House Price Contagion and U.S. City Migration Networks

Gregor Schubert
UCLA Anderson *
September 19, 2022

(NEWLY REVISED - COMMENTS WELCOME)

Abstract

Why do local shocks spread between housing markets? This paper proposes an explanation of house price contagion based on migration spillovers between U.S. cities: Increases in house prices as a result of local economic shocks and housing supply constraints drive out-migration to other cities. These migration flows are more likely to affect cities with stronger pre-existing migration links to the origin cities, and increase house prices at those destinations. I derive a measure of a city's migration exposure to shocks in other housing markets from a static location choice model, and use the network structure of inter-city migration to develop an instrument for identifying the size of these causal spillover effects between cities: I find that an increase in other cities' house prices by 10 log points causes a 4.0 log point house price increase after 5 years in a city with average migration exposure (5.9 log points in top quartile exposure cities), and a 2.2 log point increase in its population. Moreover, mortgage lending and construction in the focal city respond strongly to housing market shocks in other cities. I quantify the importance of accounting for these migration spillovers in 3 applications related to the effect of mortgage credit supply shocks on house prices, explaining city-level house price betas, and estimating the effects of density on housing markets during the Covid pandemic. I find that incorporating indirect effects through migration exposure increases the ability of each shock to explain the variation in the housing market outcome by 34-44% relative to just focusing on the direct local effect, and also changes the estimated size of the direct effect of the shocks.

*Email: gregor.schubert@anderson.ucla.edu. Website: `https://sites.google.com/view/gregorschubert`. This paper represents a substantially revised version of a previous paper circulated under the same title in 2021. I would like to thank my advisors Ed Glaeser, Larry Katz, Jeremy Stein and Adi Sunderam for their guidance and support. This paper also benefited immensely from discussions with Pol Antras, Adrien Bilal, John Campbell, Gabriel Chodorow-Reich, Marco Di Maggio, Fabian Eckert, Emmanuel Farhi, Xavier Gabaix, Caitlin Gorback, Robin Greenwood, Sam Hanson, Myrto Kalouptsidi, Robin Lee, Marc Melitz. I would also like to thank participants of the Harvard Labor Lunch, Industrial Organization Workshop, International Trade Workshop, Macro Workshop, and Finance Workshop, as well as participants of the UEA Ph.D. Workshop 2020, and participants and discussants at NBER SI Real Estate 2021, WFA 2021, EUEA 2021, EALE 2021, and ARES 2021 for many helpful comments and suggestions. I gratefully acknowledge receiving financial support from the Bradley Foundation, Harvard Business School, and the John R. Meyer Dissertation Fellowship of the Harvard Joint Center for Housing Studies, and from the UCLA Ziman Center for Real Estate while working on this research.

1 Introduction

House price growth during national boom and bust cycles varies greatly across cities in the U.S. During the 2000-2007 housing boom, real house prices in Boise, ID, grew by 41%, while house price growth in similarly sized Midwestern cities like Wichita, KS, or Jackson, MS, was only 2% and 9%. Even after accounting for the direct effect of land constraints and their interaction with national macroeconomic trends, these geographic differences are substantial as Panel (a) of Figure 1 shows. Spatial gaps in cities' co-movement with national house price trends also persist over longer time periods. The map in Panel (b) of Figure 1 shows the local correlation with the national cycle (the house price "beta") over the 1990-2017 period, which ranges from 0.4 at the 10th percentile to 1.9 at the 90th. How do shocks to local housing markets spread between cities? And what affects the strength of this contagion and the correlation in housing cycles across cities?

This paper aims to answer these questions by providing evidence for the importance of *migration* as an important causal channel for spillovers between cities. In particular, I make a number of key contributions: First, I document that migration networks between cities are persistent and are good predictors of housing market correlations between city pairs. Second, I derive a theoretical measure of a city's exposure to spillovers from housing market shocks in other cities from a static location choice model. Moreover, I empirically estimate the strength of this causal spillover effect by developing an IV identification approach based on network exposure to shocks in other cities. Third, I demonstrate the importance of this insight by showing the quantitative importance of accounting for migration exposure to shocks in other locations in three applications.

The proposed mechanism for how house price growth spills over between cities is the following: First, local shocks to housing demand in a city drive up local house prices if housing supply is constrained. This causes some workers to leave the city in pursuit of more affordable housing elsewhere. These outflows are more likely to affect cities with strong migration links to the origin city – and the increased population inflows drive up those destinations' housing demand and prices. As a result of such migration spillovers, high house price growth may spread even to cities that do not receive fundamental shocks themselves, but are highly exposed to outflows from cities that do. Returning to the example from above, the high house price growth in Boise, ID, can in part be explained by the fact that its population grew by 13% over 2000-2007, while that of Wichita, KS, and Jackson, MS, stagnated. One reason for this is that Boise is an important destination for people leaving expensive booming cities in California, such as Los Angeles, San Francisco, or San Jose. These three metro areas alone were responsible for 2000-2007 net migration into Boise corresponding to almost 4% of its population.

In the first part of the paper, I start by establishing stylized facts about the networks of migration between U.S. cities: I show that, for most cities, only a small number of other cities are destinations or origins for substantial numbers of migrant outflows or inflows. Moreover, this set of relevant other cities does not change much over time. This persistence can be explained by migration costs that are a function of enduring city characteristics. I also provide new evidence that migration links individually outperform in predicting high inter-city house price correlations when compared to measures of physical distance, social network links, trade flows, or similarity in industry structure between cities.

Then, I develop an empirical approach to identify spillover effects between cities. I use a static location choice model with migration costs to derive the theoretically predicted exposure of a city to spillovers from shocks in other housing market with migration links of varying strength. I take this model-implied measure of migration exposure to shocks to the data and use the network structure of migration flows to construct each city's exposure to other cities' housing markets. Then, I estimate spillover effects from the degree to which a focal city's house price growth co-moves with house price shocks in other migration-linked cities. To identify exogenous variation in house prices in a city's migration network, I construct an instrument consisting of Bartik (1991)-type wage shift-share shocks in *other* cities interacting with their land constraints – and exposure to those shocks through long-distance migration network links. In the estimation I control for any *direct* effects of the same shift-share shocks on each focal city to ensure that any estimated effects come only from indirect spillovers via migration exposure to other cities' housing markets.

Using this "network IV" approach, I find that causal network spillover effects are sizeable. By the 5th year, a 10 log point increase in house prices in all other migration-connected cities results through the migration spillover channel in a 4.0 log point increase in house prices for a city with average migration exposure, varying over the interquartile range of migration exposure from 0.5 to 5.9 log points. The corresponding estimated spillover effect on the focal city population is a 2.2 log point increase (with an IQ range of 0.3-3.2 log points). In addition to varying across cities with different migration exposures, these spillover effects on house prices are stronger in cities that are more land-constrained: an additional 1 SD of local land unavailable for construction ($\sim$22 ppt change) increases the 5-year spillover effect by $\sim$19% relative to an unconstrained city, raising the house price impact from a 10 log point house price shock on an average migration exposure city by 0.8 ppt. I show that these spillover effect estimates are robust to using mortgage credit shocks instead of wage shocks as the source of network house price variation, changing the minimum distance of cities included in the long-distance migration network, and excluding the largest CZs.

