effects. These fixed effects absorb any common variation within a county that is due to a regional shock, regardless of whether the shock is region-specific or correlated with shocks in other regions.

4.3.1 Retail chains' network of stores or proximity channel?

Thus far, I have abstracted from the role of geographic proximity in the propagation of shocks. However, it is important to disentangle whether my results are explained specifically by the network of retail chains or by trade relationships due to proximity that correlate with the network of retail chains.

[29] Even though in the main analysis, I cluster standard errors at the state-level, there could be errors correlated across regions independent of geographic location.

I begin by showing that even shocks in out-of-state counties are propagated through the network of retail chains. I construct the network-weighted change in house prices, excluding changes in house prices in counties that are in the same state. I then estimate the leave-state-out version of my main specification (Equation 4.5):

$$\Delta \log(P_c)^{07-11} = \beta_1 \Delta \log(HP_c)^{07-11} + \beta_2 \sum_{k \neq State_c} \omega_{c,k} \Delta \log(HP_k)^{07-11} + \epsilon_c, \tag{4.6}$$

where $State_c$ denotes the state in which county c is located. Similar to my main analysis, I construct a leave-state-out instrument: $\sum_{k \neq State_c} \omega_{c,k} WLRI_k$.

Results are reported in Table 2. Across a variety of specifications, I find that local prices are sensitive to shocks in distant counties (out-of-state) that are linked by the network of retail chains. In addition, I cannot reject the hypothesis that the coefficient for out-of-state shocks is equal to the coefficient of my main specification. This is suggestive evidence that shocks emanating from out of state are as important as shocks from nearby regions.

Table 2: Propagation of house price-induced local shocks to out-of-state counties

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.056***	0.057***	0.056***	0.125***	0.126***	0.121***
	(0.014)	(0.013)	(0.016)	(0.027)	(0.024)	(0.028)
Chain-linked Δ Log HP (other counties)		0.078***	0.078***		0.121***	0.129***
(excl state)		(0.023)	(0.023)		(0.032)	(0.032)
Panel B: First Stage						
F-statistic				21.358	11.032	9.712
County controls	no	no	yes	no	no	yes
Observations	910	910	910	910	910	910

A unit of observation is a county. The dependent variable is the percentage change in the county price index between 2007 and 2011 ($\Delta \log(P_c)$). County's Δ Log HP is the percentage change in house prices between 2007 and 2011. Chain-linked Δ Log HP (out-of-state counties) is the network-weighted percentage change in house prices in other counties, excluding the state in which the county is located ($\sum_{k \neq State} \omega_{ck} \Delta \log(HP_k)$) as defined in Equation 4.6. Columns (1) to (3) report results for OLS estimations. County-level controls in columns (3) and (6) include changes in log wages, changes in log number of retail establishments and changes in log employment. Columns (4) to (6) report results, after instrumenting the two main endogeneous variables. I instrument local percentage change in house prices with local WLRI. I instrument network-weighted percentage change in house prices ($\sum_{k \neq State} \omega_{ck} \Delta \log(HP_k)$) in out-of-state counties with network-weighted WLRI in out-of-state counties ($\sum_{k \neq State} \omega_{ck} WLRI_k$). Panel A report the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Retail Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a). County-level macroeconomic variables are obtained from the Bureau of Labor Statistics.

After showing that the shocks propagate to distant regions through the network of retail chains, I now directly study the question of to what extent propagation of local shocks is explained by the retail chains that serve the county and to what extent it is explained by trade relationships due to proximity. Comparing these two channels has two purposes. First, it allows me to check that my results are not explained by other networks that happen to be correlated with the network of retail chains. This is specially important, given that the correlation between county-level linkages by the network of retail chains and county-level linkages due to proximity is 17%. Second, exploring the role of proximity in shaping the propagation of shocks is interesting in its own right, since it allows me to compare my new channel of propagation of shocks with the more traditional channel of propagation of shocks through trade relationships between nearby regions.

Define,

$$\text{prox-weighted}\Delta\log(HP)_c^{others} = \sum_{k\neq c} \delta_{ck}^{prox} \Delta\log(HP)_k^{07-11},$$

where

$$\delta_{ck}^{prox} = \frac{d_{ck}^{-\theta} Population_{k,2007}}{\sum_k d_{ck}^{-\theta} Population_{k,2007}}$$

where d_{ck} is the distance in miles between county c and k.[30] I assume $\theta = 1$ as it is standard in the trade literature. Intuitively, the closer county c and county k are to each other and the bigger county k is, the more a shock in county k will affect county c. [31]

Then, I add this variable to my main Equation 4.5 and estimate the following equation:

$$\Delta\log(P_c)^{07-11} = \beta_1 \Delta\log(HP_c)^{07-11} + \beta_2 \sum_{k\neq c} \omega_{c,k} \Delta\log(HP_k)^{07-11} + \beta_3 \sum_{k\neq c} \delta_{ck}^{prox} \Delta\log(HP)_k^{07-11} + \epsilon_c, \tag{4.7}$$

In addition to the instruments discussed in previous sections, I instrument Prox-weighted $\Delta\log(HP)_c^{others}$ with prox-weighted $WLRI$. Formally,

$$\text{Prox-weighted } WLRI_c = \sum_{k\neq c} \delta_{ck}^{prox} WLRI,$$

Table 3 reports results for different versions of equation 4.7. Columns (1) to (3) present OLS estimates. Columns (4) to (6) report IV estimates. In all cases, the instruments have predictive power. The first row reports the coefficient associated with the proximity-weighted shocks. The second

[30] The matrix of distances between each pair of counties is obtained from the County Distance Database, provided by the National Bureau of Economic Research and compiled by Jean Roth (2014)

[31] Note that a first order approximation to market access of a county in standard trade models (e.g. Donaldson and Hornbeck (2016)) is given by:

$$MA_c = \sum_k \tau_{ck}^{-\theta} Population_k.$$

Hence, the weights serve as a proxy for the relative importance of county k in county c's market access. Alternatively, in Appendix D.1.10, I construct the weights based on trade flows between county c and county k at the state-level.

row reports the coefficient associated with the local effect of the shock. The third row reports the coefficient associated with the retail chains networks. In Column (1), I show the role of propagation through trade relationships due to geographic proximity. A 10% drop in house prices in nearby regions is associated with a 1.25% lower retail price index in the county. More importantly, in Column (3), I add the retail chains' channel of propagation of shocks. The results are striking. Once I take into account the propagation of shocks through the network of retail chains, the coefficient related to geographic proximity is no longer significant. In contrast, propagation of shocks through retail chains' network of stores remains positive and significant. Columns (4) to (6) repeat the analysis, but instrumenting each of the variables. I arrive at the same conclusions. Specifically, I conclude that, controlling for proximity-weighted changes in house prices and local changes in house prices, the elasticity of local retail prices with respect to shocks in distant counties -linked by the same retail chains- is 0.16.

These results indicate that county-level retail prices depend more on shocks to the retail chains that happen to serve the county than on shocks to neighboring counties.

Table 3: Channels of propagation of shocks: retail chains versus geographic proximity

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instruments		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
Prox-weighted Δ Log HP (other counties)	0.125***	0.085***	0.020	0.212***	0.072	-0.139
	(0.009)	(0.015)	(0.016)	(0.050)	(0.097)	(0.092)
County's Δ Log HP		0.024***	0.025***		0.153***	0.099*
		(0.007)	(0.007)		(0.053)	(0.054)
Chain-linked Δ Log HP (other counties)			0.083***			0.166***
			(0.009)			(0.035)
Panel B: First Stage						
F-statistic				12.122	26.215	12.726
County controls	yes	yes	yes	yes	yes	yes
Observations	910	910	910	910	910	910
R-squared	0.172	0.182	0.250	0.088	0.123	0.074

A unit of observation is a county. The dependent variable is the percentage change in the county price index between 2007 and 2011 ($\Delta \log(P_c)$). County's Δ Log HP is the percentage change in house prices between 2007 and 2011. Chain-linked Δ Log HP (other counties) is the network-weighted percentage change in house prices in other counties ($\sum_{k\neq c} \omega_{ck} \Delta \log(HP_k)$) as defined in Equation 4.3. Trade-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, where the weights are given by proximity between counties: $\sum_{k\neq c} \delta^{prox}_{c,k} \Delta \log(HP_k)^{07-11}$. Columns (1) to (3) report results for OLS estimations. Columns (4) to (6) report results, after instrumenting the three main endogeneous variables. I instrument local percentage change in house prices with local WLRI. I instrument Chain-Linked Δ Log HP (other counties) with Chain-Linked WLRI in other counties ($\sum_{k\neq c} \omega_{ck} WLRI_k$). I instrument Trade-Linked Δ Log HP (other counties) with Trade-Linked WLRI in other counties ($\sum_{k\neq c} \delta^{prox}_{ck} WLRI_k$). Panel A report the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Retail Scanner Data. Housing price data is obtained from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a).

4.3.2 The effect of retail chains' network of stores on other economic outcomes

If my results were explained by common shocks to regions in which the retail chains operate, then we would expect co-movement not only in retail prices, but also in other economic outcomes. For example, if results were explained by trade relationships between counties in which retail chains operate, we would expect house price movements to affect wages in distant counties that are linked through these networks. In contrast, it is less clear why the network of retail chains would affect wages in distant counties. I test this by estimating the following version of Equation 4.5:

$$\Delta \log(Y_c)^{07-11} = \beta_1 \Delta \log(HP_c)^{07-11} + \beta_2 \sum_{k \neq c} \omega_{c,k} \Delta \log(HP_k)^{07-11} + X_c + \Delta \epsilon_c, \tag{4.8}$$

where Y_c are the three alternative outcomes: county-level employment, number of retail establishments and wages, respectively. Results are reported in Table 4. Column (2) reports the effect on employment. Although house price-induced local shocks affect local economic outcomes, these shocks do not affect labor market outcomes in distant counties. This is consistent with the network of retail chains being the driver of propagation of shocks.

Note that this results has another interesting implication. As retail chains affect the propagation of shocks to prices in distant regions, but not to wages, the chains' spatial networks of stores have consequences for real income.

Table 4: Propagation of house price-induced local shocks to other economic outcomes

	County's Δ Log Price Index	County's Δ Log L	County's Δ Log # of Establishments	County's Δ Log Wages
	(1)	(2)	(3)	(4)
Panel A: Second Stage				
County's Δ Log HP	0.074**	0.155	0.126*	0.309*
	(0.034)	(0.103)	(0.076)	(0.162)
Chain-linked Δ Log HP (others)	0.129***	-0.029	0.070	-0.114
	(0.039)	(0.111)	(0.058)	(0.170)
Panel B: First Stage				
F-statistics	18.642	18.642	18.642	18.642
Observations	910	910	910	910
R-squared	0.071	0.077	0.109	0.094

A unit of observation is a county. Each column has a different dependent variable: Column (1) shows $\Delta \log(P_c)$. Column (2) shows the percentage change in the county's employment rate between 2007 and 2011 ($\Delta \log(L_c)$). Column (3) shows the percentage change in the county's number of retail establishments. Column (4) shows the percentage change in wages. County's Δ Log HP is the percentage change in house prices between 2007 and 2011. Chain-linked Δ Log HP (other counties) is the network-weighted percentage change in house prices in other counties ($\sum_{k \neq c} \omega_{ck} \Delta \log(HP_k)$), as defined in Equation 4.3. All columns report IV estimates. I instrument local percentage change in house prices with local WLRI. I instrument network-weighted percentage change in house prices ($\sum_{k \neq c} \omega_{ck} \Delta log(HP_k)$) with network-weighted WLRI ($\sum_{k \neq c} \omega_{ck} WLRI_k$). Panel A report the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Retail Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a). County-level macroeconomic variables are obtained from the BLS.

4.3.3 Retail by county level analysis

In order to address remaining concerns about separating the effect of the network of retail chains from common regional shocks, I turn to more granular data at the county-by-chain level. This allows me to include county-level fixed effects to absorb any common variation within a county that is due to a regional shock, regardless of whether the shock is region-specific or correlated with shocks in other regions. Note that, for example, spillovers from one county to another because of trade relationships due to proximity are county specific and will be accounted for by these fixed effects.

