House Prices, Quantities, and the Propagation of Shocks when Developers Are Imperfectly Competitive

Jack Favilukis, Elena S. Pikulina, Tianping Wu*

March 25, 2026

Abstract

We create a novel dataset identifying the quantity, location, and timing of home sales by national and local U.S. developers in order to study concentration in housing construction markets. Larger markets and more regulated markets are more concentrated, suggesting the importance of fixed entry or operating costs. Concentration matters, with more concentrated CBSAs having higher prices and lower quantities. Negative price shocks originating in one CBSA cause developers to shrink or exit other markets, thus propagating through the CBSA-developer network, reducing competition and raising prices in connected CBSAs. These patterns provide new evidence on how imperfect competition in housing construction and financial frictions propagate local housing shocks across space.

Keywords: housing construction, competition, regulation, internal capital markets

*All authors are affiliated with the Sauder School of Business, University of British Columbia. E-mails: jack.favilukis@@sauder.ubc.ca; elena.pikulina@sauder.ubc.ca; tianping.wu@sauder.ubc.ca.

1 Introduction

Housing affordability is one of the most salient issues facing households today. When thinking about housing supply, much of the academic research has focused on land use regulations and construction costs. For example, Goolsbee and Syverson (2023) document a large decline in construction sector productivity and Baum-Snow and Duranton (2025)'s survey of the literature focuses on the importance of supply constraints. Yet, relatively little is known about the competitive structure of the construction industry. Our goal is to fill this gap.

We begin by constructing a novel dataset which identifies the quantity, location, timing, and seller name of new homes sold in the U.S. To do so, we search the ATTOM housing transaction data for names of large and well known developers, as well as any other sellers whose name indicates they are likely to be a developer. Using this dataset, we construct the Herfindahl-Hirschman Index (HHI) for all U.S. CBSAs where data is available to explore whether the housing construction market is competitive. We also construct a CBSA-developer network to study how shocks transmit across connected CBSAs. We document multiple new and interesting facts about the U.S. construction industry, as well as a more nuanced view on some of the previously known facts. Overall, our findings suggest that competition is far from perfect and that concentration leads to higher prices and lower supply.

First, the strongest covariate with concentration, measured by HHI, is a CBSA's population – all else equal, more populated CBSAs are more competitive. This is unlikely to simply be due to larger markets consisting of segmented sub-markets because individual developers tend to have operations across large swaths of even the biggest CBSAs. It is consistent with the importance of fixed costs, either entry or operating.

Second, concentration is strongly positively related to land use regulation, measured by

the Wharton Residential Land Use Regulation Index (WRLURI), with exactions,[1] approval delays, and affordable housing playing the strongest role. Again, this is consistent with the importance of fixed costs. The positive association between WRLURI and concentration is entirely driven by smaller developers suggesting they are less capable than national developers at overcoming these fixed costs. Thus, our results imply that regulatory constraints are twofold setback for affordability – not only do they increase costs directly, they also reduce competition, further increasing costs.

Third, we investigate the increasing trend in concentration that has been documented previously by Ahluwalia et al. (2022) and Quintero (2023). We show that since 2011, the equal weighted average CBSA has not experienced an increase in concentration (as measured by HHI), but the population weighted average CBSA has, thus concentration has grown in larger and faster growing CBSAs. Much of the increase is explained by the increase in concentration within the set of larger national developers; concentration within smaller local developers has not changed, though their market share has fallen somewhat. Although mergers are positively related to the rise in concentration, there are too few of them to explain much of the rise. On the other hand, net exits explain a significant fraction of the increase in HHI in the cross-section. This is consistent with the time series evidence as much of the rise in HHI and the increase in exits happened in the aftermath of the housing bust in 2008.

Fourth, we show that places with more concentrated developer markets have higher prices and lower quantities, and that this relationship is likely to be causal. In OLS regressions, a one standard deviation increase in a CBSA's concentration is associated with a 0.26 standard deviation increase in its housing price and a 0.09 standard deviation decrease in the number of new units built. While the OLS results may not be causal, we propose an instrument based on the number of other CBSAs nearby. The intuition is that CBSA locations were determined

[1]Exactions are requirements that developers pay fees or provide public amenities.

long ago and are likely unrelated to current supply and demand conditions. However, a CBSA with more CBSAs nearby increases the pool of developers willing to commute to this CBSA for jobs. With this instrument, all three of our price measures remain positively and statistically related to HHI, all three of our quantity measures remain negatively related to HHI, though only one is statistically significant.[2]

Fifth, we build a CBSA-developer network, with CBSAs being the nodes, connected by overlapping developers. We show that shocks transmit across this network and are very persistent, with competition, prices, and quantities affected years later. Consider a developer who operates in Los Angeles and Dallas and a second developer who operates in Los Angeles and Las Vegas. Las Vegas experienced a severe fall in house prices following the housing bust in 2008, but Dallas did not.[3] We show that the developer who operates in Los Angeles and Las Vegas cuts construction in Los Angeles by far more than the developer who operates in Los Angeles and Dallas. This is despite both developers facing the same demand conditions in Los Angeles. This is consistent with the importance of internal capital markets, with cash flow shocks originating in one segment affecting investment in a firm's other segments. We then aggregate this to the CBSA level and show that CBSAs less exposed (through the developer network) to places like Dallas and more exposed to places like Las Vegas in the period 2005-2010 saw significantly larger increases in concentration post-2010, and saw significantly larger price growth post-2010. This is consistent with competition being imperfect and reinforces our earlier results.

Literature review Our paper's main focus is on competition in the housing construction industry. There has been a renewed academic, journalistic, and policy interest in declining

[2]We measure prices as either the price of new homes sold by developers in our data, the average price in the CBSA from the American Community Survey, or the average rent in the CBSA from the American Community Survey. We measure quantity as either the number of units sold by developers in our data, the lot area of units sold by developers in our data, or the total number of housing units in the CBSA from the U.S. Census.

[3]Between 2005 and 2010 – the window over which we define the shock – the Las Vegas all-transactoins house price index fell by 47.5%, while the Dallas index rose by 7%.

competition, with Autor et al. (2017), Covarrubias et al. (2020), and De Loecker et al. (2020) among the prominent academic examples. However, relatively little has been written about competition in housing construction – we conjecture this is due to limited availability of data. The two most closely related papers to our work are Kroft et al. (2025) and Quintero (2023).[4]

Kroft et al. (2025) study imperfect competition in both the product and labor market of the construction sector with a focus on how the two interact. They find that construction firms have market power in both markets. However, their methodology is different, with our focus on CBSAs and theirs on firms. First, Kroft et al. (2025) use changes in wages for firms who win government procurement auctions to identify labor market power. Second, they extend the methodology of De Loecker et al. (2020) to identify construction firms' market power and markups in the product market by combining the estimates of labor market power with firm-level information on labor share and intermediate input share of revenues. In contrast, we focus on CBSA-level concentration measures and relate them to CBSA-level prices and quantities.

Our work shares most in common with Quintero (2023), who uses proprietary data collected by the data provider Metrostudy. Like us, Metrostudy uses transactions to identify builders, however their sample covers just parts of the U.S. Northeast for the period 2005-2016.[5] In contrast, our data spans the period 2000-2023 and covers the vast majority of the U.S. population.[6] The longer and wider sample allows us to draw more general and more

[4]There is also an older literature studying the industrial organization of the construction industry, these studies include Herzog (1963), Cassimatis (1969), Grebler (1973a), Grebler (1973b), Markusen and Scheffman (1977), Eccles (1981), Quigley (1982), Landis (1986), Carroll (1988), and Somerville (1999). However, we believe that the significant improvements in quality and availability of data since then as well as the economic and political importance of housing affordability makes renewed focus on this topic worthwhile.

[5]Specifically, the sample covers places in northeastern metros (Allentown, Atlantic City, Baltimore, Dover, Philadelphia, Salisbury, Trenton, Vineland, Washington DC, Winchester, and parts of New York but not New York City), which includes 137 local housing markets with a total population of 9.3 million.

[6]For the year 2000, we are able to compute CBSA level HHI for CBSAs with 190 million people, equivalent to 67% of the U.S. population; starting in 2005, our coverage increases to above 89% of the U.S. population.

nuanced conclusions. Quintero (2023) finds that more concentrated CBSAs have lower volume, fewer units, and greater price volatility; we find that more concentrated CBSAs have fewer units but also higher prices and rents. Quintero (2023) finds that concentration has increased over time and conjectures that this is due to mergers and bankruptcies after 2008, as well as tax-loss carry back provisions in the 2009 stimulus favoring larger builders. We show that after 2009 there has been no increase in concentration for the average CBSA or within the set of smaller local builders, thus the increase can be fully attributed to increasing concentration in larger CBSAs and among larger, national builders. We also show that mergers played a minor role in the increase, while exits after 2008 and organic growth by the bigger builders were much more important. Additionally, we show that larger and less regulated places are the most competitive, highlighting the importance of fixed costs; we also build a CBSA-developer network and show that shocks propagate through this network, with long term effects on competition, prices, and investment – these findings are new to the literature.

Because of similarities in methodology, studies by van Straelen (2023) and Kim (2025) are also closely related, though they do not focus on competition or investment. Like us, both use transaction data to identify developers. Both also use shocks in one market where a developer operates to study the effect on the developer's behavior in other markets. However, both are interested in the developer's price setting behavior rather than competition. van Straelen (2023) shows that during the 2008 housing bust, individual developers who suffered negative price shocks in one market cut prices in other markets in order to raise cash, especially if they were financially constrained. Kim (2025) shows that well diversified and financially unconstrained developers were less likely to cut prices and structurally estimates the cost of financial constraints. The price cuts studied are shorter term effects for units already under construction. In contrast, we show that following negative price shocks in a market where the developer operates, the same developer cuts investment in non-shocked markets and that

this effect lasts for many years. Furthermore, we show that these non-shocked markets see long term increases in developer concentration and house prices.

D'Amico et al. (2024) study the decline in productivity in the construction sector. They argue that larger firms are more productive, but that regulation favors smaller firms, with increased regulation leading to a productivity decline. They find that in the cross-section, stronger regulation is associated with lower receipts for the average firm and lower fraction of employment in larger firms. At first blush, this appears to contradict our finding that regulation leads to more concentration. However, the two are not necessarily at odds because their result is driven largely by subcontractors who specialize in certain jobs but do not build entire homes, while ours is driven by firms who build and sell entire homes and who may hire these subcontractors.[7] Since our focus is on product market competition, concentration among firms who actually build and sell entire homes is the more relevant quantity.[8]

Several papers have studied spillovers in housing markets. The most relevant paper for our work is DeFusco et al. (2018) which documents regional spillovers in house prices – during the housing boom of the early 2000s, housing markets experiencing a boom made it more likely that nearby markets would also experience a boom; they suggest a role for non-rational factors. We also show that shocks spill over, but as a result of the developer network; since nearby housing markets are likely to share developers, our results offer a rational explanation for their findings. Campbell et al. (2011) and Benmelech et al. (2018) also document spillovers in real estate markets, but at a far shorter geographical scale. Bailey et al. (2018) show that

[7]D'Amico et al. (2024) show that firms with 500+ employees are responsible for just 4% of employment but 23% of units built, while firms with 0-4 employees are responsible for 40% of employment but just 7% of units built. They explain that smaller firms are often subcontractors who specialize in plumbing, drywall, etc., but do not build entire homes. We conjecture that stronger regulation reduces the size of these small firms, or increases the fraction of work that large firms subcontract to small firms, consistent with their findings. At the same time stronger regulation may reduce the number of small or midsize general contractors leading to higher concentration within the set of firms who build and sell entire homes, consistent with our findings. Indeed, we find that the positive relationship between regulation and concentration is mostly driven by the dearth of smaller local developers in more regulated places.

[8]However, D'Amico et al. (2024)'s finding of reduced concentration among subcontractors may give large firms more labor market power, which is the focus of Kroft et al. (2025).

when someone's friends on a social network experience housing appreciation in a far away market, that person becomes more likely to purchase a home.

Our data suggests that the spillovers we document are caused by firms' reliance on internal capital markets, leading to propagation of shocks through the developer network. If firms heavily rely on internal capital markets, then positive (negative) cash flow shocks to one segment will increase (reduce) investment to other segments of a multi-segment firm. This may be efficient – if external financing costs are very high – or inefficient – when firms practice corporate socialism as in Scharfstein and Stein (2000). There is an extensive literature documenting that internal capital markets matter and that shocks originating in one location or segment can have effects on investment (Lamont (1997), Shin and Stulz (1998)), employment (Giroud and Mueller (2019)), loan supply (Peek and Rosengren (2000)), and pricing decisions (Ge (2022), van Straelen (2023), Kim (2025)). Our empirical result showing that developers experiencing negative shocks in one market cut investment in other markets contributes to this literature.

2 Data

2.1 Constructing the Developer Transactions Dataset

To carry out our analysis, we construct a dataset of new home sales and link each sale to the name of a developer in a CBSA. Equipped with this dataset, we can then construct an HHI for each CBSA. Because some developers operate across multiple CBSAs, we can also use this measure to create a CBSA-developer network. At a high level, we use transaction data to identify all transactions whose seller is a developer and within these transactions, we group them by year and developer.