In addition, I estimate the spillover effects on mortgage credit and construction – which support the proposed migration mechanism. Mortgage loan originations jump up by 46 log points and remain elevated for several years in response to a 10 log point house price shock throughout the migration network. Furthermore, the house price response to migration spillovers can partly be explained by congestion in the construction sector: construction increases after a shock and permit volume remains 51 log points above its initial level even after 5 years in response to a 10 log point house price shock in other cities. As additional evidence of construction lags, I show that housing vacancies decline in the cities most affected by housing booms in the 2000s and 2010s while regional time-to-build delays in housing construction increase procyclically.

How important is it to account for these migration spillover effects? I consider the effect of accounting for migration exposure effects in three different applications related to housing markets: (1) The effect of mortgage credit supply shocks (based on Loutskina and Strahan (2015)) on house prices. (2) Explaining the cross-sectional variation in house price betas (as shown in Figure 1, panel B) that is driven by land constraints. (3) Estimating the effects of the Covid pandemic on housing markets through density discounts. The common insight across these applications is that incorporating indirect effects through migration exposure to a characteristic in *other* cities increases the incremental explanatory power of each approach by 34-44% relative to just focusing on the direct local effect. Moreover, without accounting for migration spillovers, the parameter estimates suffer from omitted variable bias – so controlling for the model-predicted migration exposures can substantially change the estimated effect sizes for the shocks under consideration. Constructing controls for geographic spillovers through migration channels is therefore advisable in studies of housing market responses to shocks that affect many locations simultaneously - and can be done with very little effort.[1]

[1]In fact, off-the-shelf estimates of migration exposures between geographies as used in this paper can be obtained from the author via email or his website.

These findings also have important implications for policy-makers and other stakeholders in real estate markets: Workers' ability to move across cities affects the links between cities' housing markets and the effectiveness of local housing market policies. My analysis suggests that migration links to other cities that are experiencing economic shocks can help both real estate investors and urban planners to anticipate subsequent increases in local housing demand. Moreover, to the degree that the migration spillover mechanism explains part of the volatility in house prices in U.S. cities like Phoenix, AZ, or Boise, ID, it also informs financial regulators and macroeconomic policy-makers evaluating to what degree local asset price changes are the result of speculation or credit supply changes rather than driven by demand fundamentals that may originate in *other* cities.

Related literature. This work builds on a number of papers that have documented housing market connections between cities: DeFusco, Ding, Ferreira, and Gyourko (2015) show that contagion appears to have played a role in the expansion of the 2000s boom across local housing markets. In contrast to this paper, they focus on closest neighbors and do not consider migration as an important spillover channel in their analysis. My analysis also complements Chinco and Mayer (2016), who show that out-of-town speculators play an important role in driving house price dynamics in the mid-2000s. If spillover migrants are accompanied by out-of-town speculators, my paper provides one explanation for why some cities are more likely to be the target of out-of-town speculators than others. Moreover, the *cross-city* migration mechanism that I am studying is similar to the *within-city* gentrification dynamics shown by Guerrieri, Hartley, and Hurst (2013). They show that neighborhoods that are closer to rich neighborhoods are more likely to experience house price increases after positive economic shocks to a city, providing implicit evidence of within-city migration as the driving force of this dynamic. In contrast, I study the spillover effect of positive shocks *across* proximate cities, where "proximity" is defined by the ease of migration. Bailey, Cao, Kuchler, and Stroebel (2018a) also study network connections between cities – in the form of social media friendships – and their effect on individual house purchases. However, they focus on heterogeneity in online social networks among different residents *within* one county. In this paper, I try to explain overall house price dynamics for a large network of cities using a mechanism explicitly focusing on the effect of movers, which may complement spillovers through social media connections. Cotter, Gabriel, and Roll (2015) document large correlations in housing returns between cities, and Sinai and Souleles (2013) showed that there are positive house price correlations across migration-linked cities. I provide a mechanism for how migration links can cause these house price correlations, as well as new evidence that compares migration links to other measures of inter-city links.

This paper is also closely tied to a large literature on differences in house price growth and mortgage credit origination across geographic areas.[2] For example, a rapid expansion in mortgage credit in some areas that drove house price increases during the run-up to the housing boom of the 2000s has been tied to the presence of subprime borrowers (Mian and Sufi, 2009), bank deregulation (Favara and Imbs, 2015), credit supply subsidies from government-sponsored enterprises in conjunction with differences in financial integration (Loutskina and Strahan, 2015), and changes in conforming-loan limits (Adelino, Schoar, and Severino, 2015). Moreover, regional differences in house price growth have been shown to arise from the interaction of national changes in real interest rates with local housing supply constraints (Himmelberg, Mayer, and Sinai, 2005; Glaeser, Got-

[2]There is also a voluminous literature trying to explain variations in the housing cycle within cities (Mian and Sufi, 2009), or the size of the national cycle (e.g Landvoigt (2017); Kaplan, Mitman, and Violante (2020)) which does not explicitly focus on the cross-sectional geography of differences in the housing cycle across cities.

tlieb, and Gyourko, 2012), and a concentration of high-income households in supply-constrained cities (Gyourko, Mayer, and Sinai, 2013). More generally, Chodorow-Reich, Guren, and McQuade (2021) show that local housing cycles can be driven by shocks to local fundamentals. This paper shows that all of these local shocks can also have indirect effects through migration spillovers on other cities that are not directly exposed to them. At the same time, I propose differences in migration exposure to house price and credit shocks in other cities as a new explanation for the geographic variation in the expansion of mortgage credit and house price growth.

In this paper, I also add to a literature that links migration flows and house price dynamics. For example, Howard (2020) shows that migration flows into a city stimulate the local economy and house prices, and Boustan (2010) shows that the arrival of Blacks during the Great Migration had a negative effect on city housing demand by whites. While this literature focuses on the effect on the city receiving migrants, I explicitly connect house price dynamics in origin and destination cities and quantify the importance of such spillovers in studying housing markets and understanding regional house price dynamics.

While the intuition that shocks can "spill over" between cities, as workers substitute between locations, is common to many papers with spatial equilibrium dynamics, my contribution consists of both quantifying the size of these causal spillover effects through a migration channel, and applying this intuition to understanding historical patterns of house price contagion and cross-sectional variation in housing booms, where migration links have previously not been considered as a driving force.

2 U.S. City Migration Networks: Descriptive Evidence

This section shows that there are meaningful differences in migration networks between U.S. cities, what drives these differences, and that they are persistent over time.

2.1 Migration networks between U.S. cities are sparse and persistent

When I am talking about the "U.S. city migration network", I am referring to the size of migration links between commuting zone ("city") nodes. These links are characterized by the number of migrants flowing from an origin city to a destination city. Moreover, we can distinguish between inflow and outflow networks, which can be very different for the same reference city.