Formally, I estimate the following equation:

$$\Delta \log(P_{rc})^{07-11} = \beta_2 \sum_{k \neq c} S_{r,k} \Delta \log(HP_k)^{07-11} + \gamma_c + X_r^{2007} + \epsilon_{rc}, \tag{4.9}$$

where, as defined in equation 4.1, the weights S_{rk} are the share of the retail chain r's national sales that take place in county k. The dependent variable is the percentage change in the price index for continuing varieties of retail chain r in county c.[32] I aggregate stores of a retail chain in a given county to the retail chain by county level. Therefore, $\Delta \log(P_{rc})^{07-11}$ denotes the percentage change in price index for retail chain r in county c from 2007 to 2011.[33] Analogously to the county-level analysis, I also instrument the local changes in house prices and the store-level network-weighted changes in house prices with WLRI and store-level network-weighted WLRI, respectively. I cluster standard errors at both the retail chain and the state level.

Table 5 report results for the OLS (Columns 1 to 3) and the IV estimations (Columns 4 to 6). Focus first in the OLS estimates. In Column (1) I report results without including county fixed effects. In Column (2), I include county fixed effects. Hence, I compare stores operating in the same county that are exposed to the same regional shocks, but that are exposed differently to shocks in other locations because they belong to different retail chains. We can observe that elasticity of retail chain prices with respect to house prices in other regions is similar to the corresponding elasticity without county fixed effects. In Column (3), I add controls at the retail chain level to account for differential trends for different retail chains. Reassuringly, the elasticity of store-level prices with respect to house prices in other locations remains stable. We observe a similar pattern when analyzing the IV estimates. In my preferred specification (Column 6), I include county fixed effects, as well as the initial amount of national sales of the retail chain as a control. I conclude that the elasticity of store-level prices with respect to house price changes in other counties where the same retail chain operates is 0.18. Note that this result is also consistent with Handbury and Moshary (2020)'s finding that retail chains highly exposed to a national school lunch program reduced prices across all outlets.

[32] In appendix B, I explain in detail the assumptions to construct the price index.

[33] Results are similar if I take each store of the retail chain in a county separately. I present the retail chain-by-county level estimates to avoid repeating similar observations and be conservative with the standard errors.

Table 5: Propagation of local shocks: within county analysis

	Dep Variable: Chain's Δ Log Price in county c					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.029***			0.070***		
	(0.009)			(0.023)		
Store $\Delta \log HP(others)_{rc}$	0.082***	0.080***	0.094***	0.136***	0.185***	0.183***
	(0.026)	(0.027)	(0.032)	(0.046)	(0.043)	(0.037)
Panel B: First Stage						
F-statistic				11.184	19.351	19.932
County controls	yes	-	-	yes	-	-
Retail chain controls	no	no	yes	no	no	yes
County FE	no	yes	yes	no	yes	yes
Observations	3,747	3,747	3,747	3,747	3,747	3,747
R-squared	0.165	0.601	0.612	0.047	0.495	0.497
# retailers	84	84	84	84	84	84
# counties	910	910	910	910	910	910

A unit of observation is a retail chain by county. The dependent variable is the percentage change in the retail chain by county price index between 2007 and 2011 ($\Delta \log(P_{rc})$). County's Δ Log HP is the county-level percentage change in house prices between 2007 and 2011. Store $\Delta \log HP(others)_{rc}$ is the store-level network-weighted percentage change in house prices in other counties ($\sum_{k \neq c} S_{rk} \Delta \log(HP_k)$), as defined in Equation 4.1. County-level controls include log change in wages, employment and number of retail establishment. Retail chain-level controls include log national sales at 2007. In Columns (4) to (6), I instrument the local percentage change in house prices with the local WLRI and I instrument store-level network-weighted percentage change in house prices in other counties ($\sum_{k \neq c} S_{rk} \Delta \log(HP_k)$) with store-level network-weighted WLRI in other counties ($\sum_{k \neq c} S_{rk} WLRI_k$). Panel A report the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at both the retail-chain level and state-level in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Retail Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a). County-level macroeconomic variables are obtained from the Bureau of Labor Statistics.

In appendix D.1, I explore the sensitivity of my results to alternative specifications, and test for alternative stories. In appendix D.1.3 I control for similarity-weighted house price changes in other counties, based on income, population, household debt, and education. The main coefficient remains stable (between 0.129 and 0.137). In Appendix D.1.8, I explore whether my results could be explained by common regional shocks that affect differently retail chains that cater to different income groups. I show that my main conclusions remain unchanged after comparing price changes of stores in a given county that cater to individuals of similar income group.

Until here, I have shown that retail chains play a key role in shaping the propagation of shocks through space. The network of retail chains connect counties economically, beyond the linkages that arise from, for example, proximity in space. I now turn to a model of retail chains' pricing decisions to rationalize this findings and investigate the quantitative implications of this new channel of propagation of shocks.

5 A model of retail chains' pricing decisions

I build and estimate a simple model of retail chains' pricing decisions to quantitatively interpret my findings, test for uniform pricing strategies and evaluate the role of retail chains in distributing economic shocks across the economy. In this section, I introduce the main ingredients of the model and in the next section I make further assumptions to take the model to the data and conduct quantitative analysis.

5.1 Demand

Consider a country that has a finite number of markets, $c = 1, .., C$. The market definition I use for my benchmark case is the county, so retail chains only compete for consumers within a county.[34]

Define Ω_c as the set of active retail chains in county c, p_{rc} as the price of retail r in county c, and P_c as the price index in county c. Assuming direct separability, the demand for retail chain r in county c when prices are $\mathbf{p} \equiv \{p_{rc}\}_{r \in \Omega_c}$ is given by:

$$q_{rc} = D(p_{rc}/P_c),$$

where $D(x) \in C^2(x)$ is twice differentiable with $D'(x) < 0$ and partial price elasticity given by $\sigma_c = \frac{\partial q_{rc}}{\partial p_{rc}} \frac{p_{rc}}{q_{rc}}$. I allow for variable markups by letting σ_c vary with demand conditions in county c.

I now turn to the price setting of retail chains. I consider two possible pricing strategies: uniform pricing and flexible pricing. First, I briefly discuss the standard case of flexible pricing in order to contrast the conclusions under uniform pricing. Then, I solve the model for uniform pricing: a case in which firms are constrained to set the same price in all their markets.

5.2 Retail chains pricing decisions

I assume that retail chains engage in monopolistic competition. In what follows, it is useful to define Ω_r as the set of counties in which chain r is active and Ω_c as the subset of retail chains active in county c. Since in the empirical analysis I do not observe adjustments on the extensive margin, I assume that there is no entry and exit of stores (for a discussion on the extensive margin, see Appendix D.2). I assume that retail chains' costs are national and do not depend on local conditions: $c_{rc} = c_r$. This is consistent with the fact that more than 80% of retail chains' marginal costs are related to wholesale costs of tradable goods, which are generally not produced locally (Stroebel and Vavra (2019)). In addition, it allows me to focus on the demand channel of the shocks; which has been emphasized by previous papers that study the collapse in house prices during the Great Recession.[35]

[34] My results will be robust to using a more aggregate measure of market (the MSA) or the truncated zip code (3-digit zip code) as the relevant market definitions instead of the county.

[35] Using PromoData from Nielsen, Stroebel and Vavra (2019) show that most of marginal cost of retail chains are wholesale costs that do not vary significantly across markets. In addition, they show that these costs do not react to local house price changes.

In appendix E.2, I solve the model allowing for local costs.

5.2.1 Flexible pricing

In the flexible pricing case, firms discriminate prices across markets, so each store operates as an independent business unit. In particular, the maximization problem is given by,

$$\max_{p_{rc}} \sum_{c\in\Omega_r} \pi_{rc} = \sum_{c\in\Omega_r} (p_{rc} - c_r)q_{rc}.$$

This implies,

$$p_{rc}^{flex} = \frac{\sigma_c}{\sigma_{c-1}} c_r \quad \forall \quad c \in \Omega_r.$$

Note that under flexible pricing, the price in county c is independent from demand shocks in other markets $k \neq c$. For simplicity, I aggregate prices at the county-level with a Laspeyers Geometric weighted average price index, where the weights are given by l_{rc}. Then, the county-level price index for flexible prices is given by:

$$P_{ct}^{Flex} = \prod_{rct\in\Omega_c} \left(\frac{\sigma_c}{\sigma_c - 1} c_r \right)^{l_{rc}}$$

There are two main takeaways from flexible pricing: a) demand shocks in county k, do not affect price index of county c; and b) the spatial network of the dominant chains in the county ($S_r = \{S_{rc} \; for \; c = 1, ..., c, ..., C\}$) does not affect the price index of the county.

5.2.2 Uniform Pricing

Now consider the case of uniform pricing. For simplicity, I take uniform pricing as a constraint that retail chains have. This could be rationalized in a menu-cost model in which retail chains need to pay a fixed cost of adjustment to set different prices across regions. In this case, retail chains set a uniform price across their markets. In particular, the retail chain solves:

$$\max_{\bar{p}_r} \sum_{c\in\Omega_r} \pi_{rc} = \sum_{c\in\Omega_r} (\bar{p}_r - c_r)q_{rc}$$

Solving the maximization problem, we get that the optimal price of retail chain r is,

$$p_{rc}^{u} = \bar{p}_r = c_r \frac{\sum_{k\in\Omega_r} \sigma_k S_{rk}}{\sum_{k\in\Omega_r} S_{rk}(\sigma_k - 1)} \quad \forall c \in \Omega_r, \tag{5.1}$$

where, as in the empirical section, I define S_{rc} as the share of sales of retail chain r that take place in county c.

$$S_{rc} = \frac{\bar{p}_r q_{rc}}{\sum\limits_{k \in \Omega_r} \bar{p}_r q_{rc}} = \frac{\text{Sales of chain r in county c}}{\text{National sales of chain r}}.$$

Note that under uniform pricing:

1. p^u_{rc} depends on demand conditions in every market $k \neq c$.
2. A retail chain weights more demand conditions in markets where S_{rk} is larger.

Equation 5.1 anticipates the main mechanism through which uniform pricing operates. The price of a retail chain in a county is a weighted average of demand conditions in each of the markets where the retail chain operates, where the weights are increasing in the share of the retail chain's sales that take place in the market. Hence, retail chains' prices react more in response to shocks in its larger markets.[36]

After characterizing the retail chain's optimal prices under uniform pricing, I explore the theoretical implications at the county-level. To do so, I aggregate optimal prices in order to construct a county-level price index. For simplicity, I construct a Geometric Laspeyers Price Index:

$$P^U_{ct} = \prod_{rct \in \Omega_c} \left(\frac{P_{rct}}{P_{rc0}} \right)^{l_{rc}},$$

where, as in the empirical section, the weights l_{rc} represent the share of retail chain r in county c annual revenues. Replacing the optimal uniform prices, I obtain the theoretical county level price index under uniform pricing:

$$P^U_{ct} = \prod_{rct \in \Omega_c} \left(\frac{\sum\limits_{k \in \Omega_r} \sigma_k S_{rk}}{\sum\limits_{k \in \Omega_r} S_{rk}(\sigma_k - 1)} c_r \right)^{l_{rc}}$$

By total differentiating the county-level price index under uniform pricing, I can explore theoretically the sources of variation of this price index:

$$d \log P^U_{ct} = - \left[\sum_{r \in \Omega_c} l_{rc} \theta_{rc} \right] d \log \sigma_c - \sum_{k \neq c \in \Omega_r} \sum_{r \in \Omega_c} l_{rc} \theta_{rk} d \log \sigma_k + \sum_{r \in \Omega_c} l_{rc} d \log c_r \tag{5.2}$$

where

$$\theta_{rk} = \frac{S_{rk}\sigma_k}{\left[\sum\limits_{k \in \Omega_r} S_{rk}\sigma_k \right] \left[\sum\limits_{k \in \Omega_r} S_{rk}(\sigma_k - 1) \right]}.$$

[36] Note that if I define the markup in market c as $\mu_c = \frac{\sigma_c}{\sigma_c - 1}$, I can re-write equation 5.1 as a weighted average of markups in the different markets; where the weights are given by $S_{rk}(\sigma_k - 1)$:

$$p^u_{rc} = \bar{P}_r = c_r \frac{\sum_{k \in \Omega_r} \mu_k S_{rk}(\sigma_k - 1)}{\sum_{k \in \Omega_r} S_{rk}(\sigma_k - 1)} \quad \forall c \in \Omega_r.$$

The weights θ_{rk} determine the importance of county k for retail chain r. Note that θ_{rk} is increasing in S_{rk}, but it is not necessarily equal. The weights θ_{rk} would be proportional to S_{rk} only if the initial demand elasticity is equal for every county. In particular, if $\tilde{\sigma} = \sigma_c$ for all markets c, then $\theta_{rk} = \frac{1}{\tilde{\sigma}-1} S_{rk}$. Where $\tilde{\sigma}$ denotes the average demand elasticity across counties. In the next section, we explore the implications that cross-county heterogeneity in initial demand elasticity has for my empirical analysis.