To construct our dataset we use ATTOM data, which is similar to the CoreLogic and

Zillow datasets widely used by academics. ATTOM contains data on millions of housing transactions in the U.S. This data is collected from public records such as deeds and the vast majority of transactions are single family homes. Though some of the transactions are from the 1970s, the coverage is relatively poor until the late 1990s. For each transaction, ATTOM reports the location, price, names of the buyer and seller, as well as other information. We identify the transaction as developer transaction – equivalently a new home sale – if the seller's name matches one of a list of 737 names of large national developers from Builder Magazine or one of their subsidiaries, or if the seller's name contains one of a list of keywords including "build", "homes", "constr", and "develop" – we define the latter group as local developers. We are also careful to exclude many false positives, for example a seller named "Homer Simpson" would otherwise be identified as a developer. Appendix section A provides additional details on this process.

Our final dataset contains 10,495,677 transactions between the years 2000 and 2023, of which 53% are by national developers. The top panel of Figure 1 shows the number of transactions in each year we identified to be made by developers and compares it to new units of single family homes sold for the entire U.S. from the U.S. Census. The second panel does the same but restricts the transactions only to ones by large national developers, comparing them to closings by the same developers provided by Builder Magazine. The third panel shows a ratio of our transactions relative to the Census, and our national transactions relative to Builder Magazine. It shows that we identify about 30% of total transactions in 2000, which rises to 75% by 2007 and remains fairly flat after this; the match rate for large national builders is fairly similar to the overall match rate. ATTOM does not cover all CBSAs in the U.S., for example in 2000, CBSAs with transactions in ATTOM contained 67% of U.S. population, while by 2023 this share rose to 89%. Therefore within the covered CBSAs, our match rate is likely higher than the one implied by this figure. The bottom panel compares the average price in our data to the average price of new homes from the

Census – the two co-move very closely.

We also confirm that our coverage does not appear to be biased in the cross-section. For each CBSA, we compute total housing permits from 2000 to 2023 (from the Census), scaled by the number of units in 2000; we compare it to each CBSA's total new home sales identified by us in ATTOM data from 2000 to 2023, scaled by units in 2000. The correlation between the two is 0.89 and regressing our transactions on permits leads to an R^2 of 0.79; adding various controls to the regression only increases the R^2 to 0.81, with most being statistically insignificant. Thus, in the cross-section, our measure of new units is highly correlated with new construction in the Census and the differences are small and not highly related to CBSA characteristics.

Each transaction contains the name of the seller and the location of the property. Additionally, many but not all transactions report the sales price and various home characteristics are reported in the linked tax assessor data, though we do not use this data in most of our analysis. We focus on units sold and construct an HHI for each CBSA i in year t based on the number of units sold by each developer operating in that CBSA. Specifically, we compute:

$$HHI_{i,t} = \sum_{j} \left(\frac{u_{j,i,t}}{\sum_{k} u_{k,i,t}} \right)^2 \tag{1}$$

for CBSA i where developer j sells $u_{j,i}$ units. This is our key variable of interest. We require that a CBSA has at least five developer transactions in a year for us to compute its HHI that year. For the 925 CBSA in our 24 year sample (22,200 data points), we have 14,275 observations where we can compute HHI. Most of our analysis focuses on larger CBSAs with at least 100,000 residents in the year 2000, there are 372 such CBSAs (8928 data points over 24 years), for these CBSAs we have 7,744 observations where we can compute HHI. 182 CBSAs have HHI available in every year, of these 168 are in the large group.

Since our focus is on competition, mergers are an important consideration. For example in 2009, Pulte acquired Centex, merging two of the largest developers in the U.S. If Centex were to immediately begin selling under the Pulte name, there would be no need to adjust our approach. However, the Centex brand name remains to this day and transactions may have Pulte or Centex as the name of the seller. To identify mergers, we read through the financial statements, websites, and Wikipedia pages of the largest 106 firms (including all public firms) in our data to identify the names of subsidiaries and the timing of any merger and acquisition activity. We identify 58 mergers or acquisitions by large firms, with the firms being acquired accounting for 0.6% of all large firm housing transactions in the year of acquisition. After the acquisition, all transactions with the acquired firm in the name are assigned to the acquirer, for example all sales by Centex were assigned to Pulte starting in 2009.

It is also useful to decompose HHI into three components, an HHI computed only within national builders operating in i, and HHI computed only within local builders operating in i, and the national builder market share. These three components relate to HHI in the following way: $HHI = (1-s)^2 HHI_{loc} + s^2 HHI_{nat}$ where s is the national builder's share.

We also use transactions by developers to compute each CBSA's average price of new built homes, total number of new built homes, and the total lot area of new built homes – these will be among our outcome variables of interest.

2.2 Additional Data

We collect annual house prices and rents at the county level from IPUMS and aggregate them to CBSA level for the period 2000-2023.[9] We collect data on annual population and income at

[9]IPUMS aggregates and makes it easy to use multiple public datasets, the data we use comes from the Census and the American Community Survey (ACS). County data is available in 2000, as well as for larger counties annually starting in 2005 and for smaller counties in ACS 5 year samples. We use the ACS 5 year samples to impute (approximate) annual rents and prices for the smaller counties. We interpolate between

the county level from BEA and data on housing units from Census county files. We aggregate them to CBSA level for 2000-2023. We also collect data on the average temperature in each CBSA and each CBSA's land area. Finally, we use data on regulatory constraints from the WRLURI, collected through surveys by Gyourko et al. (2008) and Gyourko et al. (2021). This index uses survey data to estimate the severity of various types of land use regulations at the local level. It contains twelve subcomponents (e.g. local zoning, supply restriction, exaction, and affordable housing) and also the weighted average of the subcomponents. Data on prices, rents, and housing units will be among our outcome variables of interest, while the other variables are used as controls.

3 Empirical Analysis

3.1 Descriptive Statistics of HHI

The average HHI across all CBSAs and years is 0.158 with a standard deviation of 0.136; this is analogous to just six equal sized developers in a CBSA. However, most of our analysis is restricted to CBSAs which had a population of at least 100,000 people in the year 2000. These are somewhat more competitive, with a mean HHI of 0.106 and a standard deviation of 0.092. The strongest covariate with HHI is a CBSA's population, with more populous CBSAs having a more competitive construction sector with significantly smaller HHI. Figure 2 illustrates the univariate relationship between a CBSA's logarithm of average population and its average HHI for the period 2000-2023. The first column of Table 1 regresses the average HHI in a CBSA over our sample on the CBSA's logarithm of the average population, the logarithm of average income, the average temperature, and the logarithm of the land

2000 and 2005 to get missing county level prices and rents for 2001-2004. We then aggregate counties to CBSAs.[10] Finally, for 378 larger CBSAs, we use Freddy Mac house price growth to replace the 2001-2004 interpolated prices by more precise estimates. Since in our main results, we focus on larger CBSAs, most of these interpolations and imputations do not appear in our sample.

area. HHI is strongly negatively associated with population, with a t-statistic of -5.53. HHI is also significantly higher in warmer places with a t-statistic of 3.07, while income and land area are not significantly related to HHI.

There are two obvious potential explanations for why HHI may be lower in more populous CBSAs. First, the relationship may be spurious: a populous CBSA may simply be a collection of segmented sub-markets for housing construction, with each sub-market having a higher HHI but the measured HHI when combining sub-markets appears lower. For example, suppose that Orange and Los Angeles counties, which make up the Los Angeles CBSA, are segmented and each have six developers operating there. Then the HHI of Orange and Los Angeles counties would be 0.17, but the measured HHI of Los Angeles CBSA – an irrelevant quantity if these markets are segmented – would be 0.08. We argue that construction markets within larger CBSA are unlikely to be too segmented.

While it is true that developers are more likely to develop near where they already develop, the geographical distance they are willing to cover is significant relative to the size of the CBSA. Appendix Figure A.1 shows that the spread in distance between a developer's projects within a CBSA is increasing in the CBSAs population (top two panels) and land area (middle two panels). The bottom two panels show that for the average developer, the spread is between 25% and 50% of the CBSA's diameter, however for some developers, this spread is comparable to the size of the CBSA. For example, in Los Angeles – which is one of the larger CBSAs (either by population or land area) with nearly 100 miles separating San Clemente in the south from San Fernando in the north – the average national (local) developer has projects 39 (13) miles apart, while some national and local developers have projects more than 90 miles apart.

Appendix Figure A.2 shows that developers are also willing to travel some distance outside of their CBSAs for projects – this is important for our instrumental variable, which we discuss in Section 3.4. More than 6% of a CBSA's local developers also operate in *other*

CBSAs that are a distance of 30 miles from the original CBSA and 2% of local developers operate in *other* CBSAs 90 miles away. Not surprisingly, national developers are willing to cover much larger distances. More than 25% of a CBSA's national developers operate in *other* CBSAs that are a distance of 30 miles from the original CBSA and 18% of national developers operate in *other* CBSAs 90 miles away. Appendix Section B provides additional discussion of this point.

Second, a negative relationship between CBSA size and HHI is consistent with a standard model of imperfect competition and fixed operating or entry costs. If product market competition is imperfect, than holding the number of competitors constant, operating in a larger market is more profitable because developers can charge the same markup but increase volume. Since national developers choose which market to enter, holding constant entry or fixed operating costs, more developers will enter larger markets until expected profits are equalized. Similarly, since local households choose whether to become developers, holding constant entry or fixed operating costs, more local households will become developers in larger markets. On the other hand, holding market size fixed, more developers will enter lower cost markets.

We now provide additional evidence for the importance of fixed entry or operating costs. To do so, we utilize the 2018 version of the WRLURI, developed by Gyourko et al. (2021). Since many of these regulations reduce profits either by restricting what a developer can do (e.g. minimum lot sizes or inclusionary zoning) or by requiring additional effort or cost from the developer (e.g. approval delays and exaction fees), these regulations create an additional mix of variable and fixed costs for developers. Column fourteen of Table 1 reestimates the relationship between HHI and MSA characteristics but now includes the weighted average of the regulations, provided by Gyourko et al. (2021). Inclusion of WRLURI only strengthens the negative association of HHI with size, but WRLURI itself is strongly positively associated

with HHI – more regulated places have fewer developers and are less competitive.[11] Columns two through thirteen of Table 1 use the individual sub-components of the index instead; most are positive but the three that stand out are exactions, affordable housing, approval delays.[12]

Appendix Tables A.1, A.2, and A.3 repeat this analysis but separately for the three sub-components of HHI – the HHI of local developers within a CBSA, the HHI of national developers within a CBSA, and the national share within a CBSA. These tables show that the positive relationship between HHI and regulation is mostly driven by the HHI of local developers. The share of development done by national developers is also positively related to the weighted average of WRLURI, though the relationship is not as strong as for local developer HHI. Two sub-components stand out – exactions and approval delays are especially important for increasing the share of national developers and decreasing competition within local developers. Thus, these two frictions are especially costly for small local developers, leading to fewer local developers operating in regulated CBSAs. With fewer local developers, market share is taken by national developers, who are better able to navigate regulation. On the other hand, the weighted average WRLURI is positively but insignificantly related to the HHI of national developers. The only sub-component of WRLURI that is statistically significantly related to the HHI of national developers is affordable housing – perhaps this is a regulation national developers find more difficult to avoid.

[11]We have also explored the relationship between HHI and the Saiz (2010) supply elasticity. We find the relationship to be negative – inelastic places have higher HHI – but not statistically significant, with a correlation of -10%. Because the Saiz (2010) measure is available for only 95 CBSAs, we have chosen to not report these results or include this measure as a control.

[12]In unreported results we have also used the 2006 of the WRLURI index (Gyourko et al. (2008)). The results are similar and the weighted average index is statistically significant, though the significance is lower than using the 2018 version, especially for some of the sub-indices. This may be because the 2018 version is more comprehensive and therefore less noisy.

3.2 Trend in HHI

In this section, we investigate the time trend in concentration. Quintero (2023) studies concentration for the period 2005-2016 using data covering parts of the U.S. Northeast region and finds that concentration has increased.[13] He conjectures this is due to exits following the 2008 financial crisis and a subsequent waive of mergers. Similarly, Ahluwalia et al. (2022) use the reported number of transactions from Builder Magazine to show that the share of new home sales by the top two builders has increased from 6.2% in 2002 to 15.2% in 2020 (35.5% to 49.3% for the top 100 builders).

In our larger sample, we find a similar general trend although the conclusion is more nuanced. The top panel of Figure 3 plots a time series of HHI averaged across all U.S. CBSAs for which we are able to compute HHI in each year, this includes 189 CBSAs covering approximately 68% of the U.S. population. Consistent with Quintero (2023) and Ahluwalia et al. (2022), we find a large rise in concentration following the financial crisis, with HHI going from 0.09 in 2007 to 0.14 in 2011. However, after 2011, HHI declines for the average CBSA to 0.125 in 2023. On the other hand, a population weighted average HHI rises almost monotonically from 0.05 in 2005 to 0.08 in 2023. Thus, HHI did rise from the point of view of the average buyer, but it rose due to increased concentration in larger and faster growing markets.

In the next three panels of Figure 3, we separate HHI into its three components: national HHI, local HHI, and national share. In the time series, nearly all of the growth in population weighted HHI since 2005 can be attributed to growth in national HHI, furthermore, much of that growth happened between 2007 and 2012. The HHI of local developers has actually fallen over time and the total market share of national developers has fluctuated significantly,

[13] His data covers the metros of Philadelphia, Washington DC, Baltimore, Allentown, Atlantic City, Dover, Salisbury, Trenton, Winchester, Vineland, and parts of New York. These markets have a population of 9.3 million people.

but is only modestly higher in 2023 compared to 2005. Thus, larger concentration within the set of national developers is responsible for much of the change in HHI.