To illustrate the geographic variation in migration networks, Figure A1 shows the migration network for the 2000-2007 housing boom period for inflows into Boston (left panel) and Dallas (right panel). For comparison, all inflows have been scaled to be shares of all inflows into the city from the continental U.S., with darker colors indicating a greater share of inflows coming from that origin commuting zone. Comparing the two maps, four things are immediately obvious: (1) Larger population centers naturally have a greater migration impact on other cities. (2) Migration networks can differ substantially between cities. (3) Migration flows are not limited to within-region flows, as many strong links extend to the farthest corners of the U.S. (4) Migration networks are sparse – that is, a majority of possible CZ-to-CZ links are too small to be recorded in the IRS data, while large flows concentrate into a small number of origin CZs for each destination. To show that these patterns are not unique to Boston and Dallas, Appendix Figure A1 shows the migration networks for a number of other large CZs.

Number of migration links. We can also document these connectivity patterns more systematically by computing the average number of inflow links that cities have, where a "link" is defined as an annual migrant inflow from another city recorded by the IRS.[3] These network characteristics are shown in Table 1 for the 1990-2010 period.[4] In the first row, I document that CZs on average receive migrants from 30 other CZs per year from a potential total of 722 CZs in the continental U.S.,[5] which shows that migration networks are sparse. That is, sizeable migration only occurs between particular city pairs rather than being widely dispersed.

Persistence. These migration links are also highly persistent over time – indicating that they are not driven by temporary shocks: even at a 10-year horizon, on average 80% of the origin CZs of incoming flows remain the same. Moreover, persistent migration networks do not only exist at a local level: the second and third rows of Table 1 show that, in an average year, cities receive inflows from 26 CZs that are at least 50 miles away, and even at 150 miles distance, the average city still has 21 links, of which 62% persist 10 years later. Comparing median and mean links, we can see that migration connectivity with other cities is skewed: many cities only have a handful of long-distance migration links, while a few (usually large) cities are widely connected.

Predicting "superstar" city house price dynamics. As an example of how this persistence contributes to understanding historical house price dynamics, consider, for example, the role of "superstar cities" (Gyourko et al., 2013) in the boom of the 2000s: As I show in Appendix E.2, by focusing on the cities that were experiencing high housing demand together with binding supply constraints before the year 2000– and identifying the historic destinations for migration out of these cities– one would have done very well at identifying the second-tier cities, such as Las Vegas, NV, or Phoenix, AZ, that ended up experiencing unusually large housing booms in the post-2000 period. That is, understanding that migration links are persistent over time helps in understanding why particular cities end up booming at the same time while others do not.

Characteristics of movers. In order to understand which population groups are driving U.S. migration patterns in general, I consider evidence from the Current Population Survey in Appendix E.4 on the demographics of inter-county movers in the U.S. I find that inter-county migration is driven mainly by young workers, and by workers who are employed. In contrast, older age groups' share of migration is less than their population share and retirees make up less than 10% of intercounty movers.

2.2 Migration links are predicted by city similarity

Why are migration networks persistent? One explanation may be that migrants belonging to a particular demographic group are attracted to cities where other members of their group have gone before. There is a wealth of evidence that this is the case for *international* migration. For instance, Bartel (1989) shows that immigrants to the U.S. tend to cluster in cities that already have a high share of immigrants of the same ethnicity.[6] A similar mechanism of path dependence has been found for historical domestic migration patterns (see, e.g. Boustan (2010)). A possible explanation for persistent migration patterns may be time-invariant city characteristics that shape both historical and later moving patterns in a similar way. For instance, Fishback, Horrace, and

[3]The threshold for the IRS to record the name of origin cities are flows corresponding to at least 10 tax returns.

[4]In 2011, the IRS changed its methodology for computing gross migration rates, leading to noise in the data, which is why I do not include post-2011 here. However, net migration rates do not seem to be affected by the methodology change, which is why I use data through 2017 in the regressions.

[5]Based on 1990 CZ definitions.

[6]In fact, this insight has seeded a large empirical literature in economics which uses historical immigration flows between geographies as predictors of later flows (e.g. Card (2001)). One proposed explanation for why immigrants follow in each other's footsteps is that group connections to previous migrants decrease uncertainty about the destination and lower moving and adjustment costs (Carrington, Detragiache, and Vishwanath, 1996).

Kantor (2006) show that natural amenities played a role in U.S. domestic migration in the 1930s.[7]

As I am not aware of a recent study that systematically explores the correlates of migration costs over the last two decades, I analyse whether this is the case in the U.S. in this section.

Migration gravity model. There are two separate descriptive questions to answer: (1) What factors correlate with a city's attractiveness to migrants from *any* location? (2) Holding city attractiveness constant, *which* cities are more likely to be part of the same migration network than others?

To answers these questions, I first estimate a simple gravity model of migration of the form

$$\ln Mig_t^{i \to j} = \alpha_t + \underbrace{\gamma_1 \ln dist_{ij} + \beta' \|\mathbf{X_j} - \mathbf{X_i}\|}_{\text{Migration cost}} + \underbrace{\theta_{it} + \theta_{jt}}_{\text{City attractiveness}} + \epsilon_{ijt} \tag{1}$$

where I decompose log annual migration flows between each city pair into a national trend α_t, various bilateral distance and difference parameters that capture the cost of migrating between cities; and fixed effects that capture the overall attractiveness of a city, i.e. its overall tendency to receive or send migrants. I estimate this equation using Poisson Pseudo-Maximum Likelihood on the total sample of 1990-2017 IRS migration flows between CZs. To explore differences by education group, I also estimate the migration cost parameters on a shorter 2005-2017 panel of flows by college and non-college workers, which is imputed from ACS and IRS data as described in Appendix D.7.

Migration cost determinants. Migration costs are parameterized to depend on various bilateral components. The most basic proxy for geographical distance is log physical distance between origin and destination. In addition, I allow migration costs to depend on cities being in the same state or Census region, as well as interactions of these institutional boundaries and physical distance. Informed by the migration literature discussed above, I also include absolute differences between cities in the population share of non-mainline Christian denomination[8] and ethnicity shares (as proxies for culture), water access and mean January temperatures (climate and natural amenities), and 2-digit industry shares (economic structure).[9]

The estimated coefficients for the migration cost determinants are shown in Table 2, with negative coefficients indicating that greater distance or absolute difference imposes a greater cost on migrants. Overall, almost all the included city distances and differences have a significant negative effect on the ease of bilateral migration.[10] This rationalizes persistent differences in migration networks between cities: for a given level of city attractiveness, migrants are more likely to flow to and from cities that are more similar, or closer to one another.[11] In Appendix Section E.1,

[7]For the more recent period of the 1990s, Glaeser and Shapiro (2001) find that cities with warmer and drier weather were able to attract more migrants, and Glaeser, Kolko, and Saiz (2001) show that, in addition to weather, U.S. city growth since the 1970s also correlates with consumption amenities, such as coastal proximity, restaurants, and performance venues.

[8]This measure of cultural differences is suggested by Saiz (2010), who uses it as a proxy for a culture that is less likely to impose regulatory restrictions on housing supply.

[9]Both differences in ethnicity shares and industry shares are computed as vector distances between the origin and destination cities' vectors of shares.