Define the theoretical exposure of county c to county k as,

$$\kappa_{ck} = \sum_{r \in \Omega_c} l_{rc} \theta_{rk} \tag{5.3}$$

Equation 5.2 summarizes the three sources of variation for the county-level price index. First, local demand shocks will affect the local consumer price index, as long as the county is important for its dominant chains (high κ_{cc}). Second, demand shocks in other locations will also affect the local price index. The extent to which demand shocks in a county k affect county c price index depend on how important is market k (θ_{rk}) for dominant chains (l_{rc}) in county c (l_{rc}). Finally, changes in the national costs of retail chains that operate in the county (c_r) will also affect the county-level price index.

In the next section, I use Equation 5.2 to take the model to the data in order to 1) quantitatively test for uniform pricing, 2) rationalize the reduced-form estimates and 3) conduct counterfactual analysis.

6 Quantitative analysis

In the empirical section, I evaluated the propagation of shocks throughout the different counties, by analyzing the effect of a house prices shock in counties $k \neq c$ on the retail price index in county c. In this section, I conduct a model-based analysis to interpret the results through the lens of the model, quantitatively test for uniform pricing, and conduct counterfactual analysis.

6.1 Taking the model to the data

In this section, I introduce assumptions about the relationship between house price changes and the parameters in the model in order to rationalize the main findings in the empirical section.

In order to emphasize the demand channel of the Great Recession, I assume that the elasticity of demand varies with house prices. Let,

$$\beta^H = -\frac{\partial \log \sigma_c}{\partial \log HP_c} \tag{6.1}$$

In principle, β^H could have any sign. However, I expect $\beta^H > 0$, reflecting that when a county is exposed to a house price-induced negative demand shock, it becomes relatively poorer and the price elasticity increases (as, for example, in Simonovska (2015)). Note that I relax the CES assumptions by letting the demand elasticity vary with house prices, but I assume that it varies in the same proportion in every county (β^H is constant across counties).

I assume that national costs of the retail chain are uncorrelated with local house price shocks to retail chains. Several pieces of evidence suggest that this may be a reasonable assumption. For example, Stroebel and Vavra (2019) use data on wholesale costs and show that geographic variation in the costs is minimal. In addition, they show that marginal costs are not affected by local house price shocks. They present further evidence suggesting that neither variation in labor costs nor variation in retail rents play a significant role.[37]

Plugg ing Equation 6.1 in Equation 5.2 to make explicit that the demand elasticity is a function of house prices,

$$d \log P_{ct} = \frac{\beta^H}{\tilde{\sigma} - 1} \kappa_{cc} d \log HP_c + \frac{\beta^H}{\tilde{\sigma} - 1} \sum_{k \neq c \in \Omega_r} \kappa_{ck} d \log HP_k + \sum_{r \in \Omega_c} l_{rc} d \log c_r, \tag{6.2}$$

Equation 6.2 resembles my main empirical Equation 4.5 (which I re-write below), with two important differences that I account for.

$$d \log P_c = \beta_1 d \log HP_c + \beta_2 \sum_{k \neq c \in \Omega_r} \omega_{ck} d \log HP_k + \epsilon_c \tag{6.3}$$

The first difference is related to the direct effect of a house price shock (see blue terms). The model highlights the fact that the effect of local shocks on local retail prices is heterogeneous. It is a function of how important the local consumer market is in the national sales of the retailers who enter the local consumption basket (κ_{cc}).[38] This has two important implications. First, it implies that uniform pricing strategies can be consistent with large local elasticity of prices with respect to demand shocks; as long as the shocks are in regions with high κ_{cc}. Given that counties with higher κ_{cc} were more affected by the house price slump, this helps reconcile uniform pricing strategies (DellaVigna and Gentzkow (2019)) with large local elasticity of retail prices with respect to house prices (Stroebel and Vavra (2019)).[39] Second, it implies that researchers interested in the structural parameter β^H (or the elasticity of σ to other economic shocks) should weight the local shocks by the importance of the own county for the dominant retail chains operating in the county.

The second difference is related to the construction of the weights (see terms in red). Although the theoretical weights θ_{rk} are increasing in the empirical weights S_{rk}, they are not exactly equal. That is, $\kappa_{ck} \neq \omega_{ck}$. In particular, the weights θ_{rk} adjust S_{rk} by initial heterogeneity in demand elasticity across counties. In order to check the sensitivity of my results, I adjust for heterogeneity in initial elasticity of demand and construct the empirical counter-part of the weights θ_{rk}. To do so, I use county-level estimates of σ_c by quartile of population from a paper by Hottman (2014): 3.9 for the

[37] In Appendix E.2, I formally discuss the relevance and implications of this assumption.

[38] κ_{cc} varies greatly across US counties. The mean is 0.03 and ranges from 0.0001 in the 5th percentile to 0.12 in the 95th percentile, with a standard deviation of 0.06.

[39] A remaining concern about Stroebel and Vavra (2019) estimates is that shocks in other counties are correlated with shocks in the own county. Hence, the direct effect (the elasticity of local retail prices with respect to local house prices) in their paper also captures part of the propagation from counties connected by the network of retail chains.

1st and 2nd quartiles, 4.5 for the 3rd quartile and 4.8 for the 4th quartile of population (average: 4.1). Results from running my main specification after adjusting ω_{ck} for heterogeneity in initial demand elasticity are reported in Table A15 of the appendix. Across a variety of specifications, the coefficient lies within one standard error of the coefficient in my main specification.

I now proceed to construct the empirical counterpart of Equation 6.2 in order to test for uniform pricing and recover the structural parameter β^H.

6.2 A test for uniform pricing

The theoretical Equation 6.2 provides i) a quantitative test for uniform pricing; and, at the same time, ii) a methodology to recover the structural parameter β^H. If retail chains prices respond uniformly across markets, then the direct effect -weighted by the own county exposure (κ_{cc})- should be equal to the indirect effect from counties connected by the network of retail chains: $\beta_1 = \beta_2 = \frac{\beta_H}{\tilde{\sigma}-1}$.

I use county-level demand elasticities from Hottman (2014) and adjust the local change in house prices with κ_{cc} in order to construct the empirical counterpart of theoretical equation 6.2. I then run my main empirical specification, but using the empirical counterpart of Equation 6.2. Results are reported in Table 6. First, note that I cannot reject the hypothesis that the coefficient for the direct effect and for effect from propagation are statistically equal (I report the t-test in the last row of Panel A). That is, I cannot reject the hypothesis that the uniform pricing model accounts for the observed co-movements in county-level prices. Second, by scaling the estimate in column (6) by the average demand elasticity ($\tilde{\sigma}$-1=3.1), I obtain an estimate for β^H: $\hat{\beta_H} = 0.48$. I use this estimate for counterfactual analysis in the next section.

Table 6: A test for uniform pricing: $\beta_1 = \beta_2$?

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP (adjusted)	0.314***	0.108*	0.106*	0.609***	0.168	0.179
	(0.095)	(0.059)	(0.058)	(0.116)	(0.112)	(0.114)
Chain-linked Log Δ HP (other counties)		0.109***	0.108***		0.155***	0.155***
		(0.027)	(0.028)		(0.047)	(0.049)
Observations	910	910	910	910	910	910
R-squared	0.035	0.235	0.249	0.006	0.118	0.132
County controls	no	no	yes	no	no	yes
p-value $\beta_1 = \beta_2$	-	0.989	0.979	-	0.830	0.922
Panel B: First Stage						
First Stage F-stat				13.340	11.348	11.127

A unit of observation is a county. The dependent variable is the percentage change in the county price index between 2007 and 2011 ($\Delta \log(P_c)$). The table reports results of estimating the empirical counterpart of Equation 6.2. County's $\Delta LogHP$ is adjusted by κ_{cc}. Columns (1) to (3) report results for OLS estimations. Columns (4) to (6) report results, after instrumenting the two main endogeneous variables. I instrument ($\sum_{k \neq c} \theta_{ck} \Delta \log(HP_k)$) with network-weighted WLRI in other counties ($\sum_{k \neq c} \theta_{ck} WLRI_k$). Panel A reports the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively.

6.3 Counterfactuals

We now have the theoretical structure, the estimates of the key parameter $\beta^H = 0.48$, demand elasticities estimated in previous papers and the data to carry out a quantitative assessment of the influence of retail chains on the spread of demand shocks during the Great Recession.

In particular, I hold constant changes in retail chains' costs ($d \log c_r = 0$) and use Equation 6.2 from the calibrated model to evaluate how the demand shocks during the Great Recession would have affected cross-county price changes under different scenarios.

I propose two set of counterfactuals to evaluate the implications of uniform pricing and the geographic distribution of retail chains for the cross-county dispersion of inflation and for the distribution of the effect of the shocks. First, I compare different pricing strategies. Second, I change the spatial networks of retail chains by mergers and de-mergers of the largest retail chains.

6.3.1 Counterfactual 1: Pricing strategy

In this counterfactual, I map the shares S_{rk} and l_{rc} to retail chains in the data and analyze two polar opposite pricing strategies: uniform pricing (benchmark) and flexible pricing.

I use these counterfactuals to quantitatively answer two questions: What would have been the the cross-county dispersion of inflation rates during the Great Recession if pricing strategies were different? Which consumers would have benefited from the different pricing strategies?

The effect on cross-county dispersion of inflation rates

I study the quantitative implications of uniform pricing for the cross-county dispersion of inflation rates. Table **??** reports the average change in retail prices during the Great Recession by quintiles of changes in local house prices under the two alternative pricing strategies. The first quintile represents counties more directly exposed to the house price slump.[40] The first column reports the change in consumer retail prices for the uniform pricing case, while the second column reports the change in consumer retail prices for the flexible pricing case.

Note that the average inflation rate of the U.S. under the two alternative pricing strategies is similar. However, there are important differences in the dispersion of the inflation rate induced by the demand shocks during the Great Recession. Comparing the coefficient of variation (CV), we observe that the dispersion in cross-county inflation under uniform pricing is 40% lower than under flexible pricing. The lower cross-county dispersion of inflation indicates that uniform pricing smoothed-out the local effect of shocks that took place during the Great Recession. First, it attenuated the effect of the local shocks on local prices. Intensifying the effect of the crisis in the local market. Second, regardless of the local crisis, counties connected to affected counties experienced a decline in their consumer prices. So in the context of a negative demand shock, retail chains' centralized pricing decisions can actually intensify the adverse consequence of the local crisis. If Miami is hit hard by the recession, retail chains do not reduce prices as much there and, instead, they reduce prices in counties like Houston that were less affected by the crisis.[41] Specifically,

[40] In Figure A4 of Appendix F.1 I plot the distribution of changes in prices for each of the alternative scenarios.

[41] This may not be the case for workers in the retail sector. While consumers of retail chains exposed to negative shocks in other counties benefit from lower prices, their workers might lose their jobs (See Giroud and Mueller (2019)).

compared to flexible pricing, the reduction in consumer prices under uniform pricing was 1.25 percent points lower for the top county quintile most affected by the house price slump, while it was 2.23 percent points higher for the county quintile least affected by the house price slump. Examples of counties that would have been better off with flexible pricing include Riverside (CA), Clark County (Nevada), Maricopa (Arizona) and Miami-Dade (Florida). In contrast, St. Louis (Minessota), Aroostook (Maine), Graham (Arizona), Houston (Texas), Philadelphia (Pennsylvania) are examples of counties that benefited from uniform pricing during the Great Recession (See Figure A6 of Appendix F.1).

Table 7: Inflation Rates

Quintile of Δ House Price	Uniform ΔP_c^u	Flexible ΔP_c^{flex}	Δ^{Losses} ΔP_c^u-ΔP_c^{flex}
Most affected	-4.19%	-5.44%	1.25%
2	-2.67%	-2.07%	-0.60%
3	-2.61%	-1.55%	-1.06%
4	-2.56%	-1.13%	-1.43%
Least Affected	-2.59%	-0.36%	-2.23%
Mean	-3.42%	-3.22%	-0.20%
SD	1.52%	2.75%	
CV	44.44%	85.47%	

The quintiles are defined in terms of the house price changes. The largest drop in house prices corresponds to the first quintile. Observations are weighted by population in 2007.