In the first three columns of Table 2 we estimate a panel regression of the one year change in HHI on the one year change in either national HHI, local HHI, or the national share. In the cross-section, all three measures are positively associated with local HHI being the strongest.[14]

We next explore whether mergers or exits may be responsible for the rise in developer HHI. The bottom panel of Figure 3 plots the fraction of all transactions affected by mergers and by exits. Entry was very high leading up to the financial crisis, exits then spiked in 2008 leading to negative net entry until turning positive again in 2012. Mergers – at least ones we are able to identify – were also high around 2009, then again in 2018. However, the fraction of transactions affected by mergers – around 4% at its peak and near zero most years – was far lower than the variation in net entry.

We next regress the annual change in HHI on the part of HHI that can be attributed to mergers,[15] on net exits, and on the change in HHI of incumbent firms that neither merged, entered, or exited. Column five shows that mergers are statistically significantly related to change in HHI, but the R^2 is tiny. Thus, mergers along are unlikely to be responsible for most of the HHI growth. Conditional on a merger occurring, the explanatory power is much greater, but relatively few transactions are affected by mergers.[16] An important caveat is that our list of mergers may be incomplete.

[14]Recall that these three measures can be combined in a non-linear way to exactly compute HHI. In the four column we include all three measures in a multi-variate regression, the only reason the R^2 is not 100% is linearity.

[15]Specifically, for each year and CBSA, we compute $HHI_{i,t}^{nomerg}$ based on the seller's ownership at $t-1$. For example, Pulte and Centex, which merged in 2009, are grouped together when we compute $HHI_{i,2009}$, but are treated as separate firms when we compute $HHI_{i,2009}^{nomerg}$. $HHI_{i,2009} - HHI_{i,2009}^{nomerg}$ represents the part of HHI in 2009 explained by merger activity between 2008 and 2009.

[16]The correlation between $HHI_{t,i} - HHI_{t-1,i}$ and $HHI_{t,i} - HHI_{t,i}^{nomerg}$ is 25% when conditioning on a CBSA-year observation having merger activity, it rises to 54% if we also restrict the sample to larger CBSAs. However, the correlation is just 2% if we do not condition on a merger occurring, and 7% for larger CBSAs.

In column six, we regress the change in HHI on net exits.[17] Net exits explain 16% of the variation in HHI growth. In column seven, we regress the change in HHI on the change in HHI computed only among incumbents. This measure rises when large incumbent firms grow and small incumbent firms shrink and it explains 33% of HHI growth. Finally, in column eight we include mergers, exits, and incumbent HHI. We conclude that the primary contributor to HHI was growth by firms already in the market, with exit and entry also contributing significantly (especially following the financial crisis), but that mergers played a minor role.

3.3 HHI, Prices, and Quantities

In this section, we investigate the relationship between the concentration of developers in a CBSA, and house prices and quantities in that CBSA. We begin with OLS analysis and discuss potential endogeneity in Sections 3.4 and 3.5. We restrict our analysis to CBSAs with a population of at least 100,000 in the year 2000.

Table 4 shows the results of cross-sectional OLS regressions of various measures of a CBSA's price or quantity of housing on its HHI and controls, which include log income, log population, average temperature, and log land area. The level of observation is a CBSA, with prices, quantities, HHI, income, and population computed by taking the annual values, scaling by the aggregate quantity that year, and then averaging over all available years. Depending on the variable of interest, there are between 308 and 319 CBSAs in the sample.

The top panel presents results for three measures of prices: the logarithm of the average price of homes sold by developers identified from our transaction data, the average price using IPUMS/ACS data, and the average rent using IPUMS/ACS data. HHI is positively

[17] We define exits as the total transactions in a CBSA at $t-1$ by firms that did not operate in that CBSA at t, scaled by the maximum of total units sold in the CBSA between t and $t-1$. We define entries as the total transaction in a CBSA at t by firms that did not operate in that CBSA at $t-1$, scaled by the maximum of total units sold in the CBSA between t and $t-1$. Net exits is the difference between the two.

and significantly associated with all three measures of prices, with t-statistics between 5.15 and 5.85 in columns two, five, and eight. A one standard deviation increase in HHI is associated with a 0.26 standard deviation increase in the average CBSA price and a 0.22 standard deviation increase in rent.[18]

The bottom panel presents results for three measures of quantities: the logarithm of the number of new units sold identified from our transaction data, the total lot area of new units sold from our transaction data, and the total number of housing units from the U.S. Census. HHI is negatively and significantly associated with all three measures of housing quantities, with t-statistics between -2.56 and -4.94 in columns two, five, and eight. A one standard deviation increase in HHI is associated with a 0.09 standard deviation increase in the number of new units built.[19]

Table 5 presents regressions similar to Table 4 but uses unbalanced panel data rather than a single cross-section. The level of observation is a CBSA-year with an average of between 263 CBSAs and 284 observations per year in the sample, depending on the specification. We regress a CBSA's log price or log rent or log housing units that year on HHI and controls that year; we also include year fixed effects. These results are similar to the simple cross-sectional regressions in Table 4, with t-statistics of between 4.78 and 5.08 for prices and between 2.29 and 6.32 for quantities.

[18]In these regressions, price, rent and HHI are all scaled by the average price and HHI. Scaled HHI has a standard deviation of 0.68, equivalent to going from a mean unscaled HHI of 0.1 to 0.168. Thus, a one standard deviation increase in scaled HHI is associated with a $0.68 \times 0.15 = 0.10$ increase in log scaled price, which has a standard deviation of 0.39; 0.15 is the coefficient on HHI in column five of Panel A. Similarly, a one standard deviation increase in scaled HHI is associated with a $0.68 \times 0.064 = 0.044$ increase in log scaled rent, which has a standard deviation of 0.2; 0.064 is the coefficient on HHI in column eight of Panel A.

[19]Scaled HHI has a standard deviation of 0.68. Thus a one standard deviation increase in scaled HHI is associated with a $0.68 \times 0.185 = 0.126$ increase in log scaled rent, which has a standard deviation of 1.36; 0.185 is the coefficient on HHI in column two of Panel B.

3.4 Instrumenting for HHI

While the results in Section 3.3 suggest that market power may be important, it is possible that these results are affected by endogeneity or omitted variables. It is useful to think through how these problems might arise. Suppose markets are frictionless and competitive but developers anticipate which CBSAs will have highest demand and shift operations there. Then such CBSAs will appear to be more competitive and have lower HHIs, they will have more supply as construction responds to higher demand. However, if costs are convex, they will have higher prices as supply cannot fully meet the demand and if the marginal cost is constant then there should be no effect on prices. Other stories in which markets are competitive but in which entry, exit, or M&A respond to demand would also imply prices and quantities moving in the same direction. These kinds of stories are not consistent with the OLS results above and cannot explain our findings.

However, suppose certain CBSAs have high demand and also high costs, for example if CBSAs in desirable locations also have more regulations or more difficult to build geographies. Suppose our controls (e.g. average temperature and land area) were not sufficient to fully capture these differences in costs and demand. Then, these CBSAs may have higher prices and lower quantities even after controls. Finally, suppose that it was also the case that fewer developers choose to operate in such CBSAs but for reasons unrelated to market power. Then HHI will be positively associated with prices and negatively with quantities, even though market power plays no role.

In this section, we propose an instrument for HHI, while in Section 3.5 we explore causality by using the propagation of shocks across the CBSA-developer network. Earlier we showed that fixed entry costs appear to be important for determining the number of developers in a CBSA. We posit that these entry costs are lower for developers who are already operating in nearby CBSAs. Therefore, CBSAs with other CBSAs nearby should be more

competitive because developers established in nearby CBSAs are willing to commute to these CBSAs for jobs. Indeed, Table A.2 shows that for both national and local developers, there is significant overlap in developers between nearby CBSAs and this overlap decreases in distance. At the same time, the geographical distribution of CBSAs was determined long ago and should not affect prices or quantities, other than through developer competition.

Table 3 shows the results of the instrumental variable first stage, where we regress each CBSAs average HHI across all years on log income, log population, average temperature, log land area, and on the instrument. For each CBSA, the instrument counts the number of CBSAs (excluding the focal CBSA) within 50 miles of the focal CBSA. Consistent with earlier results in Table 1, population and temperature are significantly related to HHI, while income and land area are not. The instrument is negatively related to HHI, with a t-statistic of 3.58, therefore CBSAs with many other CBSAs nearby tend to have more competitive developer markets. The R^2 rises from 0.150 without the instrument to 0.183 with the instrument and the F-statistic is 12.79. We compute the fitted value of HHI for each CBSA using the coefficients from this regression.

Columns (3), (6), and (9) of Tables 4 (cross-section) and 5 (unbalanced panel with year fixed effects) present the second stage results for each of the six outcome variables. These regressions are identical to the previously presented OLS results, except HHI is replaced by its fitted value from the first stage. Like with OLS, prices and rents are positively and significantly related to instrumented HHI; the three measures of quantities are negatively related to instrumented HHI, though only one of the measures is statistically significant. These results suggest that the correlation found in OLS regressions is not spurious and that concentration leads to fewer housing units and higher housing prices.

3.5 The CBSA-Developer Network

In this section, we construct a CBSA-developer network. We show that if a developer operates in multiple CBSAs, house price shocks originating in one CBSA affect the developer's output in other CBSAs. Aggregating this up, we also show that house price shocks originating in one CBSA transmit to other CBSAs when they share developers. We then show that shocks originating in one CBSA affect concentration, prices, and output in connected CBSAs.

The intuition for why shocks should transmit can be understood by thinking about financial frictions and internal capital markets. If firms heavily rely on internal capital markets, then positive (negative) cash flow shocks to one segment will increase (reduce) investment to other segments of a multi-segment firm. These kinds of internal transfers have been documented empirically by Shin and Stulz (1998) and Lamont (1997) who show that investment flows even to unproductive segments.

Suppose there are n CBSAs and let m_c denote the set of developers who operate in CBSA c. Define $u_{c,i,t}$ to be the units built by developer i in CBSA c during the period t, where t may span multiple years. Developer i's exposure to CBSA s is defined as $\xi_{s,i,t} = \frac{u_{s,i,t}}{\sum\limits_{k=1,n} u_{k,i,t}}$, this is the fraction of i's total units that are built in CBSA s. We define the exposure of CBSA c to CBSA s as $\phi_{c,s,t} = \frac{\sum\limits_{i \in m_c} u_{c,i,t}\xi_{s,i,t}}{\sum\limits_{i \in m_c} u_{c,i,t}}$, this is the average exposure of all developers to CBSA s, weighted by the development they do in CBSA c.[20]

We next compute the shock to developer i operating in CBSA c that is coming from price changes $\Delta P_{s,t+k}$ in other CBSAs. We define this shock as $\psi_{c,i,t} = \sum\limits_{s \neq c} \xi_{s,i,t} \Delta P_{s,t+k}$. This is the average price change across all CBSAs, excluding CBSA c, weighted by the fraction of

[20]For example, suppose developer 1 builds one unit in CBSA A and one unit in CBSA B, while developer 2 builds one unit in CBSA A and two units in CBSA C. Then $\xi_{A,1} = 0.5$, $\xi_{B,1} = 0.5$, $\xi_{C,1} = 0$, $\xi_{A,2} = 0.33$, $\xi_{B,2} = 0$, $\xi_{C,2} = 0.67$; $\phi_{A,A} = 0.5 \times 0.5 + 0.5 \times 0.33 = 0.42$, $\phi_{A,B} = 0.5 \times 0.5 + 0.5 \times 0.0 = 0.25$, and $\phi_{A,C} = 0.5 \times 0.0 + 0.5 \times 0.67 = 0.33$. In practice, because all local developers have an exposure of one to their one CBSA and of zero to other CBSAs, $\phi_{c,s}$ is typically large, with the average $\phi_{c,s}$ being 0.87 across all CBSAs in our sample, and 0.80 for CBSAs with a population of at least 100,000 in the year 2000.

developer i's total construction in each CBSA. Note that $\sum_{s \neq c} \xi_{s,i,t} \leq 1$, therefore a developer who mostly operates in CBSA c will have $\psi_{c,i,t}$ near zero, while developers with significant operations in CBSAs with big price increases (decreases) will have $\psi_{c,i,t}$ that is large in magnitude and positive (negative).

We also compute the shock to CBSA c coming from price changes in other CBSAs as $\psi_{c,t} = \frac{\sum_{s \neq c} \phi_{c,s,t} \Delta P_{s,t+k}}{\sum_{s=1,n} \phi_{c,s,t}}$. This is the average price change across all CBSAs, excluding CBSA c, weighted by the connection between c and each CBSA. Importantly, CBSA c's own price change is excluded in the numerator, but the weight is included in the denominator, therefore if a CBSA has small but positive exposure to many other CBSAs, but most of its developers build locally, its overall exposure to other CBSAs will be near zero. Intuitively, ψ_c is positive (negative) if many of the developers who build in CBSA c also develop a large fraction of their units in CBSAs with positive (negative) price shocks.