[10]In particular, the results for the long sample in the first column show that physical distance and crossing state boundaries significantly increase migration costs, albeit physical distance matters less conditional on leaving the state or region. Migrants are also more likely to move to areas that are culturally similar as measured by the differences in the ethnic composition and the prevalence of non-mainline Christian denominations; and workers find it easier to move to areas with a similar industry composition. Differences in local access to the water increase migration costs.

[11]The only exception is that migrants were actually *more* likely to move between cities with greater differences in

I additionally look at the drivers of variation over time in flows between cities and find that wages and house prices play an important role in understanding why cities vary in their attractiveness to migrants over time.

Heterogeneity in migration costs. The results by education group (see columns 2 and 3) highlight that the aggregate migration patterns may obscure some differences between groups. I find that physical distance and state / regional boundaries matter slightly more for non-college workers.[12] Conversely, college workers seem to put a greater weight on similar industry composition and the prevalence of non-traditional religious denominations. However, both the results here and in Appendix Section E.1 show that the drivers of migration flows are qualitatively similar between groups.

As an example of the estimated effects of differences between cities on migration, consider again migration into Boise, ID: Its industry structure is much more similar to that of Los Angeles than the difference for an average city pair.[13] The predicted effect of this similarity in industry structure on migration, based on the estimates in Table 2, is that college-educated migration flows from Los Angeles to Boise should be 31% times larger than for a city pair with average industry structure differences.[14] Similarly, the shorter distance from Los Angeles is predicted to lead to 46% greater flows of college-educated workers to Boise, ID, than to Wichita, KS, for example.[15] Of course, part of the reason for greater migration links between particular city pairs that is not captured in this regression may be historical path dependence: migration flows in the past – whatever their cause – lead to greater family and social connections between cities, and more information about opportunities at the destination, and thereby make it a more likely destination in the future. For example, already in 1980, 7% of all people in Idaho (where Boise is by far the largest city), were born in non-neighboring California, making it the most important state of origin for non-native residents.[16]

Overall, these results suggest that the persistence of migration networks is partially explained by workers seeking out cities that are physically and institutionally close, and which have similar cultures and employment structures. This rationalizes the use of historical migration shares in the reduced form analysis as sufficient statistics for the potential of later migration spillovers between particular city pairs.

January temperatures, perhaps reflecting the large ongoing movement towards the warmer regions of the American South (Glaeser and Shapiro, 2001).

[12] This is in line with findings by Molloy, Smith, and Wozniak (2011) that interstate migration rates increase with education during the 1980-2010 period.

[13] The 2-digit industry share vector distances are 0.07 for Boise-L.A. and 0.28 on average. The reason is that both cities have an unusually large share of local employment in business services industries, such as finance and information technology, which represent 24% of year 2000 employment in Boise, and 29% in Los Angeles. In contrast, these industries only represent 14% of employment in the average commuting zone. I follow Eckert, Ganapati, and Walsh (2019) in defining Business Services employment as consisting of NAICS-5 industries, that is, it includes NAICS sectors 51, 52, 53, 54, 55, and 56. The employment shares are computed from QCEW data.

[14] This is calculated as $exp(-1.291 * 0.07)/exp(-1.291 * 0.28) = 1.31$, where I use the coefficients on industry structure differences in Column 2 of Table 2.

[15] This is calculated as $exp(6.280*(-.682))/exp(6.836*(-.682)) = 1.46$, where $6.280 = ln(533)$ is the log distance in miles from L.A. to Boise, and $6.836 = ln(931)$ that from L.A. to Wichita.

[16] The question of state of birth by current state of residence was asked for the first time on the 1980 census. Overall, 50.9% of all Idaho residents were born out-of-state in 1980, and 14% of the latter were born in California. Census data is available from: `https://www.census.gov/data/tables/time-series/demo/geographic-mobility/place-of-birth-decennial.html`

2.3 Migration links predict house price correlations

Sinai and Souleles (2013) first noted that pre-Great Recession U.S. migration tended to occur between MSAs that have correlated house price cycles. This correlation can be explained by the spillover mechanism that I propose in this paper: If migrant flows through migration networks transmit local house price dynamics, we would expect house prices to be correlated more between cities that have stronger migration links. Moreover, understanding the role of migration links is of particular importance if they have predictive power for housing market correlations that goes beyond other inter-city connections.[17]

To test this predictive power of migration links empiricaly, I first compute the correlation of house price growth between pairs of CZs over 1990-2017.[18] Then, I compute for each CZ i the *weighted* average correlation with all other CZs $k \in N$, which is given by

$$E_i^m(corr(p_i, p_k)) = \sum_{k=1}^{N} w_{ik}^m \cdot corr(p_i, p_k).$$

The weights w_{ik}^m represent a particular measure m of inter-city links considered. If the inter-city link measure m better predicts pairwise house price correlations, the link-m-weighted expected correlation should be larger. [19]

My main measure of links reflects migration connections: the average share of all migrants out of city i in the IRS migration flow data that move to city k. I compare the ability of migration weights to predict house price correlations to a number of other competing measures of inter-city links: (1) An inverse-distance weighted measure that reflects geographic proximity; (2) A social connectedness index (SCI) that reflects the strength of social networks between cities (see Bailey, Cao, Kuchler, Stroebel, and Wong (2018b)); (3) A destination population-weighted measure of house price correlations that reflects the outsized role of large cities in the U.S. economy; (4) A measure of the similarity in industry structure between cities;[20] (5) An equal-weighted measure that simply reflects a city's average correlation with other cities' house prices; (6) A measure of trade flow links between cities.[21]

The expected house price correlation for each CZ under these different weights is shown in Figure 5. The graph sorts CZs by their migration-weighted house price correlations (which therefore form a line), and then plots for each CZ i the expected house price correlation with all other CZs under the alternative weights described above. A higher value indicates that the link captured by those weights is more strongly associated with a co-movement in house prices.

As is clear from the graph, the correlation in house prices between cities that have stronger migration links is almost everywhere larger than the house price correlation predicted by the other city links. This predictive outperformance suggests migration links can *not* simply be considered as proxies for one of these other city connections.[22] To quantify the predictive importance of

[17] I expand on the analysis of Sinai and Souleles (2013) by showing that migration links are in fact a *better* predictor of house price correlations than other plausible measures of inter-city links. I also include the decade since the Great Recession in the time frame, and expand the sample geographically by including commuting zones that cover the entire continental U.S.

[18] I use the log change in the FHFA repeat-sales index as my measure of house price growth.

[19] The weights are always normalized to sum to one.

[20] More details on these measures can be found in Appendix Section E.3.

[21] See Appendix Section D.5 for details on the construction of this measure, which reflects the share of a city's domestic value of outgoing trade flows that is going to another city.

[22] In Appendix Figure A7, I repeat the same analysis omitting any CZs within a 50 or 150 miles distance of one

migration links, in Appendix Section E.3, I run horserace regressions between these different link measures and find that migration links have strong and significant predictive power for house price correlations even when accounting for all other measures together.[23]

3 Empirical Approach

In this section, I use a static location choice model to derive the reduced form relationship between house price shocks in other cities and house prices in a focal city. Then, I describe how I measure this migration exposure to shocks in other cities empirically. Next, I discuss the construction of an instrument for house price changes in a city's migration network and show how I identify causal house price spillover effects between cities. Last, I introduce the data sources for the reduced-form estimation.