The effect of uniform pricing on real income inequality

In the cross-section, uniform pricing tends to exacerbate inequality, as long as the demand elasticity is higher for poorer consumers.[42] However, it is less clear how uniform pricing affects real income inequality in response to a shock. It will depend on the magnitude, sign and location of the shocks, as well on the geographic distribution of retail chains. In this section, I explore whether uniform pricing strategies during the Great Recession benefited consumers from low-income or from high-income counties.

Define the losses in purchasing power from uniform pricing relative to flexible prices as,

$$\Delta_c^{Losses} = \Delta \log(P_c^u) - \Delta \log(P_c^{flex}),$$

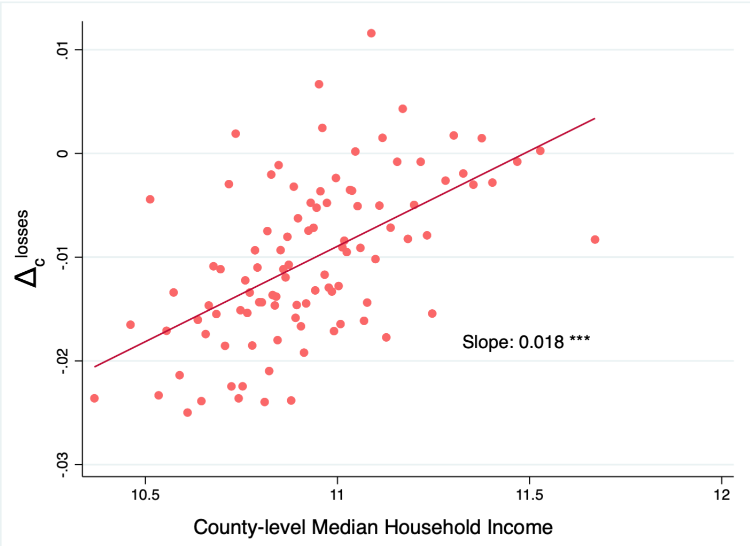
where higher values imply that county-level inflation under uniform pricing is higher than under flexible pricing. In Figure 6, I plot the binscatter of the relationship between Δ^{Losses} and the log county-level median household income. The pattern is clear. Compared to a scenario in which

[42] In calibrated example, DellaVigna and Gentzkow (2019) shows that uniform pricing increases the prices faced by consumers in the poorest decile of zip-codes.

prices are flexible, low-income counties benefited most from uniform pricing. Intuitively, low-income counties were less exposed to local drops in house prices, but still benefited from drops in local consumer prices because their retail chains also operate in counties that were more exposed to the house price slump.

It is worth mentioning that the re-distributive consequences of uniform pricing discussed above are specific to the local shocks that took place during the Great Recession. For instance, one can imagine that, in the years prior to the crisis, when house prices were increasing, uniform pricing actually benefited high-income counties. More generally, uniform pricing leads to a more synchronized inflation rate across counties. Hence, understanding how counties are connected through the retail chain networks can help a policy maker to design better-informed policies to alleviate the local effects of shocks, and also the indirect effects.

Figure 6: Uniform pricing benefited low-income counties during the Great Recession



Note: Binscatter of the relationship between the logarithm of county-level median household income (x-axis) and $\Delta_c^{losses} = (\Delta \log(P_c^U) - \Delta \log(P_c^{flex})$ (y-axis). The elasticity is 0.018 and is statistically significant at 1% level.

6.3.2 Counterfactual 2: Mergers or acquisitions

Mergers have been one of the most important public policy concerns in the design of antitrust laws. A salient feature of the retail sector is that oftentimes firms expand their network of stores by merging with (or acquiring) firms that operate in different regions. For example, in 1998, Kroger merged with the then fifth-largest grocery company Fred Meyer, along with its subsidiaries, Ralphs, QFC, and Smith's. This type of merger changes the spatial networks of retail chains, affecting the linkages between counties and, thus, the propagation of shocks.

In this counterfactual, I study what would have been the cross-county dispersion of inflation rates during the Great Recession if the ownership structure of the major retail chains had been different. Note that in the extreme scenario in which there is only one retail chain, then the cross-county dispersion of inflation after a shock would be zero. Alternatively, if all retail chains split up into local stores, the cross-county dispersion of inflation will tend to the dispersion under flexible pricing. I explore two other scenarios:

(a) *De-merger:* the largest retail chain splits up into four different retail chains, each corresponding to one of the four Census regions: West, Midwest, Northeast, South.

(b) *Merger:* merger between the largest retail chain in each of the four Census regions of the U.S. (West, Midwest, Northeast, South).

Results are presented in Table 8. First, notice that the average inflation rate is very similar in the three scenarios. However, the cross-county dispersion of inflation would have been considerably different under the alternative scenarios. As expected, when the largest firm split up into regions, the cross-county dispersion of inflation increases 5%. The reason for this is that the largest chain now only propagates shocks within regions, but not across them. In contrast, when the four largest chains in each region merge, the cross-county dispersion of inflation rates declines 12%. Intuitively, this merger expands the retail chains' spatial networks, which increases the connections between counties. This, in turn, indicates that, under uniform pricing, mergers between retail chains in different regions would have smoothed out the impact of local shocks during the Great Recession across regions even further.

Table 8: Changes in the distribution of firms: Mergers and Partitions

$\Delta \log(P_c)$	Empirical	De-merger	Merger
Mean	-3.42 %	-3.42%	-3.45%
SD	1.52%	1.59%	1.36%
CV	0.44	0.46	0.39

Note: Counties are weighted by population.

As mergers smooth out the effects of local shocks across regions, they might also have re-distributive consequences in the event of a shock. For example, during the Great Recession, I find that low-income counties would have been made better off had there been a merger between the largest retail chains in each region. Specifically, in the scenario of a merger, the poorest quintile would have experienced a 0.18 p.p lower inflation (6%) than under the benchmark case. In contrast, the richest quintile, would have experienced a 0.13 p.p. higher inflation rate (4%). The reason for this pattern is that high-income counties were more affected by shocks during that period than were low-income counties. Therefore, as mergers between retail chains in different regions intensify the linkages between those regions, the effect of the shock in high-income counties spreads to low-income counties which benefit from reduction in their retail prices.

Again, it is worth mentioning that conclusions related to the redistributive effects are specific to this period, and to the shocks I am analyzing. The more general conclusion that we obtain from this findings is that, under uniform pricing, mergers between retail chains that are located in different regions intensify the linkages between these regions, increasing the contribution of retail chains to the observed co-movement in county-level consumer prices.

7 Conclusion

I study the role of retail chains' spatial networks of stores in shaping the propagation of shocks across regions in the United States. My empirical results show that the county-level consumer price index is sensitive to shocks in distant counties in which the same retail chains operate. My results hold after controlling for trade relationships due to proximity. In fact, I show that once the network of retail chains is taken into account, the most traditional channel of propagation of shocks to nearby counties plays no additional role.

I find that the key mechanism behind the propagation of shocks is the inter-dependencies in pricing strategies between stores of a given retail chain. As retail chains set prices uniformly, they take into account demand conditions in all their markets to determine optimal prices. Hence, in response to a negative demand shock in one market, they react by changing prices in many markets.

The pricing strategies of retail chains have important consequences for the re-distribution of shocks in the economy. Uniform pricing strategies smooth-out the effect of local shocks on prices. This has two direct implications. First, uniform pricing attenuates the pro-cyclical behavior of local prices. Second, uniform pricing increases the synchronization of consumer prices across counties. In the context of a negative demand shock, uniform pricing deteriorates the situation of consumers in counties more directly affected by the shock. On this ground, I show that during the Great Recession, uniform pricing reduced the cross-county dispersion of inflation by 40% (compared to flexible pricing). Since low-income counties were relatively less directly exposed to the house price slump, they benefited from uniform pricing.

These findings have important implications for the design of policies. My paper emphasizes that retail chains' networks create linkages between regions. These linkages determine the patterns of propagation of shocks to consumer prices. Therefore, when weighting the costs and benefits of a policy, policy makers should consider not only the impact on local prices, but also the indirect impact on prices in distant regions. More broadly, when prices are rising in a region (for example, due to gentrification), the design of optimal policies should also take into account what is happening in the regions that are connected by the network of retail chains.

References

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016.

ADAMS, B. AND K. R. WILLIAMS (2019): "Zone pricing in retail oligopoly," *American Economic Journal: Microeconomics*, 11, 124–56.

ADAO, R., M. KOLESÁR, AND E. MORALES (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010.

ADELINO, M., A. SCHOAR, AND F. SEVERINO (2015): "House prices, collateral, and self-employment," *Journal of Financial Economics*, 117, 288–306.

ALVIAREZ, V. I., J. CRAVINO, AND N. RAMONDO (2020): "Firm-Embedded Productivity and Cross-Country Income Differences," Tech. rep., National Bureau of Economic Research.

BILIR, L. K. AND E. MORALES (2020): "Innovation in the global firm," *Journal of Political Economy*, 128, 1566–1625.

BOEHM, C. E., A. FLAAEN, AND N. PANDALAI-NAYAR (2019): "Input linkages and the transmission of shocks: firm-level evidence from the 2011 Tōhoku earthquake," *Review of Economics and Statistics*, 101, 60–75.

BORUSYAK, K., P. HULL, AND X. JARAVEL (2018): "Quasi-experimental shift-share research designs," Tech. rep., National Bureau of Economic Research.

BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541–585.

CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2017): "The impact of regional and sectoral productivity changes on the US economy," *The Review of economic studies*, 85, 2042–2096.

CARVALHO, V. M., M. NIREI, Y. SAITO, AND A. TAHBAZ-SALEHI (2016): "Supply chain disruptions: Evidence from the great east japan earthquake," *Columbia Business School Research Paper*.

CAVALLO, A. (2017): "Are online and offline prices similar? Evidence from large multi-channel retailers," *American Economic Review*, 107, 283–303.

CRAVINO, J. AND A. A. LEVCHENKO (2017): "Multinational firms and international business cycle transmission," *The Quarterly Journal of Economics*, 132, 921–962.

DARUICH, D. AND J. KOZLOWSKI (2020): "Macroeconomic Implications of Uniform Pricing," .

DELLAVIGNA, S. AND M. GENTZKOW (2019): "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 134, 2011–2084.

DESAI, M. A., C. F. FOLEY, AND J. R. HINES (2009): "Domestic effects of the foreign activities of US multinationals," *American Economic Journal: Economic Policy*, 1, 181–203.

DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 131, 799–858.

FEENSTRA, R. C. (1994): "New product varieties and the measurement of international prices," *American Economic Review*, 84, 157–157.

GIROUD, X. AND H. M. MUELLER (2019): "Firms' Internal Networks and Local Economic Shocks," *American Economic Review*, 109, 3617–49.

GOLDSMITH-PINKHAM, P., I. SORKIN, AND H. SWIFT (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

GYOURKO, J., A. SAIZ, AND A. SUMMERS (2008a): "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45, 693–729.

——— (2008b): "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45, 693–729.

HANDBURY, J. AND S. MOSHARY (2020): "School Food Policy Affects Everyone: Retail Responses to the National School Lunch Program," .

HOTTMAN, C. (2014): "Retail markups, misallocation, and store variety in the US," *Mimeograph, Columbia University*.

HYUN, J. AND R. KIM (2019): "Spillovers and Redistribution through Intra-Firm Networks: The Product Replacement Channel," *Available at SSRN 3480961*.

MIAN, A., K. RAO, AND A. SUFI (2013): "Household balance sheets, consumption, and the economic slump," *The Quarterly Journal of Economics*, 128, 1687–1726.

MIAN, A. AND A. SUFI (2011): "House prices, home equity-based borrowing, and the US household leverage crisis," *American Economic Review*, 101, 2132–56.

NAKAMURA, E. (2008): "Pass-through in retail and wholesale," *American Economic Review*, 98, 430–37.

SAIZ, A. (2010): "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 125, 1253–1296.

SATO, K. (1976): "The ideal log-change index number," *The Review of Economics and Statistics*, 223–228.

SIMONOVSKA, I. (2015): "Income differences and prices of tradables: Insights from an online retailer," *The Review of Economic Studies*, 82, 1612–1656.

STROEBEL, J. AND J. VAVRA (2019): "House prices, local demand, and retail prices," *Journal of Political Economy*, 127, 1391–1436.

STUMPNER, S. (2019): "Trade and the geographic spread of the great recession," *Journal of International Economics*, 119, 169–180.