Although we can compute this quantity for any time period, we focus on the developer-CBSA network in 2000-2005, just prior to the start of the financial crisis and compute $\phi_{c,s,2000-2005}$. For example, Los Angeles had an exposure to itself of 0.404, and its next five top connections were Riverside (0.066), Phoenix (0.063), Las Vegas (0.058), Dallas (0.034), and San Diego (0.031). We define the price shock as $\Delta P_{s,2010} = log(P_{s,2010}/P_{s,2005})$ to coincide with the large house price drops many CBSAs experienced during the financial crisis. Thus, the timing of the shock is that the exposures ξ and ϕ are computed based on the network structure between 2000 and 2005, while the shock is computed based on price changes between 2005 and 2010.

We begin by analyzing developers. Within each CBSA c, we separate national developers into two groups: those who received a shock $\psi_{c,i,t}$ above versus below the median. Developers with a below (above) median shock are ones who experienced relatively low (high) price growth in markets outside of c, we refer to these as negatively (positively) shocked developers. For each year, we sum up total units built by each group in the focal CBSA c, we then

aggregate across all focal CBSAs (weighting by population in the year 2000). We plot this time trend in units built (relative to the 2005 level) in Figure 4. Negatively shocked national developers are represented by the red dotted line and positively shocked by the blue dashed line; we also include local developers in this plot, although they are irrelevant for this analysis.

Prior to the start of the price shock in 2005, both types of national developers had nearly identical growth in the number of units built, which quadrupled between 2000 and 2005 as the U.S. housing market was booming. After 2005, as the housing bubble began to pop, the two lines diverge, with negatively shocked national developers cutting development in the focal CBSA by significantly more. By 2010, negatively shocked developers saw their development fall by 83% relative to 2005 levels compared to a fall of just 68% for positively shocked developers. This difference was quite persistent, with developments by negatively shocked national developers only catching up to the positively shocked developers in 2023. One may be concerned that the price shock to the focal CBSA may somehow be responsible for these results, however note that we are comparing negatively to positively shocked developers *within* the same CBSA, thus the two types of national developers receive exactly the same shock in the CBSA whose units are included in the figure.[21] This finding is a housing market analog of Lamont (1997) and Shin and Stulz (1998).

We next move onto CBSA level analysis. In a competitive construction market, even if negatively shocked developers pull out, other developers will simply expand leaving prices and quantities unchanged. On the other hand, if the market is imprefectly competitive, a reduction in output by negatively shocked developers will increase market concentration, and raise prices or lower quantities. In Table 6, we show that the data is consistent with an imperfectly competitive market.

Because annual changes are quite noisy, we split our data in half and focus on the change

[21] In unreported results, we have also split CBSAs into ones with low, medium, and high own price growth and redrawn the figure for each group. The results look similar to Figure 4, though they are noisier since each plot has fewer CBSAs included.

in CBSA c's prices, quantities, and HHI from the period 2000-2011 to the period 2012-2023 as a function of its realized shock $\psi_{c,2005-2010}$. This shock is computed based on the developer network in 2000-2005, and the realized price growth in CBSAs connected to the focal CBSA c in 2005-2010. We refer to CBSAs with a low (high) value of $\psi_{c,2005-2010}$ as negatively (positively) shocked CBSAs. For each quantity of interest, we compute the average in the first and second periods: $y_{2000-2011}$ and $y_{2012-2023}$. We then regress the growth from 2000-2011 to 2012-2023 ($log(y_{c,2012-2023}/y_{c,2000-2011})$ for prices and quantities and $y_{c,2012-2023} - y_{c,2000-2011}$ for HHI) on $\phi_{c,2005-2010}$ and various controls. One may be concerned that due to regional correlations, the shock $\phi_{c,2005-2010}$ is correlated with fundamentals in CBSA c even though it does not include the price growth in c in its computation. For this reason, it is important to include the price growth in CBSA c over the same period as the shock, $log(P_{c,2010}/P_{c,2005})$ as one of the controls. This controls for changes in c's fundamentals that may affect prices and quantities in c. We also include as controls each CBSA's log of average price and income in 2000-2011, the growth from 2000 to 2011 of price and income, average HHI in 2000-2011, and the average of the y variable in 2000-2011 (e.g. when the dependent variable is growth in units, we include log of average units in 2000-2011 as a control).

Column eight in Panel A of Table 6 shows that the shock, which is realized by 2010 affects HHI between 2012 and 2023 with a negative and significant sign. A CBSA c is negatively shocked if it is connected through the developer network to other CBSAs with negative price shocks. As discussed earlier and shown in Figure 4, after a negative shock, affected developers reduce activity in c. This reduction in activity leads to more concentration in c and an increase in HHI. The other columns in Panel A show that a negative shock realized in 2010 positively and significantly affect house prices and rents in 2012-2023. As concentration rises, the developers in c exercise more market power and charge higher prices. Panel B presents analogous results for quantities, we do not find a statistically significant relationship between the shock realized in 2010 and quantities in 2012-2023.

Thus, we have shown that prices shocks originating in other CBSAs and not connected to the fundamentals of the focal CBSA lead to less development in the focal CBSA by developers exposed to such shocks, less competition in the focal CBSA, and higher house prices in the focal CBSA.

4 Conclusion

This paper studies the industrial organization of housing development. We show that housing developers are relatively concentrated, with HHI around 0.15 for medium size metros, and around 0.06 for the largest metros. Concentration is stronger in smaller and more regulated places, suggesting the importance of fixed costs. Concentration matters for affordability – more concentrated markets have higher prices and lower housing supply. Using a CBSA-developer network, we find that shocks to concentration are persistent, with developers suffering price shocks in one market significantly cutting construction in unrelated markets; this leads to higher concentration and lower prices for years after the shock.

References

Ahluwalia, Gopal, Kermit Baker, and Kent W. Colton, 2022, Concentration in the home-building industry: Trends, strategies, and prospects, Working Paper / Technical Report –, Joint Center for Housing Studies, Harvard University.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, 2017, Concentrating on the fall of the labor share, *American Economic Review* 107, 180–85.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2018, The economic effects of social networks: Evidence from the housing market, *Journal of Political Economy* 126, 2224–2276.

Baum-Snow, Nathaniel, and Gilles Duranton, 2025, Chapter 6 - housing supply and housing affordability, in Dave Donaldson, and Stephen J. Redding, eds., *Handbook of Regional and Urban Economics*, volume 6 of *Handbook of Regional and Urban Economics*, 353–461 (Elsevier).

Benmelech, Efraim, Nittai Bergman, Anna Milanez, and Vladimir Mukharlyamov, 2018, The agglomeration of bankruptcy, *The Review of Financial Studies* 32, 2541–2586.

Campbell, John Y., Stefano Giglio, and Parag Pathak, 2011, Forced sales and house prices, *American Economic Review* 101, 2108–31.

Carroll, Barbara Wake, 1988, Market concentration in a geographically segmented market: Housebuilding in ontario, 1978–1984, *Canadian Public Policy* 14, 295–306, Housebuilding in Ontario, 1978–1984.

Cassimatis, P. J., 1969, *Economics of the Construction Industry* (The Conference Board, New York, NY), Studies in Business Economics, Report No. 111.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon, 2020, From good to bad concentration? us industries over the past 30 years, *NBER Macroeconomics Annual* 34, 1–46.

D'Amico, Leonardo, Edward Glaeser, Joseph Gyourko, William Kerr, and Giacomo Ponzetto, 2024, Why has construction productivity stagnated? The role of land-use regulation, *Working Paper 33188, National Bureau of Economic Research* .

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger, 2020, The rise of market power and the macroeconomic implications*, *The Quarterly Journal of Economics* 135, 561–644.

DeFusco, Anthony, Wenjie Ding, Fernando Ferreira, and Joseph Gyourko, 2018, The role of price spillovers in the American housing boom, *Journal of Urban Economics* 108, 72–84.

Eccles, Robert G., 1981, The quasifirm in the construction industry, *Journal of Economic Behavior & Organization* 2, 335–357.

Ge, Shan, 2022, How do financial constraints affect product pricing? Evidence from weather and life insurance premiums, *The Journal of Finance* 77, 449–503.

Giroud, Xavier, and Holger M. Mueller, 2019, Firms' internal networks and local economic shocks, *American Economic Review* 109, 3617–49.

Goolsbee, Austan, and Chad Syverson, 2023, The strange and awful path of productivity in the U.S. construction sector, Working Paper 30845, National Bureau of Economic Research.

Grebler, Leo, 1973a, Growing scale of real estate firms and projects, *Real Estate Economics* 1, 107–122.

Grebler, Leo, 1973b, *Large Scale Housing and Real Estate Firms: Analysis of a New Business Enterprise*, UCLA Housing, Real Estate and Urban Land Studies Program (Praeger, New York).

Gyourko, Joseph, Jonathan S. Hartley, and Jacob Krimmel, 2021, The local residential land use regulatory environment across U.S. housing markets: Evidence from a new Wharton index, *Journal of Urban Economics* 124.

Gyourko, Joseph, Albert Saiz, and Anita Summersl, 2008, A new measure of the local regulatory environment for housing markets: The Wharton residential land use regulatory index, *Urban Studies* 693–729.

Herzog, John P., 1963, The dynamics of large-scale housebuilding, Technical Report Research Report 22, Real Estate Research Program, Institute of Business and Economic Research, University of California, Berkeley, CA.

Kim, Seohee, 2025, Financial frictions and geographical diversification of national homebuilders, *Working paper, Duke University* .

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler, 2025, Imperfect competition and rents in labor and product markets: The case of the construction industry, *American Economic Review* 115, 2926–69.

Lamont, Owen, 1997, Cash flow and investment: Evidence from internal capital markets, *The Journal of Finance* 52, 83–109.

Landis, John D., 1986, Land regulation and the price of new housing lessons from three california cities, *Journal of the American Planning Association* 52, 9–21.

Markusen, J. R., and David T. Scheffman, 1977, *Speculation and Monopoly in Urban De-*

velopment: Analytical Foundations with Evidence for Toronto, Ontario Economic Council Research Studies, Heritage Series (University of Toronto Press).

Peek, Joe, and Eric S. Rosengren, 2000, Collateral damage: Effects of the Japanese bank crisis on real activity in the United States, *American Economic Review* 90, 30–45.

Quigley, John M., 1982, Residential construction and public policy: A progress report, Technical report, Fisher Center for Real Estate and Urban Economics, University of California.

Quintero, Luis, 2023, Fewer players, fewer homes: Concentration and the new dynamics of housing supply, *Working paper, Johns Hopkins University* .

Saiz, Albert, 2010, The geographic determinants of housing supply, *The Quarterly Journal of Economics* 125, 1253–1296.

Scharfstein, David S., and Jeremy C. Stein, 2000, The dark side of internal capital markets: Divisional rent-seeking and inefficient investment, *The Journal of Finance* 55, 2537–2564.

Shin, Hyun-Han, and René M. Stulz, 1998, Are internal capital markets efficient?, *The Quarterly Journal of Economics* 113, 531–552.

Somerville, C. Tsuriel, 1999, The industrial organization of housing supply: Market activity, land supply and the size of homebuilder firms, *Real Estate Economics* 27, 669–694.

van Straelen, Eileen, 2023, Desperate house sellers: Distress among developers, *The Review of Financial Studies* 37, 802–836.

Figure 1: Comparison of Our Data to Aggregate Data

The top panel plots the number of new single-family homes (in thousands) sold per year in the U.S. from the Census (dashed red line) and the number of ATTOM transactions in which we identify the seller as a developer (solid blue line). The second panel plots the number of closings by national developers reported in Builder magazine (dashed red line) and the number of ATTOM transactions in which we identify the seller as a national developer from the Builder magazine list (solid blue line). The third panel shows the corresponding ratios: our developer transactions relative to all Census transactions (solid blue line) and our national-developer transactions relative to all Builder magazine transactions (dashed red line). The bottom panel plots the average transaction price for developer sales in our data (solid blue line) and the average price for new units sold from the Census (dashed red line).