3.1 Conceptual framework: spillovers in a static location choice model

In this section, I motivate the functional form for the reduced form analysis by deriving a static model of inter-city migration. This model shows why house price spillovers between cities are possible, and disciplines the functional form and empirical specification to test whether there are causal effects operating via migration networks. Moreover, this model introduces the key ideas underlying the discussion of identification concerns in the reduced-form estimation and motivates the choice of control variables to include in the estimation.

Intuitively, the model reflects the following mechanism for spillovers of house prices between cities: (1) Some cities receive shocks that drive up local house prices (e.g. a technology shock increasing skilled workers' wages in large and supply-constrained cities). (2) This increase in house prices makes the city less affordable and makes local residents more likely to move to other cities. (3) These moves are more likely to have as their destination cities with strong pre-existing migration links to the origin city. (4) As a result of increased housing demand due to the population spillover, these migration-linked destination cities also experience higher house price growth. The reduced form pattern is one where, as a consequence of this spillover mechanism, we would observe house prices move together between cities that have stronger migration links.

Worker preferences. All workers live in some city $i \in N$. They have Cobb-Douglas utility over tradable consumption goods with uniform prices across all locations, and local nontradable goods, including housing, with unit cost Q_{it}. The indirect utility for a worker ω in location i can therefore be written as:

$$U_{it}(\omega) + z_i(\omega) = \ln A_{it} + \ln W_{it} - \alpha \ln P_{it} + z_i(\omega),$$

where α is the preference for nontradable goods. Here, W_{it} is the local wage and A_{it} captures residential amenities from living in i. To model the heterogeneity among workers, I assume that in addition to the common component of utility $U_{it}(\omega)$, in each period workers draw an idiosyncratic location amenity shock $z_i(\omega)$ for each location i, which is Type 1 extreme value (Gumbel) distributed with shape parameter θ. A smaller θ corresponds to less heterogeneity in idiosyncratic

another from the calculation of the averages to capture only the long-distance effect via these links - and find very similar results.

[23]The only measure that is not directly included in the horseraces is the trade flow link measure, which is only available for a small subset of CZ pairs, representing $< 10\%$ of the full sample.

preferences among workers and therefore a higher sensitivity to differences in common factors between locations. The idiosyncratic location amenity is realized right after the moving decision.

In addition, whenever a worker moves between two locations i and k, she has to pay an additive and time-invariant moving cost τ^{ik} in utility units in the same period. The worker ω therefore chooses the destination location k by solving the following problem:

$$\max_{k \in N}\{U_{kt} - \tau^{ik} + z_k(\omega)\}.$$

Moreover, workers think of these moves as being once-and-for-all and don't take into account future moves or future realizations of wages or housing costs.[24] The probability of a worker choosing to move from i to k in period t therefore takes the standard logit form

$$\mu_t^{ik} = \frac{\exp(U_{kt} - \tau^{ik})^{\frac{1}{\theta}}}{\sum_j^N \exp(U_{jt} - \tau^{ij})^{\frac{1}{\theta}}}. \tag{2}$$

Housing supply. House prices are a function of city size, i.e.

$$P_{it} = \kappa_i L_{it}^{\eta^H} e^{\varepsilon_{it}^H} \tag{3}$$

where η^H is a supply elasticity parameter (which will be allowed to vary across cities in some empirical specifications), κ_i is a location-specific housing productivity parameter, and $\varepsilon_i^H t$ is a random time-varying shock to housing supply. Both of these parameters are assumed to be exogenously determined and time-invariant. [25]

Spillovers. We can use this location choice model to derive the expected housing market spillovers between cities. [26] Going forward, lowercase letters denote log changes in the corresponding upper case variables. By using the equilibrium condition that the current local population in each city is the sum of all the city populations in the previous period that decided to move that city (or stay there), we can derive

$$l_{it} = \Delta\mathcal{M}_{i,t-1} + \theta^{-1}\left[(a_{it} - a_{it}^{nw}) + (w_{it} - w_{it}^{nw}) - \alpha(p_{it} - p_{it}^{nw})\right] \tag{4}$$

where I have defined the change in "migration access" $\Delta\mathcal{M}_{i,t-1} = \sum_k \phi_{t-1}^{i \leftarrow k} l_{k,t-1}$, where $\phi_{t-1}^{i \leftarrow k} = \mu_{t-1}^{ki} L_{k,t-2} / \sum_j \mu_{t-1}^{ji} L_{j,t-2}$ is the share of city i's population coming from city k in the previous period. The notation $x_{it}^{nw} = \sum_k \phi_{t-1}^{i \leftarrow k} \sum_j \mu_{t-1}^{kj} x_{jt}$ denotes a migration-weighted sum of the log change in characteristic X over city i's entire migration network. This expression sums first over the importance of different cities k for city i's migration flows, and then over the degree to which city k residents care about different migration destinations j. Intuitively, the population of city i will grow if it either has strong migration links to cities with growing populations, or its characteristics - amenities, wages, and house prices - are changing favourably relative to those of the cities in its migration network that it is competing with.

[24] Alternatively, one could think of workers in this static model as being born in city i, deciding to move (or not) to some location k, and then dying at the end of the period to be replaced by a descendant who does not enter their parent's utility function.

[25] I do not distinguish between rents and house prices, as is common in static spatial equilibrium models (e.g. Ahlfeldt, Redding, Sturm, and Wolf (2015); Caliendo, Dvorkin, and Parro (2019); Sturm, Heblich, and Redding (2020)).

[26] See section C.1 for the full derivation.

Together with the housing supply equation 3, we can then derive the reduced form relationship between focal city house price growth and house price shocks in its network:

$$p_{it} = \eta^{nw} p_{it}^{nw} + \xi_{it}^{p}, \tag{5}$$

where $\eta^{nw} = \frac{\alpha\eta^{H}\theta^{-1}}{1+\alpha\eta^{H}\theta^{-1}}$ is the network spillover effect. Defining $\zeta = \frac{\eta^{H}}{1+\alpha\eta^{H}\theta^{-1}}$, the residual

$$\xi_{it}^{p} = \zeta\Delta\mathcal{M}_{i,t-1} + \zeta\theta^{-1}(a_{it} - a_{it}^{nw}) + \zeta\theta^{-1}(w_{it} - w_{it}^{nw}) + \left(\frac{\zeta}{\eta^{H}}\right)\Delta\varepsilon_{it}^{H} \tag{6}$$

represents the potential confounders that I will address through control variables and the IV approach.

Intuitively, we solved for the increase in local house prices as a result of population changes that is consistent with equilibrium location choices. That is, after observing the change in house prices in city i in response to period t migration, the marginal migrant is indifferent between city i and their origin city. Note that migrants in this model are not "surprised" by the house prices changes in their destination city in response to migration, but rather take the response of house prices in their destination city into account when deciding to move. Based on the derivation above, the strength of the spillover effect η^{nw} depends on the share of expenditures devoted to housing, the slope of the housing supply function, and the inverse of heterogeneity in idiosyncratic location preferences.

Identification issues arise because residual changes in local house prices ξ_{it}^{p} may be driven by residual changes in local amenities or labor productivity relative to the migration network's characteristics, or local shocks to housing supply - and these are addressed in detail in the empirical approach.