VARTIA, Y. O. (1976): "Ideal log-change index numbers," *scandinavian Journal of statistics*, 121–126.

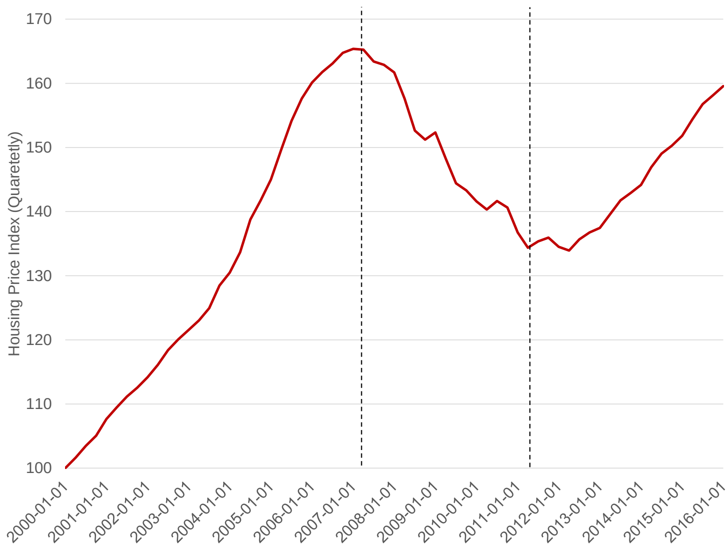
8 Appendix

A Data

A.1 House prices Data

Figure A1 depicts the evolution of the national house price index. We can observe the collapse in house prices started in the beginning of 2007. House prices continued declining until the second semester of 2011. Throughout that period, house prices dropped 18%.

Figure A1: Evolution of House Price Index



Note: av

A.2 Differences between counties in sample and out of sample

For my main analysis, I only consider counties for which I have information on change in house prices (2208), Nielsen Scanner Data (2300), and Wharton Land Regulation Index (910). Hence, in my baseline specification, I consider only 910 counties for which I have information about the WLRI. In Table A1, I show that these 910 counties represent 70% of the population and 70% of the sales in the Nielsen Data. These counties where hit slightly more by the crisis and had a smaller initial unemployment rate.

Table A1: Counties with and without data on Wharton Land Regulation Index

Data on WLRI avaiable	Counties #	Population Total (millions)	Sales Total (millions)	$\Delta Log(HP)^{2007-2011}$ % (mean)	Unemployment rate (mean)
No	1298	86.64	13798.59	-10.50	5.07
Yes	910	200.21	33638.94	-13.41	4.83
Total	2208	286.86	47437.53	-11.71	-4.97

B Price Indices: Main analysis.

As the main focus of the paper is on a) prices of existing products that are similar within chain across stores, and b) variation of price indices across time, we include an item only if it has positive sales in 2007 and 2011. We track the price of identical items (UPC-store combinations) across time, so that changes in quality or issues with comparing nonidentical products are less relevant for our results.

We construct two price indices, at two different levels of aggregation: A) A price index for each County (P_{ct}), and a retail chain by county price Index (P_{rct}). I describe them in detail in next sections.

B.1 County level Price index

We construct county-level price index in two steps. We first construct a product-module level price index. Ignoring the introduction of new varieties, the exact price index of the CES utility function for product module m in county c is as in Sato (1976) and Vartia (1976):[43]

$$P_{mct} = \Pi_{u \in I_{mc}} \left(\frac{P_{umct}}{P_{umct-1}} \right)^{w_{umct}},$$

where

$$w_{ucmt} = \frac{(s_{umct} - s_{umct-1}) / (ln(s_{umct}) - ln(s_{umct-1}))}{\sum_{v \in I_{mc}} (s_{vmct} - s_{vmct-1}) / (ln(s_{vmct}) - ln(s_{vmct-1}))} \;\; ; \;\; s_{umct} = \frac{P_{umct} Q_{umct}}{\sum_{v \in I_{mc}} P_{vmct} Q_{vmct}},$$

and I_{mc} is the set of varieties in product module m in county c that are consumed in both years (continuers). The weights w are ideal log-change weights and they are county specific to allow for

[43]This price index is consistent with the following utility function:

$$U_c(y_c) = \prod_{m \in R_c} \left[\sum_{u \in R_{mc}} q_{umc}^{\frac{\sigma_m(y_c)-1}{\sigma_m(y_c)}} \right]^{\alpha_{mc} \frac{\sigma_m(y_c)}{\sigma_m(y_c)-1}}$$

Consumer behavior features multi-stage budgeting in two stages. In the first stage, consumers in a county decide which of 1000 product modules to buy from based on the product module price index. In the second stage, conditional on the product module, consumers decide which variety to purchase; where variety is defined as a store-barcode combination (eg. 12 oz. Coke in 7-eleven).

spatial variation the relative weight of an item. [44]

I then construct an overall county-specific price index by weighting these category price indices by the revenue share of a particular product module in the initial year,

$$P_{ct} = \Pi_m \left(\frac{P_{mct}}{P_{mct-1}} \right)^{w_{mct-1}},$$

where

$$w_{mct-1} = \frac{\sum_{u \in m} Sales_{umct-1}}{\sum_u Sales_{uct-1}}$$

In figure A2, I report the histogram of county-level inflation rate between 2007 and 2011. There is substantial heterogeneity in inflation rates by county in the US. The population-weighted average inflation rate of the counties in our sample is 11.09%, which contrast well with the variation in the food at home official CPI from BLS. Table A2 reports inflation rate for the 10 counties with highest inflation rate and for the 10 counties with the lowest inflation rate (excluding outliers in the top/bottom 1 percent).

Figure A2: Histogram: County's percentage change in retail prices between 2007 and 2011

[44]Note that they are always bounded between the shares of spending in period t and period t-1.

Table A2: Counties with lowest and highest inflation rate

County	County's Δ Log Price Index 07-11
Lowest Inflation Rate	
Contra Costa County, California	0.067
Bonner County, Idaho	0.067
Letcher County, Kentucky	0.068
Prowers County, Colorado	0.068
Buchanan County, Virginia	0.068
Leelanau County, Michigan	0.068
San Francisco County, California	0.068
Marin County, California	0.069
Sonoma County, California	0.069
Pike County, Alabama	0.069
Highest Inflation Rate	
Richmond County, North Carolina	0.180
Ben Hill County, Georgia	0.180
Jasper County, Iowa	0.180
Grundy County, Iowa	0.181
McCormick County, South Carolina	0.181
Lawrence County, Arkansas	0.182
Atkinson County, Georgia	0.182
Terrell County, Georgia	0.184
Chautauqua County, New York	0.185
Dickinson County, Iowa	0.185
U.S population-weighted average	0.111

Retail chain by county Price Index

Similarly, we construct a price index at the retail chain by county level.

First, we first construct a product-module level price index within retail chain in a county.

$$P_{mrct} = \Pi_{u \in I_{mrc}} \left(\frac{P_{umrct}}{P_{umrct-1}} \right)^{w_{umrct}},$$

where

$$w_{umrct} = \frac{(s_{umrct} - s_{umrct-1}) / (ln(s_{umrct}) - ln(s_{umrct-1}))}{\sum_{v \in I_{mrc}} (s_{vmrct} - s_{vmrct-1}) / (ln(s_{vmrct}) - ln(s_{vmrct-1}))} \;\; ; \;\; s_{umrct} = \frac{P_{umrct} Q_{umrct}}{\sum_{v \in I_{mrc}} P_{vmrct} Q_{vmrct}},$$

and I_{mrc} is the set of varieties in product module m, sold by retailer chain r that are consumed in both years (continuers). The weights w are ideal log-change weights and they are retailer chain specific to allow for variation in the importance of items in different retailer chains.

We then construct a retail chain by county price index by weighting these category price indices by the revenue share in $t-1$ of a particular product module in the retail chain in the county.

$$P_{rct} = \Pi_m \left(\frac{P_{mrct}}{P_{mrct-1}} \right)^{w_{mrct-1}},$$

where

$$w_{mrct-1} = \frac{\sum_{u \in m} Sales_{umrct-1}}{\sum_u Sales_{urct-1}}$$

C Geographic Distribution of Retail Chains

Nielsen does not disclose the names of the chains in the data. In order to to provide a broad view of the geographic distribution of retail chains, in Figure A3 I use AggData from 2006 to illustrate the geographic distribution of some popular retail chains .[45]

Figure A3: Geographic Distribution of retail chains

Source: AggData for 2006

D Empirical Analysis

D.1 Robustness Checks

D.1.1 Alternative county-level controls

I begin by exploring whether the main IV estimates are stable across specifications with different county-level controls. Results are reported in Table A3. Reassuringly, the coefficient associated with chain-linked percentage change in house prices in other counties remains stable across different specifications. This suggests that, once I control for county's house price growth, the instrument is not correlated with other county-level variables in the error term.

[45]Some of these chains are in my data, while some of them are not. This is an illustrative example.

Table A3: Validity of identification assumption: Adding controls to IV estimations

Dep Variable: County's Δ Log Price Index						
	IV: WLRI instrument					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.074**	0.075**	0.078**	0.076**	0.070	0.057
	(0.034)	(0.036)	(0.038)	(0.037)	(0.045)	(0.046)
Chain-linked Δ Log HP (other counties)	0.129***	0.128***	0.132***	0.135***	0.138***	0.147***
	(0.039)	(0.041)	(0.041)	(0.040)	(0.042)	(0.045)
County's Δ Log Wages		-0.002	0.005	0.031**	0.032**	0.024*
		(0.016)	(0.013)	(0.014)	(0.014)	(0.013)
County's Δ Log # Establishments			-0.036	-0.020	-0.020	-0.018
			(0.025)	(0.026)	(0.025)	(0.022)
County's Δ Log Employment				-0.028*	-0.028**	-0.029**
				(0.014)	(0.014)	(0.013)
County's Log Retail Sales in 2007					-0.000	-0.000
					(0.001)	(0.001)
County's Log Market Access 2007						-0.006
						(0.004)
Panel B: First Stage						
F-stat	22.100	20.746	14.286	14.479	13.172	13.223
Observations	910	910	910	910	910	910
R-squared	0.059	0.062	0.048	0.060	0.077	0.096

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the percentage change in the county price index between 2007 and 2011. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (1) to (6) report results, after instrumenting county exposure to house price shocks through the network of retailer chains with county exposure to WLRI through the network of retailer chains. Panel A report the second stage coefficients. Robust standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.2 Test of Overidentifying assumptions: Saiz Housing Supply elasticity instrument

As a further check, I explore whether the main results change if I use another common instrument in the literature. In particular, I instrument county's percentage change in house price with housing supply elasticity (HSE, hereafter) constructed by Saiz (2010). Saiz (2010) uses geographic information of the metropolitan area to measure how easy is constructing new houses (e.g: areas with a flat topology are assigned with a higher elasticity). Naturally, I also use the Saiz HSE and the weights ω_{ck} to construct an instrument for network-weighted percentage change in house price in other counties.[46] Results are reported in Table A4. The first two columns present OLS coefficients. Column (3) and (4) reports results using the WRLI instrument, as in my main specification. Columns (5) and (6) instrument the endogeneous variables with the Saiz HSE. The main coeffi-

[46] Saiz HSE is at the MSA level so I have less variability. In addition, it is available for a lower number of counties. For this reason, the WLRI instrument is my baseline specification.

cient increases modestly. Columns (7) and (8) present results including the 4 instruments. Again, throughout specifications, the coefficient remains stable between 0.129 and 0.146. In addition, I cannot reject the hypothesis that the coefficients obtained by using the different instruments are different.

Table A4: Validity of identification assumption: Saiz (2010) housing supply elasticity instrument

	Dep Variable: County's Δ Log Price Index					
	IV: WLRI instrument		IV: HSE (Saiz) instrument		IV: All instruments	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.074**	0.076**	0.042*	0.040	0.054**	0.055*
	(0.034)	(0.037)	(0.023)	(0.026)	(0.027)	(0.031)
Chain-linked Δ Log HP (other counties)	0.129***	0.135***	0.139***	0.145***	0.140***	0.146***
	(0.039)	(0.040)	(0.037)	(0.039)	(0.035)	(0.036)
Panel B: First Stage						
First Stage F-stat	18.642	14.756	13.868	10.525	11.558	8.787
Observations	910	910	903	903	903	903
R-squared	0.082	0.074	0.183	0.185	0.143	0.140
County controls	no	yes	no	yes	no	yes

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retail chains distribution of stores: $\sum_{k \neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (8) report results, after instrumenting county exposure to house price shocks through the network of retailer chains. In column (3) to (4), I instrument with exposure to WLRI, as in the baseline specification. In columns (5) and (6), I instrument with exposure to Saiz Housing supply elasticity. Columns (7) and (8) use the 4 instruments. Panel A report the second stage coefficients. Robust standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.3 Robustness Checks: Controlling for similarity-weighted house price changes in other counties

In Table A5, I provide another test for the possibility of common regional shocks at the county-level. Following the strategy in Giroud and Mueller (2019), I control for similarity-weighted house price changes in other counties. The weights place more weight on those counties that have a smaller absolute difference with respect to the county, based on different county-level covariates. The idea is that counties which are more similar in population, income, household debt and education are more likely to be exposed to the same county-level shocks and co-move in retail prices.