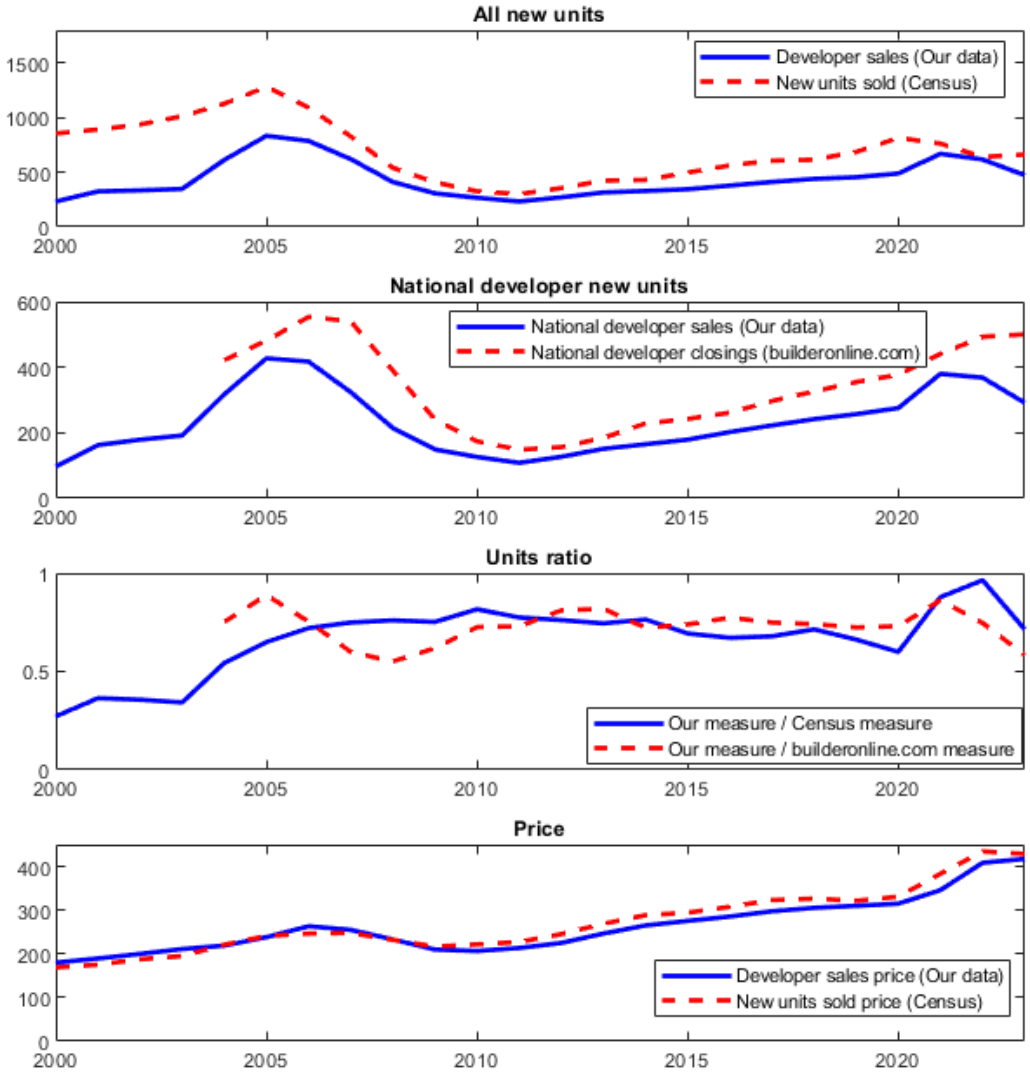


Figure 2: CBSA population and HHI

This figure presents a scatter plot of the logarithm of a CBSA's population in 2000 against its HHI. We separate CBSAs with populations below 100,000 residents (dots) and those above 100,000 residents (hollow circles) because, in most of our analysis, we only include CBSAs in the latter group.

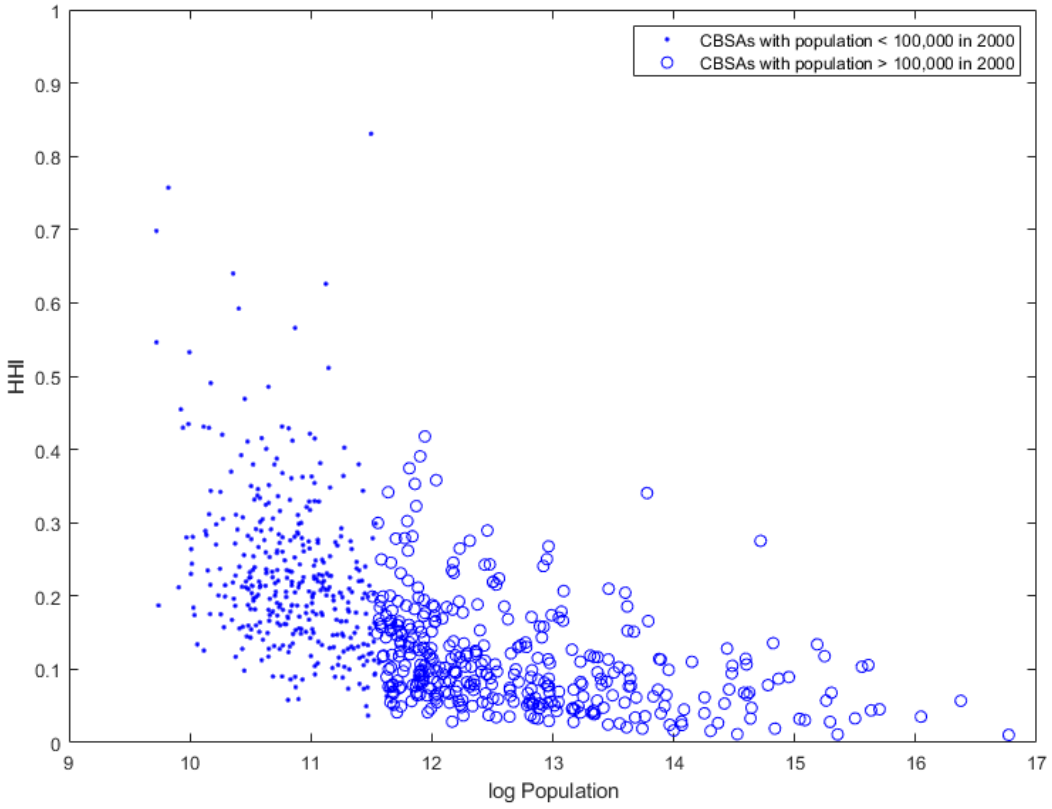


Figure 3: Competition Over Time

This figure plots time trends in several variables of interest. The top four panels plot HHI and its three sub-components: HHI of national developers, HHI of local developers, and the market share of national developers. Each is measured at the CBSA level, then aggregated up to national level either by simple average (solid blue line) or population weighted average (dashed red line). We include only the 186 CBSAs for which we can compute HHI in every single year between 2000 and 2023, these cover 67% of the U.S. population in the year 2000. The bottom panel plots each year's number of units sold by merged firms as a fraction of all units sold (solid blue line), and the number of units sold by firms that enter minus firms that leave as a fraction of total units sold (dashed red line). To compute units by firms that leave, we sum up all the units in $t-1$ sold by firms that do not operate in a market at t; for firms that enter, we sum up all the units sold in t by firms that do not operate in a market at $t-1$; we divide each of these by the maximum of total units sold at $t-1$ or t in that market.

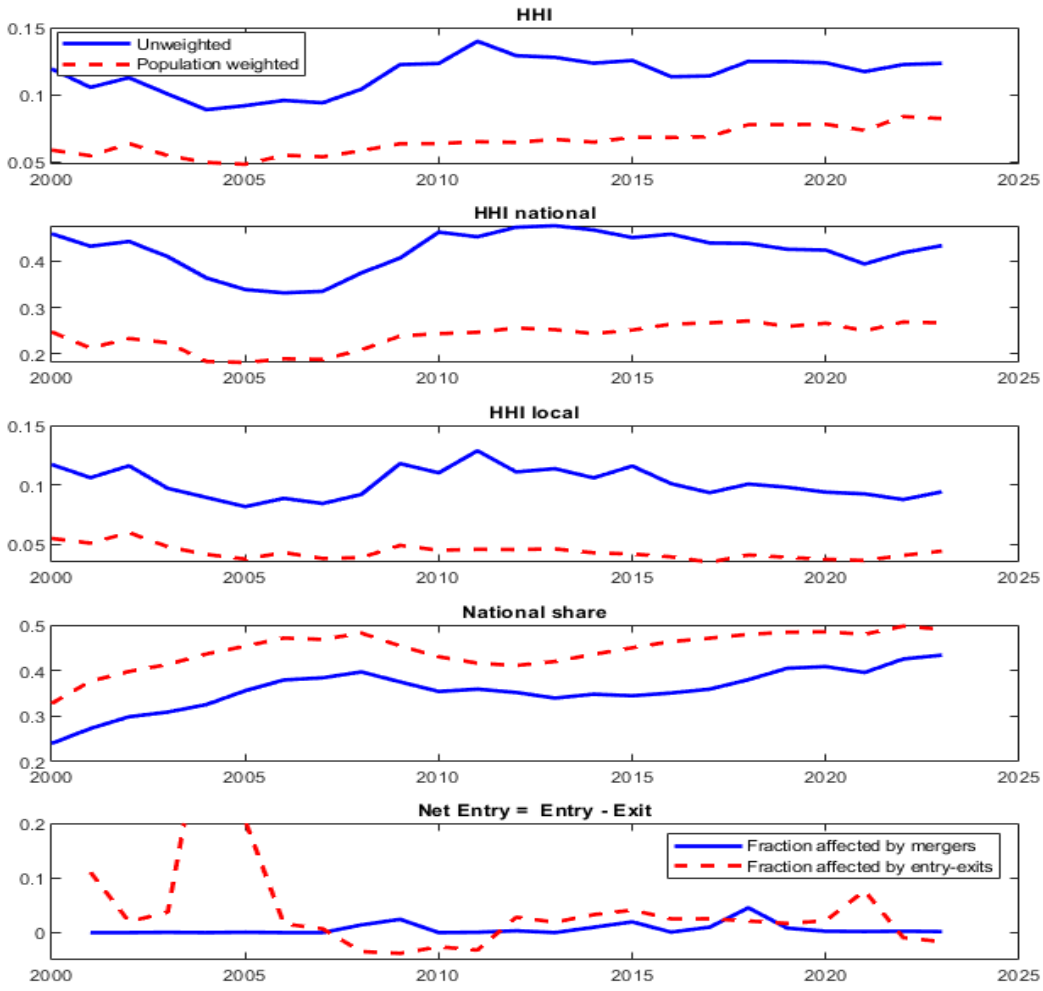


Figure 4: Developer Output in Focal CBSA After Price Shock in Other CBSAs

This figure plots a time series of units sold by three groups of developers, with each series scaled by its 2005 level. The solid black line represents local developers and is constructed by summing units sold by local developers in each CBSA and year, then taking a population-weighted average across CBSAs. The dashed blue line represents national developers who received an above-median shock $\psi_{c,i}$, where the shock is defined in Section 3.5. For each CBSA c and year t, we sum up the units sold in c by national developers whose shock from other CBSAs where they develop ($\psi_{c,i}$) was above the median of all national developers operating in c. We then take a population weighted average across all CBSAs. The dotted red line is computed analogously but includes national developers whose shock is below the median.

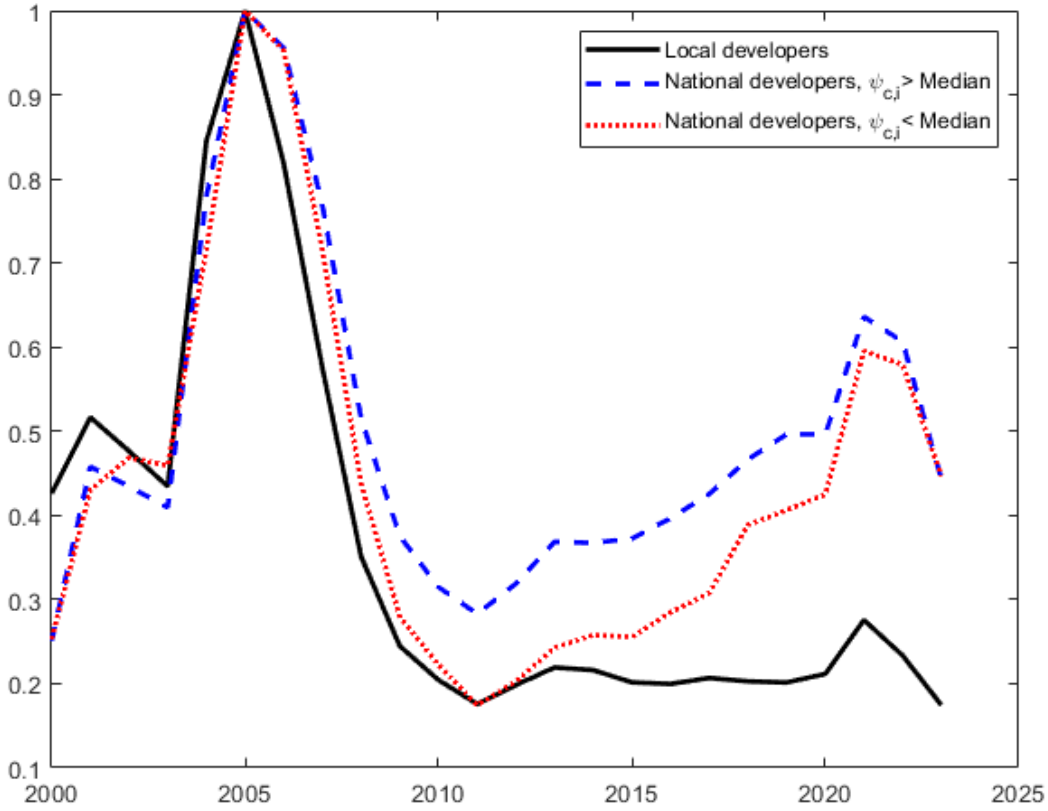


Table 1: HHI Covariates

This table presents cross-sectional regressions of the HHI of the developers' market share within an CBSA on CBSA characteristics. The characteristics include log population, log income, average temperature, log land area, and components of the 2018 WRLURI index. HHI, population, and income are averaged over the period 2000-2023. Column (1) presents results without including WRLURI. The following columns include the WRLURI components one at a time with the components being: (2) political pressure, (3) political involvement, (4) court involvement, (5) local approval, (6) local zoning, (7) local assembly, (8) supply restrictions, (9) density restrictions, (10) open space, (11) exaction, (12) affordable housing, (13) approval delay, and (14) weighted average. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pop	-0.274	-0.270	-0.269	-0.275	-0.274	-0.267	-0.269	-0.273	-0.266	-0.295	-0.281	-0.272	-0.313	-0.297
	(-5.526)	(-5.518)	(-5.484)	(-5.604)	(-5.576)	(-5.454)	(-5.468)	(-5.583)	(-5.412)	(-6.145)	(-5.877)	(-5.831)	(-6.579)	(-6.276)
Inc	0.366	0.210	0.296	0.296	0.309	0.318	0.300	0.272	0.343	0.236	0.243	-0.023	0.050	0.096
	(1.478)	(0.863)	(1.249)	(1.251)	(1.305)	(1.346)	(1.251)	(1.148)	(1.437)	(1.023)	(1.052)	(-0.100)	(0.216)	(0.414)
Temp	0.015	0.008	0.008	0.009	0.009	0.009	0.009	0.010	0.009	0.010	0.007	0.009	0.011	0.007
	(3.057)	(1.650)	(1.699)	(1.844)	(1.939)	(1.938)	(1.840)	(1.969)	(1.900)	(2.042)	(1.373)	(1.920)	(2.296)	(1.595)
Land Area	-0.027	0.002	0.004	0.009	0.022	0.013	0.017	0.026	0.011	0.011	0.000	0.032	0.038	0.003
	(-0.501)	(0.028)	(0.068)	(0.162)	(0.403)	(0.237)	(0.298)	(0.474)	(0.198)	(0.200)	(0.000)	(0.610)	(0.710)	(0.061)
WRLURI		0.045	0.065	0.045	0.031	0.056	-0.036	0.130	-0.039	0.387	0.353	0.755	0.060	0.227
		(1.728)	(1.502)	(1.390)	(1.114)	(1.455)	(-0.353)	(1.781)	(-0.987)	(4.016)	(3.861)	(5.244)	(5.090)	(4.908)
R^2	0.150	0.165	0.163	0.162	0.159	0.162	0.156	0.166	0.158	0.206	0.202	0.238	0.234	0.229
CBSAs	319	258	258	258	258	258	258	258	258	258	258	258	258	258

Table 2: What Explains Changes in HHI?