3.2 Baseline network spillover specification

To measure the spillover effects of house price changes in other cities, we need a measure of the size of a focal city's migration network exposure to house price changes in other cities. This measure should reflect exposure to changes in other cities' characteristics through migration flows. The conceptual framework presented above implies that the reduced form local effect of house price shocks in other cities is proportional to the network house price growth term

$$p_{it}^{nw} = \underbrace{\sum_{j:j\neq i} \psi^{ij}}_{\substack{\text{Migration}\\ \text{exposure}}} p_{jt}, \tag{7}$$

where the model-implied migration exposure weights can be written as

$$\psi^{ij} = \sum_{k\in N} \underbrace{\phi^{i\leftarrow k}}_{\substack{\text{In-migration}\\ \text{share for } i}} \underbrace{\mu^{k\rightarrow j}}_{\substack{\text{Out-migration}\\ \text{share for } k}} . \tag{8}$$

The empirical implementation of the network price shock measure in equation 7 only differs from the version in the location choice model by explicitly excluding the focal city i from the network, such that we don't include indirect effects of city i's shocks that affect other cities' option sets because they would be mechanically correlated with its house prices. However, even if we included these indirect effects, they would be of second-order importance because they are always weighted by the product two small migration flow shares. In practice, in the estimation I only use network variation that excludes the focal city, and the IV approach thus explicitly prevents this kind of mechanical correlation issue from arising.

Intuitively, the expression in equation 8 shows that migration exposure of city i to changes in house prices in some other city j depends on two variables: (1) the importance of city j as a destination for migrants from all other cities k (captured by the the outmigration share $\mu^{k\rightarrow j}$) and (2) the degree to which redirected migrants from city k will increase inflows into city i if city j becomes more expensive. (the inmigration share $\phi^{i\leftarrow k}$). These two components capture the fact that, in a network setting, migration from city k to city i will not only depend on shocks to either of these cities, but also on changes occurring in other locations that indirectly compete with them for migrants.

Empirical migration exposure. Both of the components of equation 8 can be directly computed from data on gross migration flows between geographic areas, for which I use migration data from the Internal Revenue Service (see section 3.7 for details). In order to avoid concerns over endogenous changes in these migration exposure weights over time, I hold the migration network weights fixed across years at baseline period values. In particular, I average inter-city migration flows over 1990-1995 reported in IRS migration data to construct the migration weights $\psi^{ij}_{\text{`90-`95}}$. This practice of holding the migration weights fixed at a baseline period level has a strong precedent in the literature using historical migration shift-share instruments (Altonji and Card, 1989; Boustan, 2010; Howard, 2020; Derenoncourt, 2019). [27]

There is a possibility that local shocks could affect neighboring commuting zones directly through changes in commuting patterns or overlapping housing markets – without a need for a migration channel. For instance, shocks to the New York financial sector may affect house prices in some parts of Northern New Jersey directly through the incomes of long-distance commuters. I control for this possibility by excluding from any network measures commuting zones that are too close to one another in terms of physical distance. In the most stringent baseline specification, I will exclude any CZs that contain counties of which the county center is less than 150 miles away from *any* county center in commuting zone i. This ensures that the estimated house price spillovers operate at a distance that requires long-distance migration, such that a worker could not retain their local job when moving. [28]

Regression specification. The baseline spillover effect that we are interested in is the effect of network house price changes p^{nw}_{it} defined above on house price growth in city i. I estimate the following reduced-form relationship:

$$p_{it} = \alpha_i + \alpha_t + \eta^{nw} p^{nw}_{it} + \beta' \Gamma_{it} + \tilde{\xi}^P_{it}, \tag{9}$$

Here, the network spillover coefficient η^{nw} is a function of constant model parameters, and Γ_{it} represents observable control variables capturing characteristics of the focal city or its migration network. These covariates are discussed in detail in Section 3.6. The error term $\tilde{\xi}^P_{it}$ captures unobservable relative changes in other drivers of location choices in the focal city relative to its migration network, i.e. amenity and wage changes, and local differences in effect coefficients.

In this specification, we are mainly interested in the network spillover effect η^{nw}. If house price growth in other cites can propagate through migration networks, we would expect those cities to

[27] See Appendix Section D.2 for further discussion on why I choose to construct weights this way.

[28] Note that this inverts the mechanism considered, for example, in DeFusco, Ding, Ferreira, and Gyourko (2018), who focus on spillovers between nearest-neighbour cities.

experience higher house price growth that are more exposed to cities with increasing house prices through their migration network. That is, positive house price growth spillovers should be reflected in $\tilde{\eta}^{nw} > 0$. Moreover, given that the postulated channel for these effects runs through migration flows that cause house price increases, we should be able to find effects in the same direction for migration flows when we replace the dependent variable by the net migration rate.

In the next section, I present an identification strategy that allows me to estimate causal spillover effects of house prices in this specification.

3.3 Construction of network instruments

The main identification concern in estimating Equation 9 arises from the fact that the house price residual $\tilde{\xi}^P_{it}$ in a location choice model is likely to be a function of focal city and migration network changes in amenities and wages.

Intuitively, if house price changes in different cities are fully offset by changes in wages or amenities, then shifts in housing cost might not cause migration, as the overall attractiveness of the city would not change. Similarly, any migration network spillover effects found may be due to a correlation in the unobserved amenity and productivity shocks in connected cities rather than causal effects of one city on another. In general, there may be omitted variable bias if $Cov(p^{nw}_{i,t-1}, \tilde{\xi}^P_{it}) \neq 0$. To address this concern, I develop a network instrumental variable approach for network price changes, which combines city-level Bartik (1991) shift-share instruments common in the house price literature with the network structure of migration flows.

City-level "Bartik" instruments. In standard location choice models, housing demand is a positive function of local labor productivity shocks – and the size of the house price response to these shocks will depend on the local house price elasticity (Glaeser, Gyourko, and Saiz, 2008). This insight motivates the use of shift-share shocks exploiting exogenous local wage changes, in combination with housing supply constraints, as instruments for house prices. The intuition behind the common use of these instruments in the house price literature is that *national* trends in wages over time in particular industries – for example due to secular trends in technology – are plausibly exogenous with regard to whether cities with high exposure to those industries are experiencing idiosyncratic *local* house price shocks over time. However, through the interaction of the resulting wage changes with local supply constraints these national wage shocks can drive local variation in house prices, making them a valid instrument.

As a preliminary step, I construct plausibly exogenous local wage shocks B_{it} by combining the local wage bill share $\tilde{\omega}_{\iota,i,t_0}$ of workers in 3-digit NAICS industries ι in a baseline period t_0 with national wage growth $\Delta \ln W^{US}_{\iota,-i,t}$ in that industry (as a proxy for industry wage trends) in the form

$$B_{it} = \sum_{\iota} \tilde{\omega}_{\iota,i,t_0} \Delta \ln W^{US}_{\iota,-i,t}.$$

In order to minimize bias from endogeneity in the local industry exposure – which might result from auto-correlated national industry shocks (Goldsmith-Pinkham, Sorkin, and Swift, 2018; Jaeger, Ruist, and Stuhler, 2018) – I fix industry exposure shares at their 1990 level. Moreover, the industry averages of log wage growth are computed as leave-one-out measures to avoid mechanical correlation between the national trend estimate and city i wages (Borusyak, Hull, and Jaravel, 2020). See Appendix Section D.3 for more details on the construction of these shift-share shocks.