Table A5: Controlling for similarity-weighted shocks

	Dep Variable: County's Δ Log Price Index				
	IV: WLRI instrument				
	(1)	(2)	(3)	(4)	(5)
Panel A: Second Stage					
County's Δ Log HP	0.079**	0.063*	0.062*	0.063*	0.064*
	(0.035)	(0.032)	(0.034)	(0.032)	(0.033)
Chain-linked Δ Log HP (other counties)	0.135***	0.136***	0.134***	0.136***	0.129***
	(0.033)	(0.032)	(0.032)	(0.032)	(0.036)
Population-weighted Δ Log HP (other counties)		0.029			
		(0.036)			
Income-weighted Δ Log HP (other counties)			-0.034		
			(0.051)		
Education-weighted Δ Log HP (other counties)				-0.032	
				(0.042)	
HH debt-weighted Δ Log HP (other counties)					0.010
					(0.020)
Panel B: First Stage					
F-statistic	13.814	13.582	13.230	13.755	13.254
Observations	910	910	910	910	910
R-squared	0.060	0.093	0.097	0.092	0.089
County controls	yes	yes	yes	yes	yes

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retail chains distribution of stores: $\sum_{k \neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (5) control for similarity-weighted changes in house prices. Robust standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.4 Robustness Checks: Different assumptions to construct the Price Index

In my baseline specification, I choose the CES exact price index for continuing varieties to construct the price index. In this section, I show that my main results remain qualitatively unchanged under different assumptions to construct the price indices. In particular, I repeat my main analysis with Laspeyers, a Paasche and a Fisher price index. Results are reported in Table A6. We can observe that main results hold under these alternative county-level price indices.

Table A6: Robustness Check: Other Price Indices

	Dep Variable: County's Δ Log Price Index			
	Exact CES	Laspeyeres	Pasche	Fischer
	(1)	(2)	(3)	(4)
Panel A: Second Stage				
County's Δ Log HP	0.068*	0.069	0.082**	0.075*
	(0.040)	(0.043)	(0.040)	(0.041)
Chain-linked Δ Log HP (other counties)	0.135***	0.140***	0.098**	0.120***
	(0.040)	(0.040)	(0.041)	(0.040)
Observations	922	922	922	922
R-squared	0.060	0.043	0.072	0.071
County controls	yes	yes	yes	yes
Panel B: First Stage				
F-stat	14.756	14.756	14.756	14.756

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retailer chains distribution of stores: $\sum_{k\neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (8) report results, after instrumenting county exposure to house price shocks through the network of retailer chains. In column (3) to (4), I instrument with exposure to WLRI, as in the baseline specification. In columns (5) and (6), I instrument with exposure to Saiz Housing supply elasticity. Columns (7) and (8) use the 4 instruments. Panel A report the second stage coefficients. Robust standard errors clustered at the state level in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.5 Robustness Checks: Alternative period of analysis: 2007-2009

In my baseline specification, I choose the period between 2007 and 2011. In this section, I show that my main results remain qualitatively unchanged when I choose other period of analysis. Table A7 replicates main table 1 for the period 2007 to 2009.

Table A7: Propagation of house price-induced local demand shocks across counties through the network of retailer chains: 2007-2009

	Dep Variable: County's Δ Log Price Index (2007-2009)					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.044	0.023	0.012	0.121***	0.073*	0.069
	(0.027)	(0.026)	(0.027)	(0.043)	(0.044)	(0.046)
Chain-linked Δ Log HP (other counties)		0.074**	0.081***		0.110***	0.117***
		(0.030)	(0.030)		(0.037)	(0.037)
Panel B: First Stage						
F-stat				20.757	14.378	13.437
Observations	910	910	910	910	910	910
County controls	no	no	yes	no	no	yes

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the percentage change in the county price index between 2007 and 2009. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2009, where the weights are defined as $\omega_{c,k}$ in the main text. Columns (1) to (3) report results for OLS estimations. Columns (4) to (6) report results, after instrumenting the two main endogeneous variables. I instrument local percentage change in house prices with WLRI. I instrument county exposure to house price shocks through the network of retailer chains with county exposure to WLRI through the network of retailer chains. Panel A report the second stage coefficients. Robust standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.6 Robustness Checks: Alternative market definitions: three digit Zip Code level

In my baseline specification, I choose the county as the relevant market. In this section, I check the robustness of my results to a different market definition. In particular, I define markets as a three digit zip code.

D.1.7 Robustness Checks: Region fixed effects

I explore the sensitivity of results to different combinations of region fixed effects. Throughout specifications, the main coefficient remains positive and significant. Note that once I include state fixed effects, the instrument becomes weak as reflected in the first stage F-statistic; so the coefficients in column (4) should be interpreted with caution.

Table A8: Robustness Check: Region fixed effects

	County's Δ Log Price Index			
	(1)	(2)	(3)	(4)
County's Δ Log HP	0.065*	0.073*	0.091***	0.203***
	(0.036)	(0.038)	(0.045)	(0.078)
Chain-linked Δ Log HP (other counties)	0.112***	0.124***	0.115***	0.083***
	(0.037)	(0.034)	(0.037)	(0.031)
County controls	yes	yes	yes	yes
4-Regions FE	no	yes	-	-
9-Divisions FE	no	no	yes	-
State FE	no	no	no	yes
First Stage F-stat	20.746	19.026	13.718	4.782
Observations	922	922	922	922
R-squared	0.062	0.059	0.020	0.043

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retailer chains distribution of stores: $\sum_{k \neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (8) report results, after instrumenting county exposure to house price shocks through the network of retailer chains. In column (3) to (4), I instrument with exposure to WLRI, as in the baseline specification. In columns (5) and (6), I instrument with exposure to Saiz Housing supply elasticity. Columns (7) and (8) use the 4 instruments. Panel A report the second stage coefficients. Robust standard errors clustered at the state level in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

D.1.8 Clientele Effects

It is possible that common regional shocks bear different effects on different type of stores, even within a county. For example, a store of a retail chain that caters to richer consumers might be affected differently compared to a store of a retail chain that caters to poorer consumers. To account for this, I construct a variable at the retail chain level that captures whether the retail chain is mostly present in rich counties or in poor counties, based on county-level median household income.

I define the indicator variable, Clientele, that takes value one if the retail chain is above the median, and 0 otherwise. And then run the main specification at the county-by-store level, but now including County by Clientele fixed effects. Thus, I compare stores within a county, catering to similar demographic groups.

Results are presented in Table A9. The main conclusions remain unchanged.

Table A9: Robustness Check: Clientele effects at the retail chain by county-level

	Dep Variable: Chain's Δ Log Price in county c			
	OLS		IV: WLRI instrument	
	(1)	(2)	(3)	(4)
Panel A: Second Stage				
Store $\Delta \log HP(others)_{rc}$	0.094***	0.079***	0.183***	0.191***
	(0.032)	(0.033)	(0.037)	(0.061)
Panel B: First Stage				
F-statistic			19.932	16.221
Retail chain controls	yes	yes	yes	yes
County FE	yes	-	yes	-
County-Clientele FE	no	yes	no	yes
Observations	3,747	3,747	3,747	3,747
R-squared	0.612	0.738	0.497	0.568
# retailers	84	84	84	84
# counties	910	910	910	910

Clustered standard errors at the state-level in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

D.1.9 Robustness check: Excluding California

I check the robustness of my results to excluding the largest state in the sample, California.

Table A10: Robustness Check: excluding CA

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.047***	0.029**	0.019	0.126***	0.072*	0.078
	(0.014)	(0.013)	(0.014)	(0.034)	(0.041)	(0.050)
Chain-linked Δ Log HP (other counties)		0.077**	0.079***		0.165***	0.174***
		(0.030)	(0.029)		(0.038)	(0.037)
Panel B: First Stage						
F-stat				20.527	16.695	11.933
Observations	881	881	881	881	881	881
County controls	no	no	yes	no	no	yes

This table repeats Table 1, but excluding Caliornia (the largest state) from the analysis. Clustered standard errors at the state-level in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

D.1.10 Robustness check: Trade Flows

Alternatively, I construct a variable that accounts for the trade flows between county c and county k:

$$\text{trade-weighted}\Delta\log(HP)_c^{others} = \sum_{k\neq c}\gamma_{ck}^{trade}\Delta\log(HP)_k^{07-11},$$

where

$$\gamma_{ck}^{trade} = \frac{tradeflow_{ck}}{\sum_k tradeflow_{ck}}$$

where $trade_{ck}$ is the trade flow (exports and imports) between the state where county c is located and the state in which county k is located.[47] Intuitively, the more county c and county k trade with each other and the bigger county k is, the more a shock in county k will affect county c.

I add this term to my main Equation 4.5, estimate it and obtain the following results for the coefficients of interest:

D.1.11 Inference with correlated errors in shift-share design

In a recent paper, Adao et al. (2019) shows that in shift-share designs, regression residuals can be correlated across regions with similar shares, independent of their geographic location. This implies that even though I cluster at the state-level in my main analysis, there could still be remaining issues with the standard errors.

In this section, I follow Adao et al. (2019) methodology to take this into account. In particular, their methodology provides inference methods that are valid under arbitrary cross-regional correlation in the regression residuals. I present results in table A11.

[47] This data was obtained from Bureau of Transportation Statistics, Office of Secretary And Federal Highway Administration U.S. Department of Transportation combining data from the 2012 Commodity Flow Survey(CFS) and other trade data from the Census Bureau.

Table A11: Adao et al. (2019) Robust Standard Errors

	Dep Variable: County's Δ Log Price Index			
	OLS		IV: WLRI instrument	
	(2)	(3)	(5)	(6)
Panel A: Second Stage				
Chain-linked Δ Log HP (other counties)	0.088***	0.089***	0.129***	0.135***
	(0.036)	(0.037)	(0.054)	(0.059)
Panel B: First Stage				
F-stat			16.642	12.756
Observations	910	910	910	910
County controls	no	yes	no	yes
Demeaned by $\Delta LogHP_c$	yes	yes	yes	yes

This table repeats Columns (1), (2), (3) and (4) of Table 1 using shift-share robust standard errors proposed by Adao et al. (2019) to conduct valid inference under arbitrary cross-regional correlation in the residuals. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

D.2 Extensive Margin Adjustments

In the main analysis, we have focused on the propagation of shocks to prices of continuing varieties. However, retail chains might also adjust their extensive margin in response to shocks. For instance, they might close stores or discontinue products.

In this section of the appendix, I explore this possibility. In my previous analysis, I showed that county level prices are sensitive to shocks in distant regions linked by the retail chains' network. This can imply changes in average markups of the county, leading to entry or exit of stores. For example, a county exposed to negative shocks in other county will experience a drop in its retail prices. This, in turn, will make competition tougher in the county. Therefore, we may observe more exit of stores due to tougher competition in counties that were linked to counties more affected by the house price drops.

In order to study the effect on the extensive margin at the county level, I begin by running my main specification with counties' percentage changes in number of stores, barcodes and varieties as dependent variable. I present results in Table A12. Although significant only at 10%, there is some suggestive evidence of exit of stores and barcodes in counties served by retail chains more exposed to drops in house prices in other counties.