This table regresses the change in HHI on several components of HHI. The regressions use panel data, with each observation being a CBSA-year. Columns 1–3 regress the change in HHI on the changes in the national and local components of HHI and on the national share; column 4 includes all three. Columns 5–7 regress the change in HHI on the part of HHI explained by mergers, on the fraction of sold homes affected by exit and entry, and on the change in HHI of incumbent builders; column 8 includes all three. The text provides definitions of all variables used in these regressions. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔHHI^{nat}	0.07			0.06				
	(20.50)			(28.82)				
ΔHHI^{loc}		0.60		0.55				
		(78.12)		(85.45)				
ΔNat. Share			0.28	0.35				
			(38.18)	(69.94)				
Merger					1.22			0.86
					(5.67)			(5.53)
Net exit						0.16		0.16
						(37.43)		(46.71)
ΔHHI^{incum}							0.58	0.58
							(60.44)	(68.19)
R^2	0.05	0.46	0.18	0.64	0.00	0.16	0.33	0.49
Obs.	317.61	317.61	288.30	288.30	317.61	317.61	317.61	317.61

Table 3: Instrumental Variable First Stage

This table presents the results of the first stage of our instrumental variable approach. The instrument is the number of other CBSAs within a 50 mile radius of the focal CBSA. We regress each CBSA's average HHI on CBSA characteristics and the instrument. The CBSA characteristics are log income, log population, average temperature, and log land area. HHI, population, and income are averaged over the period 2000-2023. t-statistics are reported in parentheses.

Inc	Pop	Temp	Land Area	IV	R^2
0.35	-0.26	0.01	-0.03		0.15
(1.48)	(-5.53)	(3.06)	(-0.50)		
0.19	-0.20	0.01	-0.16	-0.07	0.18
(0.82)	(-4.16)	(1.77)	(-2.55)	(-3.58)	F=12.79

Table 4: Market Power, Prices, and Quantities, Single Cross-Section

This table presents CBSA level cross-sectional regressions of several measures of housing prices and quantities on HHI and CBSA characteristics. The CBSA characteristics are log income, log population, average temperature, and log land area. HHI, population, income, and all dependent variables are averaged over the period 2000-2023. The top panel presents log prices, with the average price of newly built homes from our data in columns 1-3, the average price from the ACS in columns 4-6, and the average rent from the ACS in columns 7-9. The bottom panel presents log quantities, with the number of newly built homes from our data in columns 1-3, the total lot area of newly built homes in columns 4-6, and the total number of units from the U.S. Census in columns 7-9. We present both OLS and IV regressions. The IV is a fitted value of HHI where the first stage is presented in Table 3 and the instrument is the number of other CBSAs within a 50 mile radius of the focal CBSA. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS	OLS	2SLS
	Panel A: Prices regressed on HHI								
	y =Price^{New}			y =Price^{ACS}			y =Rent^{ACS}		
HHI		0.15	0.42		0.15	0.68		0.06	0.34
		(5.15)	(2.80)		(5.85)	(5.11)		(5.29)	(5.51)
Inc	1.24	1.19	1.09	1.31	1.26	1.07	0.69	0.67	0.57
	(9.96)	(9.88)	(8.16)	(11.42)	(11.48)	(8.96)	(13.06)	(13.13)	(10.45)
Pop	0.06	0.10	0.17	0.02	0.06	0.20	0.03	0.05	0.12
	(2.47)	(4.03)	(3.71)	(0.87)	(2.61)	(4.79)	(3.25)	(4.81)	(6.47)
Temp	0.00	0.00	0.00	0.00	0.00	-0.01	0.01	0.01	0.00
	(1.27)	(0.49)	(-0.88)	(0.61)	(-0.36)	(-2.87)	(5.93)	(5.19)	(1.06)
Land Area	0.01	0.01	0.02	0.03	0.04	0.05	-0.01	-0.01	0.00
	(0.28)	(0.39)	(0.67)	(1.29)	(1.53)	(2.05)	(-0.70)	(-0.58)	(0.06)
R^2	0.39	0.44	0.40	0.40	0.46	0.45	0.52	0.56	0.56
Obs./Year	308	308	308	319	319	319	319	319	319
	Panel B: Quantities regressed on HHI								
	y =Units^{New}			y =LotArea^{New}			y =Units^{Census}		
HHI		-0.18	-0.38		-0.32	-0.88		-0.03	-0.08
		(-2.85)	(-1.16)		(-4.94)	(-2.66)		(-2.56)	(-1.42)
Inc	0.14	0.20	0.27	0.03	0.14	0.34	0.14	0.15	0.17
	(0.51)	(0.75)	(0.92)	(0.12)	(0.53)	(1.14)	(3.03)	(3.26)	(3.35)
Pop	0.91	0.86	0.81	0.87	0.79	0.64	0.97	0.96	0.95
	(16.53)	(15.11)	(7.89)	(15.74)	(14.11)	(6.28)	(104.46)	(99.83)	(54.91)
Temp	0.04	0.04	0.04	0.02	0.02	0.03	0.00	0.00	0.00
	(7.28)	(7.74)	(6.26)	(3.22)	(4.13)	(4.20)	(1.70)	(2.12)	(2.21)
Land Area	0.22	0.21	0.21	0.11	0.10	0.08	0.00	0.00	0.00
	(3.74)	(3.70)	(3.54)	(1.80)	(1.73)	(1.43)	(-0.19)	(-0.26)	(-0.39)
R^2	0.71	0.72	0.71	0.64	0.67	0.65	0.99	0.99	0.99
Obs./Year	319	319	319	319	319	319	319	319	319

Table 5: Market Power, Prices, and Quantities, Panel

This table presents regressions of several measures of housing prices and quantities on HHI, CBSA characteristics, and year fixed effects. The CBSA characteristics are log income, log population, average temperature, and log land area. The top panel presents prices, with the average price of newly built homes from our data in columns 1-3, the average price from the ACS in columns 4-6, and the average rent from the ACS in columns 7-9. The bottom panel presents log quantities, with the number of newly built homes from our data in columns 1-3, the total lot area of newly built homes in columns 4-6, and the total number of units from the U.S. Census in columns 7-9. The regressions are pooled from an unbalanced panel, with a single observation being MSA-Year. The sample period is 2000-2023. We present both OLS and IV regressions. The IV is a fitted value of HHI where the first stage is presented in Table 3 and the instrument is the number of other CBSAs within a 50 mile radius of the focal CBSA; unlike some of the other variables in this regression, the IV does not vary over time. Standard errors are clustered at the CBSA level. t-statistics are reported in parentheses.

	(1) OLS	(2) OLS	(3) 2SLS	(4) OLS	(5) OLS	(6) 2SLS	(7) OLS	(8) OLS	(9) 2SLS
				Panel A: Prices regressed on HHI					
		y_t =Price$_t^{New}$			y_t =Price$_t^{ACS}$			y_t =Rent$_t^{ACS}$	
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
HHI		1.06	0.48		1.13	0.80		0.47	0.37
		(4.92)	(4.12)		(4.78)	(6.56)		(5.08)	(7.59)
Inc	1.33	1.29	1.16	1.26	1.22	1.00	0.65	0.64	0.53
	(8.45)	(8.72)	(7.00)	(7.20)	(7.49)	(5.93)	(8.81)	(9.20)	(7.62)
Pop	0.05	0.08	0.17	0.02	0.05	0.23	0.03	0.05	0.13
	(1.75)	(2.78)	(3.85)	(0.80)	(1.68)	(4.99)	(2.80)	(3.73)	(7.20)
Temp	0.00	0.00	0.00	0.00	0.00	-0.01	0.01	0.01	0.00
	(0.90)	(0.31)	(-1.80)	(0.15)	(-0.54)	(-3.89)	(5.48)	(5.17)	(0.43)
Land Area	0.02	0.03	0.04	0.03	0.03	0.05	-0.01	-0.01	0.00
	(0.78)	(0.89)	(1.23)	(1.01)	(1.24)	(1.99)	(-0.64)	(-0.53)	(0.14)
R^2	0.38	0.41	0.40	0.48	0.52	0.53	0.70	0.71	0.73
Obs./Year	262.78	262.78	262.78	284.43	284.43	284.43	284.43	284.43	284.43
				Panel B: Quantities regressed on HHI					
		y_t =Units$_t^{New}$			$y_t = \Delta$LotArea$_t^{New}$			y_t =Units$_t^{Census}$	
HHI		-1.18	-0.30		-2.60	-0.85		-0.19	-0.07
		(-3.24)	(-0.94)		(-6.32)	(-2.87)		(-2.29)	(-1.55)
Inc	0.04	0.08	0.14	-0.03	0.06	0.25	0.14	0.15	0.17
	(0.14)	(0.29)	(0.51)	(-0.10)	(0.24)	(0.96)	(2.33)	(2.40)	(2.54)
Pop	0.87	0.84	0.79	0.84	0.78	0.62	0.97	0.96	0.95
	(16.65)	(16.50)	(8.18)	(16.90)	(16.45)	(7.01)	(103.12)	(98.56)	(55.11)
Temp	0.04	0.04	0.04	0.02	0.02	0.03	0.00	0.00	0.00
	(7.01)	(7.63)	(6.04)	(2.99)	(4.05)	(4.40)	(1.83)	(2.17)	(2.48)
Land Area	0.24	0.24	0.24	0.13	0.12	0.11	0.00	-0.01	-0.01
	(4.88)	(4.90)	(4.63)	(2.39)	(2.38)	(2.03)	(-0.42)	(-0.48)	(-0.61)
R^2	0.67	0.67	0.67	0.58	0.60	0.59	0.98	0.98	0.98
Obs./Year	284.43	284.43	284.43	283.61	283.61	283.61	284.43	284.43	284.43