As an example of what might be driving common national industry trends, Eckert et al. (2019) note that the national wage growth in "skilled scalable services" like finance and communication industries in the last decades has been an important contributor to wage growth in U.S. cities where these industries are prevalent.[29] This trend, in part caused by a decline in communication costs, is an example of the identifying variation underlying the industry shift-share instruments .

To capture plausibly exogenous heterogeneity in the effect of this local exposure to national wage shocks on house prices, I interact B_{it} with a measure of local land unavailability for construction x_i^{land} from Lutz and Sand (2019).[30] This measure captures geographic constraints to marginal housing construction, which would be expected to increase the slope of the housing supply curve, and thereby increase the responsiveness of house prices to the wage shocks.[31]

Network instruments. Assuming that the shift-share "Bartik" shocks just defined represent plausibly exogenous shocks to local house prices in each city, a weighted average of these shocks occurring in *other* cities in a focal city's migration network should capture its exposure to exogenous house price changes in other cities. Put differently, those cities that have strong migration connections to locations that are more exposed to Bartik shocks will see greater exogenous house price growth in their migration network – which implies a greater spillover effect on their own house prices. Variation across cities in their network exposure to other cities' Bartik shocks can therefore identify the house price spillover effect η^{nw}.

As an example, consider a simple stylized migration network[32] between Los Angeles, Boise, Boston, and Portland (Maine) where Boise has a large migration exposure to Los Angeles for Boise, and Portland to Boston to Portland, while migration links across these pairs are negligibly small. Now, if a wage shock B_{LA} interacts with the relatively constrained housing supply x_{LA}^{land} in Los Angeles, driving up house prices, this would be expected to cause greater migration flows from Los Angeles to Boise, but only negligible flows to Portland. As a result, house prices in Boise increase, all else equal, while Portland is mostly unaffected. In the context of this example, the network IV approach uses the degree to which Los Angeles shocks lead to greater house price changes in Boise than in Portland to identify the size of causal spillover effects.

The instrument for network house price changes is constructed analogous to p_{it}^{nw} as a migration exposure-weighted average of network Bartik shocks and their interactions with land constraints:

$$bp_{it}^{nw} = \sum_{j:j\neq i} \psi^{ij}_{`90\text{-}`95} B_{jt} \cdot x_j^{land}, \tag{10}$$

where the migration network weights $\psi^{ij}_{`90\text{-}`95}$ are as defined in equation 8. Intuitively, the instrument bp_{it}^{nw} captures the predicted exogenous component of migration-weighted house price growth in city i's migration network.[33]

There is by now an extensive literature that instruments for local house prices using shift-share measures of local exposure to exogenous national productivity trends (Saks, 2008; Saiz, 2010; Paciorek, 2013; Mian, Rao, and Sufi, 2013; Guerrieri et al., 2013; Beaudry, Green, and Sand, 2014; Diamond, 2016). The innovation in this paper consists of applying those city-level shocks to *other* cities to identify exogenous house price changes in the focal city's *network* and their effect on

[29]See Appendix Figure A10 for a plot of wage trends by industry sector over time.

[30]These are comparable to the Saiz (2010) land availability measures commonly used in the literature. Lutz and Sand (2019) build on his methodology to expand the number of covered cities and, among other things, improve the measurement of land availability for overlapping city areas and coastal locations.

[31]Note that the Davidoff et al. (2016) critique of supply constraint instruments for house prices does not apply in this setting - see Appendix Section C.2 for further details on the exclusion restriction in this network IV approach.

[32]See Figure A2 for an illustration of this example.

[33]Analogous to the construction of the wage shocks B_{it}, and consistent with the definition of p_{it}^{nw}, I again hold the migration exposure shares fixed at a baseline period level, here consisting of average 1990-1995 migration shares.

the focal city's house prices through a spillover channel, rather than looking at the direct effect of the Bartik wage shocks on the focal city. As far as I know, there are few comparable papers that use a network approach to shift-share identification. The most similar ones include Bartelme (2018), who uses Bartik shocks to nearby cities as an instrument for a city's market access in a trade gravity model. Similarly, Baum-Snow, Hartley, and Lee (2019) use shift-share labor demand shocks to other neighborhoods that are accessible by commuting to identify the effect of changes in "resident market access" on life outcomes, and Severen (2021) uses commuting network links to labor demand shocks in other locations to identify model parameters. Moreover, Kotova and Zhang (2021) cite the identification strategy that I develop here and apply it in a different setting. However, I am not aware of any other work using migration links and shift-share shocks to other cities to identify spillover effects.

3.4 Network instrument identification

Industry-level reformulation of network IV. In order to make the discussion of the identification in my network IV approach more transparent, I first extend the results in Borusyak et al. (2020) to show how the distribution of industry wage shocks determines the spillover estimate. As I show in Appendix Section C.2, the IV estimator of the spillover coefficient can be written as

$$\hat{\eta}^{nw} = \frac{\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_\iota g_{\iota t}(\overline{p_{it}})^{\perp}_{\iota t}}{\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_\iota g_{\iota t}(\overline{p^{nw}_{it}})^{\perp}_{\iota t}},$$

Here, I have combined the industry exposure and migration network structure into a weight $s_\iota = \frac{1}{N}\sum_{i=1}^{N}\sum_{j:j\neq i}\psi^{ij}_{\text{`}90\text{-`}95}\tilde{\omega}_{\iota,j,\text{`}90}$ that summarizes the average across cities i of their migration network exposure to industry ι. The notation $\bar{\nu}_t$ denotes an average across cities, weighting them by their relative exposure to industry ι, and $g_{\iota t}$ is log national wage growth in industry ι. This estimator is equivalent to running an industry-level regression of weighted average house price growth (with weights given by industry exposure through migration networks) on weighted network house price growth, instrumenting for the latter with national industry wage growth.

This rewriting of the estimator in this way at the industry level clarifies the identifying variation underlying the network IV estimate: We can think of the spillover effect estimation as identifying the spillover effect from the degree to which the covariance in industry wage growth shocks with house price growth in the cities most exposed to the industry (the numerator) is higher in the cities that are "treated" in the form of having house price changes in their migration network that vary with industry wage shocks (the denominator). For instance, if cities with migration connections to technology hubs (e.g. Boise, ID) see higher house price growth precisely when tech hub house prices rise (e.g. in San Francisco) as a result of national trends in knowledge industry wages, then this variation allows us to infer a causal positive spillover effect.