Table A12: Retailer chain's extensive margin responses

VARIABLES	(1) $\Delta Stores$	(2) $\Delta Stores$	(3) $\Delta Products$	(4) $\Delta Products$	(5) $\Delta Varieties$	(6) $\Delta Varieties$
County's Δ Log HP	0.024	-0.095	0.271	0.135	0.093	-0.122
	(0.151)	(0.187)	(0.175)	(0.230)	(0.246)	(0.296)
Chain-linked Δ Log HP (other counties)		0.316*		0.361		0.570*
		(0.179)		(0.297)		(0.317)
Observations	910	910	910	910	910	910
R-squared	0.001	0.030	0.006	0.003	0.004	0.026
County controls	yes	yes	yes	yes	yes	yes
First Stage F-stat	13.358	13.862	13.358	13.862	13.358	13.862

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retailer chains distribution of stores: $\sum_{k \neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (8) report results, after instrumenting county exposure to house price shocks through the network of retailer chains. In column (3) to (4), I instrument with exposure to WLRI, as in the baseline specification. In columns (5) and (6), I instrument with exposure to Saiz Housing supply elasticity. Columns (7) and (8) use the 4 instruments. Panel A report the second stage coefficients. Robust standard errors are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

Previous table shows how the Great Recession and the geographic distribution of retailer chain's stores affected the varieties available for consumers. Given that consumers have a taste for variety, an increase in the range of available varieties should lead to a decrease in the price index. Translating the increase in product variety into welfare gains requires structural assumptions. Following Feenstra (1994) and Broda and Weinstein (2006) I assume a CES utility function to infer the inframarginal consumer surplus created or destroyed by changes in product variety from the observed spending shares on new and exiting products. I also decompose entry and exit of barcode-stores into two terms:[48] In particular, I decompose the Feenstra Ratio into two margins of adjustments:

1. Store-level Feenstra Ratio: captures consumer surplus by changes in stores available. 2. Barcode-level Feenstra Ratio: captures consumer surplus created by entry or exit of barcodes in existing stores.

I then construct an exact CES price index that considers entry and exit of new varieties:

$$\left[\Pi_{bs \in I_{mc}} \left(\frac{P_{bsmct}}{P_{bsmt-1}} \right)^{w_{bsmct}} \right] \left[\Pi_{s \in I_{mc}} \left(\frac{\lambda_{smct}}{\lambda_{smct-1}} \right)^{\frac{w_{smct}}{\sigma_s - 1}} \right] \left(\frac{\lambda_{mct}}{\lambda_{mct-1}} \right)^{\frac{1}{\sigma_m - 1}}$$

The first multiplicative term in brackets is the the CES price index for continuing products between t and $t-1$, as described in the body of the paper. We refer to the second term as the barcode-level Feesntra Ratio. The higher the expenditure share of new varieties within a product module and store, the lower λ_{smct}, implying a lower adjusted inflation rate. Intuitively, conditional on the

[48]To do so, I assume that consumers first choose the store in which they will shop and then choose the barcodes.

number of stores, the adjusted inflation rate is going to be lower if consumers spend more in new varieties. The third term is the store-level Feenstra Ratio. The higher the expenditure in new stores, the lower λ_{mct}, implying a lower adjusted inflation rate. The price index also depends on the module-specific elasticity of substitution σ_m between stores and module-store specific elasticity of substitution σ_s between barcodes in a product-module-store. As these elasticities grow, the additional terms converge to one and the bias in the price index goes to zero. The intuition is that when existing varieties are close substitutes to new or disappearing varieties, price changes in the set of existing products perfectly reflect price changes for exiting and new varieties.

Now, I estimate my main specification with each of the terms above as dependent variable. Table A13 report the results. Columnn (1) report results for the price index for continuing varities, as in the main analysis of the paper. Column (2), reports the effect of retailer chains networks on changes in product variety within existing stores. Column (3) report results for the effect on changes in available stores. Finally, column (4) provides the exact price index combining the three terms. We do not observe any effect of the network of retail chains on entry and exit of stores and barcodes. However, in column (4), we observe that the main conclusions of the paper remain when we adjust the price index to account for entry and exit of varieties.

Table A13: County level Price Index adjusted for entry and exit of varieties

	Δ Log Price Index (continuers)	Feenstra Ratio (barcodes)	Feenstra Ratio (stores)	Δ Log Price Index (adjusted)
	(1)	(2)	(3)	(4)
Panel A: Second Stage				
County's Δ Log HP	0.069*	-0.034	0.035	0.071
	(0.040)	(0.025)	(0.034)	(0.067)
Chain-linked Δ Log HP (other counties)	0.119***	0.034	-0.008	0.147**
	(0.039)	(0.034)	(0.033)	(0.071)
Observations	922	922	922	922
County controls	yes	yes	yes	yes
First Stage F-stat	14.479	14.479	14.479	14.479

A unit of observation is a county. I define a chain to be a unique combination of two identifiers in the Nielsen data: parent code and retailer code. The dependent variable is the change in county price index between 2007 and 2011: $\Delta log(HP)_c^{07-11}$. The average of the dependent variable is 11.8%. Chain-linked Δ Log HP (other counties) refers to the weighted average percentage change in house prices in other counties, linked through retailer chains distribution of stores: $\sum_{k \neq c} \omega_{c,k} \Delta log(HP_k)^{07-11}$. Chain-linked Δ Log HP (other counties) is the percentage change in house prices between 2007 and 2011, where the weights are the county share of the chain national sales. Columns (2) to (8) report results, after instrumenting county exposure to house price shocks through the network of retailer chains. In column (3) to (4), I instrument with exposure to WLRI, as in the baseline specification. In columns (5) and (6), I instrument with exposure to Saiz Housing supply elasticity. Columns (7) and (8) use the 4 instruments. Panel A report the second stage coefficients. Robust standard errors are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively. Source: Prices, quantities and sales are obtained from Nielsen Scanner Data. Housing price data is obtained from from the Federal Housing Finance agency (FHFA). Wharton Land Regulation Index from Gyourko et al. (2008a), Housing supply elasticity is obtained from Saiz (2010).

E Model

E.1 National costs and error term

In this section, I discuss the the implication for identification in the model, when costs are at the retail level.

$$dlogP_{ct} = \beta^H \left[\sum_{r \in \Omega_c} \theta_{rc} S_{rc} \right] dlogHP_c + \beta^H \sum_{k \neq c \in \Omega_r} \sum_{r \in \Omega_c} l_{rc} \theta_{rk} dlogHP_k + \sum_{r \in \Omega_c} l_{rc} dlogC_r \tag{E.1}$$

As the cost of retail chain is not observed, it lies in the error term. Hence, the exogeneity assumption to identify β^H is given by,

$$E\Big[\sum_{r \in \Omega_c} l_{rc} dlogC_r \Big| \Big(\sum_{k \neq c \in \Omega_r} \sum_{r \in \Omega_c} l_{rc} \theta_{rk} dlogHP_k \Big) \Big] = 0 \tag{E.2}$$

This simply reflects that, conditional on own house price changes, the evolution of the marginal cost in county c cannot be correlated with the weighted average of the change in house prices in other counties k. For exposition, define $\Delta C_r = \sum_c \lambda_{rc} \Delta A_c$ as a weighted average of unit costs (A_c) across counties in the U.S; where λ_{rc} are the share of costs of retail r that comes from county c (e.g.: wholesale products that buys in c). The threat is common shocks to costs to regions in which the retail chain operates that also correlate with house price changes. Then there are a couple of intuitive *Sufficient conditions* to satisfy Condition E.2.

1. If $Cov(\Delta A_k, \Delta log(HP)_k) = 0$, then Condition E.2 is satisfied. *However, house price shocks could be correlated with productivity shocks in the county.*

2. If $Cov(\lambda_{rc}, \theta_{rc}) = 0$, then Condition E.2 is satisfied. Stroebel and Vavra (2019) present a range of evidence suggesting that the location where retail sell their products (θ_{rk}) differ from the locations where the retail buy their products from wholesales. In line with this, Hyun and Kim (2019), show that most of sales of manufacturing firms come from markets that are not where they have their plants.

In case sufficient conditions (1) and (2) do not hold, I rely on the housing supply elasticity instrument. The assumption then becomes:

$$E\Big[\sum_{r \in \Omega_c} l_{rc} dlogC_r \Big| \Big(\sum_{k \neq c \in \Omega_r} \sum_{r \in \Omega_c} l_{rc} \theta_{rk} WLRI_k \Big) \Big] = 0 \tag{E.3}$$

Note that now a sufficient condition for E.3 to hold is that: $Cov(\Delta A_k, WLRI_k) = 0$. This assumption mimics the identification assumption of papers that analyzed the local effect of the house price slump (e.g: Mian and Sufi (2011), Stroebel and Vavra (2019)). Hence, given that this condition is sufficient (but not necessary), my empirical design relies on milder assumptions.

E.2 Alternative assumptions for costs: local costs

In the main analysis, I solved the model for the case in which the retail chain has national costs $c_{rc} = c_r$. Here, I explore the case in which retail chain has local costs in each market (c_{rc}).

Solving the optimization problem, we get the optimal price in the situation where the retail chains' costs are local:

$$P_{rc} = \bar{p}_r = \frac{\sum_k \mu_k c_{rk}(\sigma_k - 1)S_{rk}}{\sum_k S_{rk}(\sigma_k - 1)} \tag{E.4}$$

where $\mu_k = \frac{\sigma_k}{\sigma_k - 1}$ is the markup in market k. From equation E.4 it is clear that the price under uniform pricing is a weighted average of the conditions (costs and markups) in the different markets that a retail chain serve.

Following the steps in the main section to aggregate the price index and total differentiate the aggregate price index to observe the sources of variation, we get:

$$dlogP^U_{ct} = -\left[\sum_{r\in\Omega_c} l_{rc}\tilde{\theta}_{rc}\psi_{rc}\right] dlog\sigma_c - \sum_{k\neq c\in\Omega_r}\sum_{r\in\Omega_c} l_{rc}\tilde{\theta}_{rk}\psi_{rk} dlog\sigma_k + \sum_{k\in\Omega_r}\sum_{r\in\Omega_c} l_{rc}\frac{S_{rk}\sigma_k c_{rk}}{\sum_k \sigma_k S_{rk} c_{rk}} dlogc_{rk} \tag{E.5}$$

where I define θ as in the main analysis

$$\tilde{\theta}_{rk} = \frac{S_{rk}\sigma_k}{\left[\sum_{k\in\Omega_r} S_{rk}\sigma_k c_{rk}\right]\left[\sum_{k\in\Omega_r} S_{rk}(\sigma_k - 1)\right]},$$

and,

$$\psi_{rc} = c_{rc}\sum_k S_{rk}\left(\sigma_k \frac{c_{rc} - c_{rk}}{c_{rc}} - 1\right)$$

Compared with the case of national costs in the main analysis, now the weights also adjust for initial differences in costs across markets (ψ_{rk}). In addition, local shocks can affect the cost term of the equation.

In the next section, I explore how propagation of cost shocks would look like.

E.3 Propagation of local cost shocks

In order to be consistent with previous papers that emphasize the demand channel of the house price shock, in the main analysis I assumed that the the house price shock affected the elasticity of demand. However, to study propagation of shocks through the network of retail chains, one could be agnostic about whether the shock is demand-driven or cost-driven.

In this section, I discuss theoretically the propagation of local shocks through the network of retail chains. Assume that σ_k is constant and the shocks are to the local costs of the firm. That is $\beta^{H-Costs} = \frac{\partial log c_{rc}}{\partial log HP_c}$. Then, Equation E.6 becomes:

$$dlogP^{U}_{ct} = \beta^{H-Costs} \sum_{r\in\Omega_c} l_{rc} \frac{S_{rc}\sigma_c c_{rc}}{\sum_k \sigma_k S_{rk} c_{rk}} dlogc_{rc} + \beta^{H-Costs} \sum_{k\neq c\in\Omega_r} \sum_{r\in\Omega_c} l_{rc} \frac{S_{rk}\sigma_k c_{rk}}{\sum_k \sigma_k S_{rk} c_{rk}} dlogc_{rk} \tag{E.6}$$

The last term corresponds to the propagation of local shocks through the network of retail chains. Note that the importance of a market is also increasing on S_{rk}. This makes more meaningful the weights ω_{ck} in the reduced form that can be used to proxy weights for both cost shocks and demand shocks.