Table 6: Changes in Prices, Quantities, and HHI

This table presents regressions of the change from t to $t+1$ in various CBSA level variables on time t controls and a shock realized at the end of t. This is a single cross-sectional regression in which t represents the period 2000-2011, and $t+1$ represents 2012-2023. The left hand side variables, represented by Δy_{t+1} are computed as the log of the average quantity for 2012-2023, divided by the average quantity in 2000-2011 for new home sale prices from our dataset, home prices from ACS, rents from ACS, new sold units from our data, lot area of new sold units from our data, and total units from the Census; for HHI we take the difference between the 2012-2023 and 2000-2011 averages. The key explanatory variable of interest is $\psi_{c,t}$, which is the shock to CBSA c, computed based on the average price changes in connected CBSAs in 2005-2010, weighted by the overlap in developers between CBSA c and these CBSAs in 2000-2005, it is defined in Section 3.5. The controls are the log of average income in 2000-2011, the log of population in 2000-2011, the growth in income and in population from 2000 to 2011, the average HHI in 2000-2011, the log of the average of the y-variable in 2000-2011, and the price growth in CBSA c from 2005 to 2010. t-statistics are reported in parentheses.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Changes in prices regressed on shock							
	$\Delta y_{t+1} = \Delta\text{Price}^{New}_{t+1}$		$\Delta y_{t+1} = \Delta\text{Price}^{ACS}_{t+1}$		$\Delta y_{t+1} = \Delta\text{Rent}_{t+1}$		$\Delta y_{t+1} = \Delta\text{HHI}_{t+1}$	
ψ_t		-7.81		-5.71		-2.27		-18.32
		(-2.65)		(-4.41)		(-2.88)		(-2.68)
Inc_t	0.37	0.37	-0.01	-0.01	0.04	0.04	-0.26	-0.31
	(3.73)	(3.83)	(-0.26)	(-0.29)	(1.36)	(1.57)	(-1.40)	(-1.67)
Pop_t	0.05	0.04	0.02	0.01	0.01	0.01	-0.06	-0.09
	(3.48)	(2.69)	(2.81)	(1.70)	(3.49)	(2.87)	(-1.92)	(-2.59)
ΔInc_t	-1.47	-1.32	2.57	2.51	1.61	1.60	12.09	12.02
	(-0.87)	(-0.79)	(3.67)	(3.68)	(3.99)	(4.00)	(3.42)	(3.44)
ΔPop_t	7.22	6.09	6.19	5.39	2.01	1.77	3.25	0.43
	(4.58)	(3.77)	(8.29)	(7.22)	(4.45)	(3.89)	(0.88)	(0.11)
HHI_t	0.08	0.06	-0.01	-0.01	0.00	0.00	-0.26	-0.29
	(3.61)	(3.06)	(-0.60)	(-1.20)	(0.66)	(0.35)	(-5.26)	(-5.80)
y_t	-0.35	-0.37	-0.07	-0.08	0.04	0.01		
	(-8.24)	(-8.62)	(-3.30)	(-3.97)	(1.25)	(0.50)		
ΔPrice_t	-0.98	-0.45	0.99	1.35	0.01	0.14	-3.45	-2.13
	(-2.38)	(-0.99)	(4.70)	(6.14)	(0.09)	(1.09)	(-3.61)	(-1.99)
R^2	0.25	0.27	0.39	0.43	0.21	0.23	0.13	0.15
Obs.	284	284	315	315	315	315	313	313
	Panel B: Changes in quantities regressed on shock							
	$\Delta y_{t+1} = \Delta\text{Units}^{New}_{t+1}$		$\Delta y_{t+1} = \Delta\text{LotArea}^{New}_{t+1}$		$\Delta y_{t+1} = \Delta\text{Units}^{Census}_{t+1}$			
ψ_t		-1.86		3.38		-0.39		
		(-0.27)		(0.50)		(-1.03)		
Inc_t	-0.02	-0.02	-0.06	-0.05	0.02	0.02		
	(-0.09)	(-0.12)	(-0.33)	(-0.28)	(2.22)	(2.07)		
Pop_t	0.26	0.26	0.28	0.29	0.00	0.00		
	(4.24)	(4.24)	(5.86)	(5.85)	(-0.11)	(-0.23)		
ΔInc_t	2.92	2.91	3.04	3.06	0.40	0.40		
	(0.82)	(0.81)	(0.85)	(0.86)	(1.96)	(1.99)		
ΔPop_t	28.56	28.47	26.59	27.09	7.57	7.50		
	(5.12)	(5.08)	(5.90)	(5.86)	(36.41)	(34.64)		
HHI_t	-0.19	-0.20	-0.17	-0.16	-0.01	-0.01		
	(-4.18)	(-4.15)	(-3.28)	(-3.08)	(-2.86)	(-2.98)		
y_t	-0.21	-0.22	-0.29	-0.29	0.00	0.00		
	(-3.98)	(-3.95)	(-6.42)	(-6.40)	(0.11)	(0.19)		
ΔPrice_t	4.04	4.18	5.15	4.90	0.21	0.24		
	(4.20)	(3.85)	(5.48)	(4.63)	(3.89)	(3.92)		
R^2	0.22	0.22	0.27	0.27	0.82	0.82		
Obs.	314	314	313	313	315	315		

Appendix

A Algorithm to Identify and Group Developers

The algorithm is largely based on finding developers by their names in the seller name field of transactions in the ATTOM data and also carefully excluding false matches. We begin by creating a list of 737 large national developers.[22] We also compiled a list of the subsidiaries of the largest 106 of these, as well as a timeline of their mergers and acquisitions.[23] We then use fuzzy matching to search all seller names for these developer names and variations.[24]

To search for smaller and local developers, we created a list of strings to search for using Partial Match,[25] as well as a list of strings to exclude.[26] After excluding duplicate transactions, this resulted in 15,500,481 observations.

We then merge this data with tax assessor data, which contains a variable for the year built, although it is available only for some of the transactions and is not always reliable.[27] We use year built for additional exclusions. For transactions under $3 million, we do not use year built, except we exclude transactions whose transaction year minus year built is greater than three and at the same time whose seller name contains the strings buy, invest, resort, or vineyard as these are likely resales. For transactions above $3 million, we exclude transactions

[22]Starting in 2004, each year Builder Magazine (builderonline.com) provides a list of the top 200 U.S. developers by total closings, this list can be found at www.builderonline.com/builder-100/builder-100-list. We combined this with an annual list of the top 25 multi-family builders provided by the National Multifamily Housing Council, this list can be found at https://www.nmhc.org/research-insight/the-nmhc-50/.

[23]To search for the names of subsidiaries we read through Exhibit 21 of 10K reports for public developers and read the Wikipedia pages and company websites of the largest 106 developers. This also allowed us to determine any merger dates for these developers.

[24]For example D. R. Horton may appear as D R Horton or DR Horton.

[25]This includes the strings homes, homebui, homeland, build, bldg, constr, bldr, resident, develop, and communit

[26]We use Partial Match to exclude sellers with strings associated with financial companies, such as loan, lend, mortg, mtg, finan, title, homesale, residential nominee, bank, fund, equity, federal home, zillow, secs, securit, insur, and national resid. Additionally Whole match was used to exclude sellers with strings fin, federal, fed, cap, capit, capital, credit, asset, hldgs, buying, trust, trustee, invest, investment, investments, portfolio, and funeral.

[27]An alternative strategy is to identify the first transaction after year built as a sale by a developer. We chose to not follow this strategy because (at least in ATTOM data), we find year built to be somewhat unreliable. For example, of all the developer transactions identified by us, 8% have year built missing, 7% have year built at least 3 years before the developer's transaction, and 13% have year built after the developer's transaction. Even if we restrict the analysis to the top national developers (Lennar Homes, KB Homes, Centex/Pulte, DR Horton, and Beazer Homes) – whose exact names we know and whose identification is more reliable – these percentages are 8%, 1% and 10%.

where the transaction's year minus the built year is greater than three. Additionally, if there multiple transactions by the same developer in the same zip code that are above $3 million and all have exactly the same price, we assign to them the transaction price divided by the number of transactions.[28] After these exclusions, 15,447,755 transactions are left.

We next drop observations where the transaction date is missing, where the transaction type is listed as a mortgage, foreclosure and auction transactions, and Quitclaim Deed observations. We also drop certain document types.[29] This leaves us with 14,293,919 observations. We then keep only properties where PropertyUseGroup is set to residential, leaving us with 11,741,946 observations. We additionally drop rows where every variable is exactly identical except TransactionID as these are likely duplicates. There are also 231,437 transactions that involve multiple builders. For such cases, if a transaction has n builders, we expand the observation into n observations with each builder receiving $1/n$th of the price and lot area. This leads to 11,954,411 total transactions.

Many national builders have common last names in the name builder names, for example Morrison Homes. To avoid including people named with the last name Morrison, for builders whose name includes a common last name, we require that the match is exact, or that the name is accompanied by the strings LLC or Home or any of the other words on our local developer string list.[30] This left us with 11,826,672 observations.

B Geographical Spread of Developers

Figure A.1 summarizes how geographically spread out developers are within a CBSA. To plot this figure, within each CBSA, for each national developer, we define the developer's

[28]This is done because sometimes a developer will developer multiple nearby properties and sell them jointly, with the price recorded for each property being the total price.

[29]We drop document types DTDT (Affidavit of death), DTIT (Intrafamily transfer), and DTGF (Gift deed).

[30]The last names are Weiland, Royce, Brookfield, Taylor, William, Walter, Morrison, Shea, Robson, Ashton, Fulton, Neumann, Bowen, Fischer, Davis, Ball, Brown, Brayson, Hunt, Young, Whittaker, Park, Suarez, Longford, Darling, Marrano, Morley, Clark, Anderson, Powers, Forrest, Roger, Crawford, Dixon, Mayer, Ginsburg, Perry, John, McCar, Elliott, Barton, Mathews, Matthews, Schumacher, Polk, Harris, Kennedy, Trammell, Jerry, Pierce, Morissette, William, Ryan, Simmons, Thornton, Shelby, Ennis, Harkins, John, Willis, Malloy, Coleman, McBee, Cooper, Benton, McCaffrey, Bloomfield, Shumaker, Armstrong, Kendall, Wynn, Cassidy, Keller, Heckman, Johnson, McGuyer, Allen, Edwin, Lombardo, Felder, Pratt, Payne, Lewis, Martin, Jacobs, Gardner, Holmes, Brien, Dickerson, Hanson, McArthur, Smith, Douglas, Caruso, McKee, Clayton, Andrews, Walker, Thomas, Robertson, Davidson, McGuinn, Adams, McKinley, McGuinn, Edwards, Nilson, Morgan, Ward, Shaw, Fischer, McGuinn, Turner, Miller, Smith, Stanley, Martin, Jordan, McCaffrey.

geographical spread as the distance between its two furthest away projects across all years. We then compute the average of the geographical spread across all national developers operating in the CBSA – this tells us how spread out within a CBSA the average national developers is. We do the same for local developers.[31] The top left panel plots the average geographical spread of developers within a CBSA against the CBSA's log population; it shows that larger CBSAs have more spread out developers. The middle left panel is similar but has log land area instead of log population on the x-axis, again developers have more spread out operations in larger CBSAs. For the largest CBSAs (either by population or land area), the average national developer has a spread of around 30 miles between its projects, and similar for the average local developer. As a point of reference, the average land area for the 10 largest CBSAs is 17,400 square miles, comparable to a circle with a diameter of 140 miles. For the bottom panel we define $d = 2 \times (LandArea/\pi)^2$, this is the diameter a CBSA would have if it was a perfect circle. A developer's geographical spread divided by d gives a rough estimate of what fraction of a CBSA a developer is willing to traverse for jobs. We plot d on the x-axis against the spread divided by d on the y-axis. The average developer traverses a half to a third of the CBSAs diameter in smaller and mid-size CBSAs, this falls to about a quarter for the largest CBSAs. However, some developers cover much more than the average. The three panels on the right are analogous to the ones on the left, but plot the spread of the most spread out rather than the average developer. Even in the largest CBSAs, some developers traverse 100% of the CBSAs diameter for jobs.

Figure A.2 shows that some developers are willing to develop projects outside of their CBSA, as long as the distance to the project isn't too large. For each CBSA we record the fraction of this CBSA's developers that also operate in other CBSAs located within a specific distance of this CBSA. We do this for various distances and also separate developers into local and national developers. More than 6% of a CBSA's local developers operate in *other* CBSAs that are a distance of 30 miles from the original CBSA and 2% of local developers operate in *other* CBSAs 90 miles away. National developers exhibit a similar pattern, but not surprisingly, they are willing to cover much larger distances. More than 25% of a CBSA's national developers operate in *other* CBSAs that are a distance of 30 miles from the original CBSA and 18% of national developers operate in *other* CBSAs 90 miles away.

[31] Because we do not keep track of local developers across years, we are unable to do this across years and instead do this separately year by year.

Figure A.1: Geographic Spread Within CBSAs

This figure shows how spread out each developer's operations are within a CBSA. Within each CBSA, we define a developer's spread as the distance between its two furthest-apart projects. We then compute the average and maximum spread for all national developers (red × markers) and, separately, for all local developers (blue circles) operating within a CBSA. The left panels plot average spread and the right panels plot maximum spread against different measures of CBSA size. In the top panel, CBSA size is defined as the logarithm of population; in the middle panel, as the logarithm of total land area. In the bottom panel, CBSA size is defined as the diameter implied by total land area, and we plot spread scaled by this diameter.

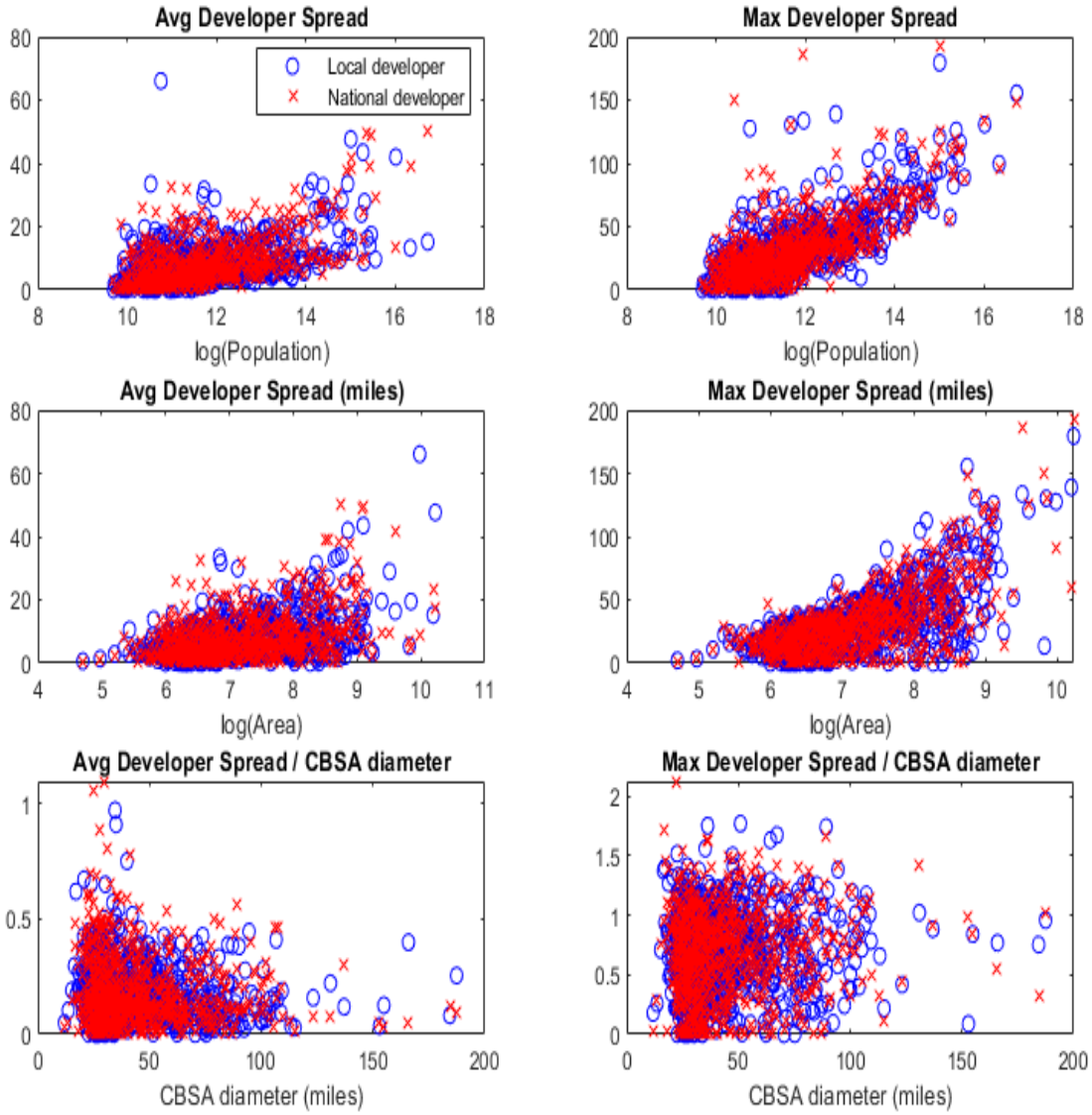


Figure A.2: Geographic Overlap Across CBSAs

For each CBSA, we compute the fraction of its developers that also operate in other CBSAs within a given distance. The top and middle panels plot the average of this measure across all CBSAs for national and local developers, respectively. The bottom panel plots the fraction of all U.S. CBSAs that lie within a given distance of a focal CBSA, averaged across all focal CBSAs.

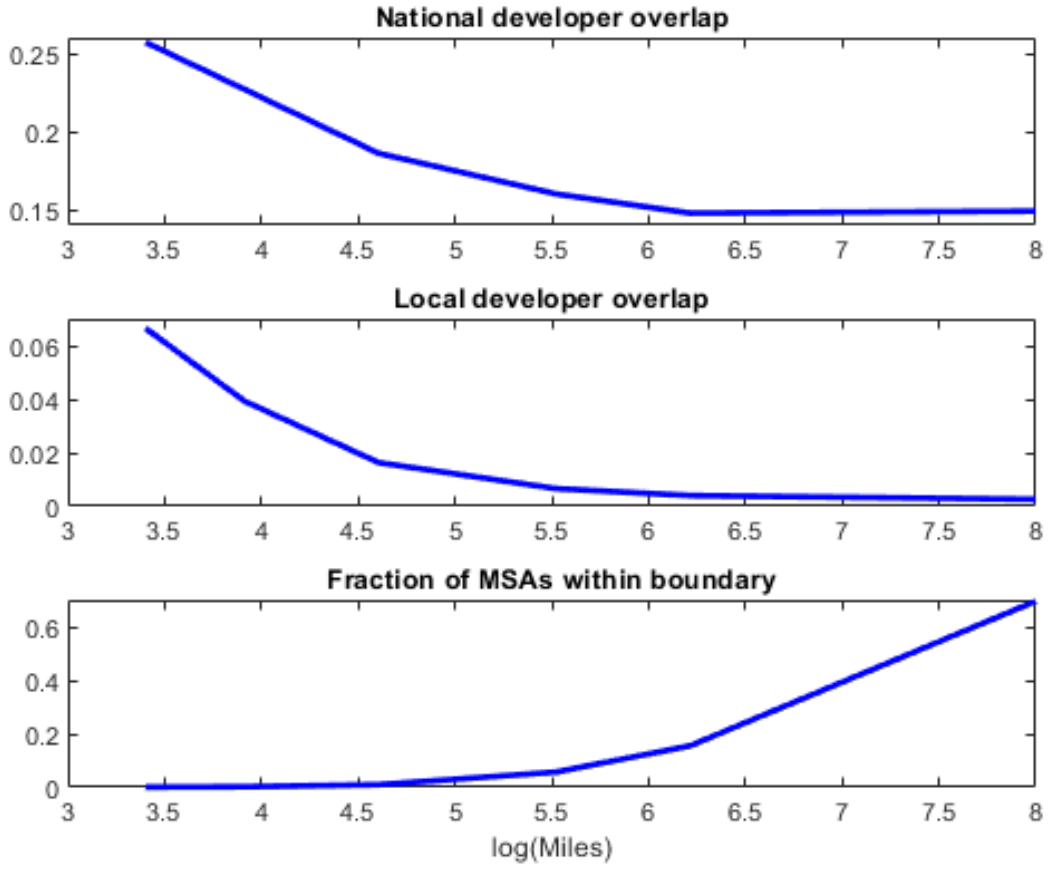


Table A.1: Local Developers HHI Covariates

This table presents cross-sectional regressions of the HHI of the local developers' market share within an CBSA on CBSA characteristics. The characteristics include log population, log income, average temperature, log land area, and components of the 2018 WRLURI index. HHI, population, and income are averaged over the period 2000-2023. Column (1) presents results without including WRLURI. The following columns include the WRLURI components one at a time with the components being: (2) political pressure, (3) political involvement, (4) court involvement, (5) local approval, (6) local zoning, (7) local assembly, (8) supply restrictions, (9) density restrictions, (10) open space, (11) exaction, (12) affordable housing, (13) approval delay, and (14) weighted average. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pop	-0.503	-0.525	-0.523	-0.528	-0.526	-0.524	-0.524	-0.532	-0.518	-0.553	-0.547	-0.529	-0.590	-0.561
	(-7.521)	(-7.268)	(-7.208)	(-7.213)	(-7.191)	(-7.185)	(-7.173)	(-7.421)	(-7.103)	(-7.647)	(-7.836)	(-7.706)	(-8.379)	(-7.930)
Inc	0.438	0.138	0.310	0.335	0.341	0.343	0.317	0.240	0.403	0.257	0.210	-0.198	-0.057	0.049
	(1.313)	(0.384)	(0.884)	(0.951)	(0.967)	(0.975)	(0.890)	(0.689)	(1.139)	(0.738)	(0.622)	(-0.575)	(-0.165)	(0.143)
Temp	0.011	0.003	0.004	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.001	0.005	0.008	0.003
	(1.719)	(0.463)	(0.553)	(0.770)	(0.783)	(0.795)	(0.723)	(0.895)	(0.782)	(0.857)	(0.091)	(0.745)	(1.134)	(0.464)
Land Area	-0.001	0.022	0.028	0.047	0.049	0.048	0.051	0.075	0.039	0.044	0.020	0.076	0.082	0.033
	(-0.009)	(0.262)	(0.344)	(0.573)	(0.598)	(0.585)	(0.618)	(0.924)	(0.479)	(0.543)	(0.252)	(0.976)	(1.051)	(0.412)
WRLURI		0.089	0.114	0.014	0.000	0.029	-0.067	0.320	-0.080	0.423	0.663	1.209	0.091	0.303
		(2.294)	(1.780)	(0.299)	(-0.002)	(0.506)	(-0.437)	(2.981)	(-1.382)	(2.916)	(4.971)	(5.709)	(5.189)	(4.387)
R^2	0.227	0.262	0.256	0.247	0.247	0.247	0.247	0.272	0.252	0.271	0.314	0.333	0.319	0.300
CBSAs	319	258	258	258	258	258	258	258	258	258	258	258	258	258

Table A.2: National Developers HHI Covariates

This table presents cross-sectional regressions of the HHI of the national developers' market share within an CBSA on CBSA characteristics. The characteristics include log population, log income, average temperature, log land area, and components of the 2018 WRLURI index. HHI, population, and income are averaged over the period 2000-2023. Column (1) presents results without including WRLURI. The following columns include the WRLURI components one at a time with the components being: (2) political pressure, (3) political involvement, (4) court involvement, (5) local approval, (6) local zoning, (7) local assembly, (8) supply restrictions, (9) density restrictions, (10) open space, (11) exaction, (12) affordable housing, (13) approval delay, and (14) weighted average. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pop	-0.340	-0.339	-0.337	-0.341	-0.343	-0.336	-0.340	-0.339	-0.339	-0.338	-0.334	-0.340	-0.351	-0.346
	(-9.137)	(-8.402)	(-8.377)	(-8.421)	(-8.514)	(-8.351)	(-8.419)	(-8.405)	(-8.375)	(-8.308)	(-8.316)	(-8.636)	(-8.613)	(-8.552)
Inc	-0.270	-0.125	-0.180	-0.175	-0.173	-0.163	-0.147	-0.180	-0.173	-0.167	-0.141	-0.367	-0.246	-0.228
	(-1.456)	(-0.621)	(-0.926)	(-0.896)	(-0.889)	(-0.841)	(-0.744)	(-0.918)	(-0.881)	(-0.855)	(-0.723)	(-1.859)	(-1.236)	(-1.154)
Temp	-0.008	-0.014	-0.015	-0.015	-0.015	-0.015	-0.014	-0.015	-0.015	-0.015	-0.014	-0.015	-0.014	-0.015
	(-2.282)	(-3.508)	(-3.807)	(-3.692)	(-3.626)	(-3.644)	(-3.547)	(-3.641)	(-3.661)	(-3.661)	(-3.385)	(-3.814)	(-3.571)	(-3.793)
Land Area	-0.004	0.029	0.015	0.020	0.031	0.021	0.022	0.026	0.024	0.023	0.029	0.033	0.030	0.020
	(-0.110)	(0.635)	(0.323)	(0.446)	(0.676)	(0.457)	(0.485)	(0.574)	(0.528)	(0.511)	(0.648)	(0.741)	(0.654)	(0.435)
WRLURI		-0.019	0.045	0.019	0.036	0.052	0.060	0.038	0.007	-0.003	-0.138	0.447	0.018	0.063
		(-0.875)	(1.279)	(0.702)	(1.561)	(1.649)	(0.704)	(0.624)	(0.220)	(-0.034)	(-1.792)	(3.686)	(1.760)	(1.592)
R^2	0.385	0.400	0.402	0.399	0.404	0.405	0.399	0.399	0.398	0.398	0.406	0.429	0.405	0.404
CBSAs	319	258	258	258	258	258	258	258	258	258	258	258	258	258

Table A.3: National Share Covariates

This table presents cross-sectional regressions of the the national developers' market share within an CBSA on CBSA characteristics. The characteristics include log population, log income, average temperature, log land area, and components of the 2018 WRLURI index. HHI, population, and income are averaged over the period 2000-2023. Column (1) presents results without including WRLURI. The following columns include the WRLURI components one at a time with the components being: (2) political pressure, (3) political involvement, (4) court involvement, (5) local approval, (6) local zoning, (7) local assembly, (8) supply restrictions, (9) density restrictions, (10) open space, (11) exaction, (12) affordable housing, (13) approval delay, and (14) weighted average. t-statistics are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pop	0.170	0.154	0.153	0.152	0.154	0.154	0.155	0.151	0.156	0.139	0.143	0.154	0.137	0.144
	(6.333)	(5.165)	(5.116)	(5.077)	(5.116)	(5.133)	(5.167)	(5.120)	(5.190)	(4.739)	(5.128)	(5.127)	(4.586)	(4.855)
Inc	0.312	0.232	0.294	0.286	0.290	0.290	0.274	0.252	0.306	0.246	0.224	0.286	0.186	0.212
	(2.329)	(1.564)	(2.030)	(1.974)	(2.005)	(2.008)	(1.875)	(1.758)	(2.105)	(1.744)	(1.661)	(1.902)	(1.280)	(1.464)
Temp	0.022	0.023	0.024	0.023	0.023	0.023	0.023	0.024	0.023	0.024	0.021	0.023	0.024	0.023
	(8.304)	(7.601)	(7.869)	(7.834)	(7.857)	(7.865)	(7.717)	(8.065)	(7.871)	(8.148)	(7.451)	(7.859)	(8.197)	(7.737)
Land Area	-0.021	0.000	0.010	0.006	0.008	0.007	0.009	0.017	0.005	0.005	-0.007	0.008	0.016	0.003
	(-0.723)	(-0.008)	(0.303)	(0.176)	(0.225)	(0.221)	(0.255)	(0.521)	(0.150)	(0.147)	(-0.227)	(0.234)	(0.491)	(0.097)
WRLURI		0.025	-0.014	0.011	0.000	0.005	-0.045	0.121	-0.021	0.218	0.332	0.009	0.024	0.081
		(1.587)	(-0.533)	(0.580)	(-0.016)	(0.219)	(-0.719)	(2.750)	(-0.883)	(3.706)	(6.212)	(0.097)	(3.200)	(2.785)
R^2	0.369	0.386	0.380	0.380	0.379	0.380	0.381	0.398	0.381	0.412	0.462	0.379	0.404	0.398
CBSAs	319	258	258	258	258	258	258	258	258	258	258	258	258	258