E.4 Quantitative analysis

Table A14: σ by quartile of population (From Hottman, 2014)

City Size Dist:	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
σ_c	3.9	3.9	4.5	4.8

Table A15: Propagation of local shocks through the network of retails chains (weights adjusted by σ)

	Dep Variable: County's Δ Log Price Index					
	OLS			IV: WLRI instrument		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Second Stage						
County's Δ Log HP	0.056***	0.050***	0.050***	0.125***	0.093**	0.096**
	(0.014)	(0.013)	(0.015)	(0.027)	(0.034)	(0.037)
Chain-linked Δ Log HP (other counties)		0.049***	0.048***		0.116**	0.126**
($\sigma_{adjusted}$)		(0.017)	(0.017)		(0.050)	(0.054)
Panel B: First Stage						
F-stat				21.358	11.660	10.810
Observations	910	910	910	910	910	910
County controls	no	no	yes	no	no	yes

A unit of observation is a county. The dependent variable is the percentage change in the county price index between 2007 and 2011 ($\Delta log(P_c)$). County's Δ Log HP is the percentage change in house prices between 2007 and 2011. Chain-linked Δ Log HP (other counties) is the network-weighted percentage change in house prices in other counties ($\sum_{k\neq c} \theta_{ck} \Delta log(HP_k)$) as defined in Equation 6.2. Columns (1) to (3) report results for OLS estimations. County-level controls in columns (3) and (6) include changes in log wages, changes in log number of retail establishments and changes in log employment. Columns (4) to (6) report results, after instrumenting the two main endogeneous variables. I instrument local percentage change in house prices with local WLRI. I instrument network-weighted percentage change in house prices ($\sum_{k\neq c} \theta_{ck} \Delta log(HP_k)$) in other counties with network-weighted WLRI in other counties ($\sum_{k\neq c} \omega_{ck} WLRI_k$).Panel A report the second stage coefficients. Panel B reports the Kleibergen-Paap (2006) rk F-statistic for the first stage. Robust standard errors clustered at the state level are in parenthesis. ***,**,* denotes statistical significance at the 1, 5, and 10 percent levels, respectively.

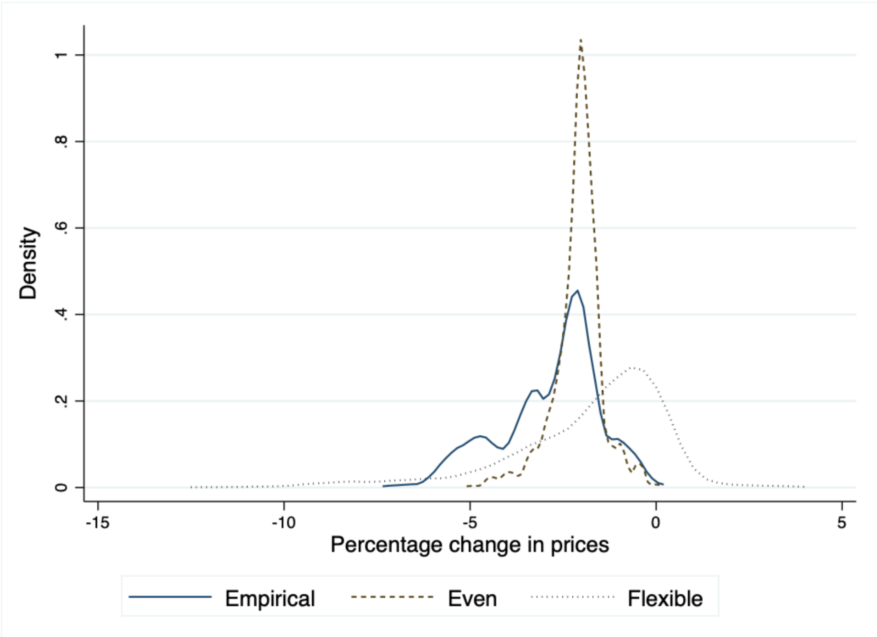
F Counterfactuals

F.1 Cross-county dispersion in inflation rate: Alternative scenarios

Figure F.1 plots the distribution of house price changes under three alternative scenarios. Uniform pricing and empirical distribution of sales (benchmark), Uniform pricing and even distribution of sales, and flexible pricing. We can observe that the dispersion is much smaller under uniform pricing.

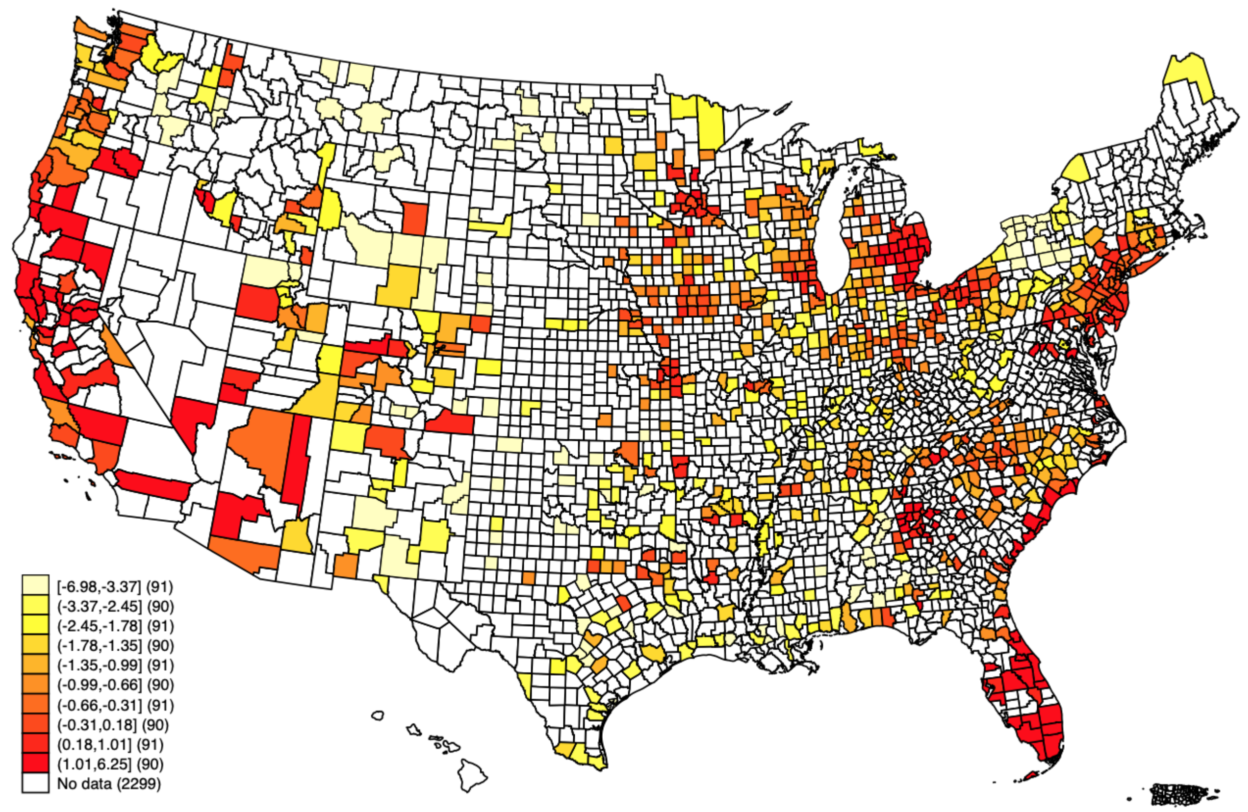
Figure A4: Cross-county dispersion in inflation rate: Alternative scenarios

Figure A5: Densities



F.2 Distributive effects of uniform pricing

Figure A6: Regional Redistribution of shocks: $\Delta \log(P_c^U) - \Delta \log(P_c^{flex})$



Note: Heat map for the difference between inflation rate under uniform pricing and inflation rate under flexible pricing for each county ($\Delta \log(P_c^U) - \Delta \log(P_c^{flex})$). The numbers are in %. The map is divided into deciles of $\Delta \log(P_c^U) - \Delta \log(P_c^{flex})$. Red indicates high values for this difference. That is, counties that were negatively affected by uniform pricing since their prices decrease less than under a counterfactual of flexible pricing.

F.3 A positive shock of 20% to Los Angeles' house prices

In this section, I will be simulating a positive house price shock of 20% in Los Angeles, California. I evaluate how the pricing strategies and the geographic distribution of the stores has implications on the effect of the shock. I choose Los Angeles because it is the county whose shocks have on average the biggest effect on other counties, and at the same time is in the top 20% richest counties in the United States.

Table A16 shows the results for the different pricing strategies. For this shock, I calculate the welfare implications measured as price changes under different price settings: uniform pricing and flexible prices. In particular, I decompose the aggregate effect on U.S. inflation in two different ways. First, In Panel A, I decompose the aggregate inflation into geographic categories: changes in local prices, changes in prices of counties in the same state and changes in prices out of state. Second, in Panel B, I decompose the aggregate inflation into income quintiles. Each county's inflation is weighted by the population of the county, so number represent the contribution of counties in that group to aggregate inflation.

Lets focus first in Panel A of Table A16. While the weighted average inflation rate is similar under uniform and flexible prices, the existence of uniform prices creates redistribution from the county affected by the shock to counties connected by the network of retail chains. In particular, under uniform pricing Los Angeles *export* half of the inflation rate to other counties. An increase of 20% in Los Angeles' house prices implies an increase of 0.14% average inflation rate in the U.S under uniform pricing. 0.07% of that inflation rate is explained by price changes in L.A, 0.05% is explained by propagation to counties outside California and 0.02% is explained by propagation to other counties within California.

When collapsed to five per capita expenditure quintiles on panel B of Table A16, I find that the richer the counties are, the bigger average effect relative to poorer counties. Two effects might play a role here. On the one had, richer counties are more connected than poor counties, so they are more likely affected by a shock somewhere else. On the other hand, retailers in richer counties depend less on other counties to set up their prices given that the sales on their own county are highly relevant. The former effect might be dominating. Moreover, the fact that the effect in weighted by population, and riches counties tend to have higher population density, contributes to getting a higher average effect for richest quintiles.

Table A16: 20% shock in Los Angeles under alternative pricing rules

Panel A: Effect by Geographic Location		
Welfare \ Price Choice	Uniform	Flexible
Δ in Prices(average)	0.1440%	0.1368%
Δ in Local Prices	0.0732%	0.1368%
Δ Prices in-state	0.0215%	0.0000%
Δ Prices out-of-state	0.0494%	0.0000%
Panel B: Effect by Income Quintile		
Welfare \ Price Choice	Uniform	Flexible
Δ in Prices(average)	0.1440%	0.1368%
Δ in Local Prices	0.0732%	0.1368%
Δ in Poorest Quintile	0.0023%	0.0000%
Δ in Second Poorest Quintile	0.0035%	0.0000%
Δ in Middle Quintile	0.0090%	0.0000%
Δ in Second Richest Quintile	0.0246%	0.0000%
Δ in Richest Quintile	0.0315%	0.0000%

Note: Effects in each county are weighted by population. As Los Angeles is on the richest quintile, we subtract the local average effect from the effect on the richest quintile.

In Table A17, I report results for alternative distributions of the stores of the retail chains. The first column is the effect on national inflation rate under the current distribution of retail chains. Column (2) analyzes an increase in retail chains' geographic dispersion of sales. The effect of the shock becomes smaller at the national level. Intuitively, relative to the current distribution, retail chains become smaller in big counties; while relatively larger in small counties. As S_{rk} are equal, retail chains in L.A. export the inflation rate in equal proportions to every county. Hence,the population weighted average inflation rate for the U.S. declines considerably. In the third column, we increase competition within a county by letting all retail chains in the county have equal shares.

The national effect of the shock is slightly smaller. This evidences that the key for the propagation of shocks is the importance of each county for the retail chain. The forth column contains both equal distribution in counties and equal importance of counties for a chain. This situation is closely to full integration and implies that the effect of a shock in L.A would impact 60% less the average inflation rate in the U.S. Finally, the last column the case of a monopoly in each of the counties. For each county, I pick the retail chain with the highest sales in the county and assign $l_{rc} = 1$ to it and zero to other retail chains in the county. The effect average national effect increases. A potential explanation for this is that when the biggest retailers in the country behave as monopolists in their counties, the propagation of the shocks increases the most. The prices are controlled by a few chains that are most likely present in Los Angeles and a lot of other counties around the United States.

Table A17: 20% shock in Los Angeles under alternative distributions

Welfare \ Distribution	**Current**	**Competition: Even Distribution**			**Monopoly**
		$S_{rk} = 1/\#c$	$L_{rk} = 1/\#r$	Both	$L_{rk} = 1$
Δ in Prices(average)	0.144%	0.027%	0.122%	0.063%	0.154%
Δ in Local Prices	0.073%	0.019%	0.085%	0.049%	0.119%
Δ Prices in-state	0.021%	0.004%	0.021%	0.011%	0.014%
Δ Prices out-of-state	0.049%	0.004%	0.015%	0.004%	0.209%