This rewriting of the network IV estimator in the form of industry-level shocks then allows me to formulate the exclusion restriction of the network approach as follows:

Network IV exclusion restriction: *If the network instrument bp^{nw}_{it} is relevant and mild regularity conditions hold (the variance matrix of control variables has full rank, and the covariance matrices of instruments and residuals with controls are bounded and exist), then the IV estimate of the spillover effect $\hat{\eta}^{nw}$ is consistent if and only if*

$$\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_\iota g_{\iota t}\bar{\xi}^{P,\perp}_{\iota t} \to^p 0.$$

Here, $\bar{\xi}^{P,\perp}_{\iota,t}$ is the error in the house price growth regression in Equation 9, residualized with regard to the control variables Γ_{it}, and averaged over cities, weighting them by their migration network exposure to industry shocks.[34]

This condition shows that, for the network IV estimate to be consistent, industry wage shocks cannot be systematically higher for those industries that have a systematically larger migration network impact on cities that are experiencing large unobserved house price shocks, conditional on control variables.

As Borusyak et al. (2020) show, this identification allows for a city's migration network to be endogenously determined – it only requires the national industry trends over time to be exogenous in the sense defined in the exclusion restriction. This would be invalidated, if, for example, cities that experience more migration flows from cities that specialize in the booming tech industry are also systematically experiencing greater idiosyncratic house price movements in a way that is not captured by their own exposure to knowledge industries or any other included control variables.

"Double Bartik" control variable. As I am using the industry structure of other cities in the same migration network to instrument for network house price shocks, there may be a concern that the network house price instruments are correlated with focal city i industry shocks if industry structure is correlated across cities that share migration links. Note that this concern is supported by the significant coefficient on industry structure in the migration cost determinants analyzed in Section 2.2: Table 2 showed that migration costs appear to be lower among cities with similar industry structures, making them more likely to have strong migration links.

To address this issue, I include focal city i's direct Bartik shocks B_{it} and $B_{it}x^{land}_{it}$ as control variables in the regression, an approach which Chodorow-Reich and Wieland (2020) titled the "double Bartik" method. Controlling directly for the focal city's own Bartik shocks effectively orthogonalizes the network instrument with regard to any direct city i exposure to national industry wage shocks. To make this concrete, consider again the Boise, ID, example: While Los Angeles and Boise have a relatively similar industry structure, it is not identical. For example, L.A. is more exposed than Boise to industries in the "information" industry sector that includes motion pictures and broadcasting (NAICS code 51), which saw real wage growth nationally in 2000-2017 that was 15% higher than average.[35] This difference in exposure to shocks means that Los Angeles will see some increases in local house prices that are not correlated with Boise's direct exposure to industry wage shocks – and these effects will be larger due to L.A. being in the top quartile of land-constrained cities. These exogenous house price increases in L.A. that are driven by industry shocks which are *not* affecting Boise directly by construction are then used to identify the spillover effect on house prices in Boise due to its exposure to migration outflows from L.A.[36]

3.5 Dynamic effect estimation

To understand the time patterns of these spillover effects, I also estimate the effect of period t shocks on city i in future periods. I estimate IV forecasting regressions that correspond to local

[34]See Appendix Section C.2 for details on the derivation of this expression.

[35]Real wage growth (corrected for CPI-U growth) in NAICS 51 was 27% over the 2000-2017 period, while the employment-weighted average of real wage growth across sectors was 12%. See Appendix Figure A10 for an illustration of these industry wage trends.

[36]Note that it is not necessary for the Bartik direct shock controls to be a perfect measure of *all* local industry shocks: what is relevant is that it captures the direct local variation that results from the national shocks used to identify the house price variation in other cities.

projections with external instruments (Jordà, 2005; Stock and Watson, 2018) of the form

$$Y_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \beta' \Gamma_{it} + \xi_{i,t-1+h}^{P}, \tag{11}$$

where the dependent variable $Y_{i,t-1+h}$ can be a flow variable, such as period $t-1+h$ house price growth, or a cumulative variable, such as the total effect on the level of house prices $\sum_{k=1}^{h} p_{i,t-1+h}$ of period t shocks, or the effect on the level of population. The vector Γ_{it} includes the same additional period t control variables as the static model (see Section 3.6), and I am again instrumenting for the time t network price growth shock using network shocks. The coefficients η_h^{nw} now represent the impulse response in period $t-1+h$ of the shock. That is, the contemporaneous impact corresponds to $h=1$, the impact on the dependent variable in the year after the shock is $h=2$, and so on.

When reporting dynamic effects, I will cumulate impulse responses on house price growth and net migration to obtain total house price level and population level effects over time. However, effects on per-period flow variables like permits or mortgage lending will be reported as per-period effects.

3.6 Additional covariates

In addition to the IV strategy described above, I also try to address any potential omitted variables bias by explicitly controlling for covariates that might confound the house price spillover effect. I use the theoretical framework developed in section 3.1 (see equation 6) to provide guidance in identifying potential confounders. Moreover, I explore different specifications to ensure the robustness of the results. These variations are detailed below.

National house price trends. An extensive literature suggest that house prices might co-move in different cities due to national house price trends, for instance caused by changes in mortgage rates or business cycle dynamics as a result of credit supply changes (López-Salido, Stein, and Zakrajšek, 2017; Mian, Sufi, and Verner, 2019). I control for this possibility by including year fixed effects in the regressions.

Regional trends. There may also be regional differences in house price trends, for instance due to the timing of settlements in the historical evolution of the U.S., resulting in higher population mobility in general due to differences in the "rootedness" of local populations, as Coate and Mangum (2019) argue. Moreover, underlying house price trends may be caused by long-run shifts in preferences for regional amenities, such as the increased attractiveness of the sunbelt as a result of air conditioning (Glaeser et al., 2001). To avoid confounding my analysis of house price-driven migration spillovers with these trends, I allow for U.S. Census region-specific time trends and city fixed effects in house price growth.

Migration access. The static location choice model in section 3.1 suggests that the size of migration flows to and from a city should depend on the population size of cities in its migration network. That is, a given change in attractiveness of a city should result in a different migration flow if "migration access" in the form of the potential migrant population changes – analogous to the role played by market access in trade models (Donaldson and Hornbeck, 2016; Allen and Donaldson, 2018). To control for this size effect, I construct the empirical counterpart to migration access $\mathcal{M}_{i,t-1}$ implied by the model as $\mathcal{M}_{i,t-1} = \sum_k \phi_{'90\text{-}'95}^{i \leftarrow k} \Delta \ln L_{k,t-1}$, and include it in the regressions.

Wage shock effects. The attractiveness of a city to migrants is not just a function of housing costs, but also of income. In fact, if a shock to labor demand creates opposite and offsetting wage and housing cost effects, we might see a city becoming more expensive but without any change in migration as real wages remain constant (Moretti, 2013). Therefore, it may be important to control for wage changes in the migration network that result from the wage shocks that I am using to identify exogenous house price changes. To do so, I construct the wage shock analogue of the network house price shock term as the weighted network average of the Bartik wage shocks without interacting them with local housing supply constraints:

$$bw_{it}^{nw} = \sum_{j:j\neq i} \psi^{ij}_{\text{'}90\text{-'}95} B_{jt} \tag{12}$$

I include bw_{it}^{nw} as a control variable in my baseline specification. Note that I am not controlling for the actual wage changes in the network here because those will be endogenous both with regard to the migration flows of interest and with regard to house price changes that might trigger cost-of-living adjustments. Controlling for actual wage changes would therefore raise a "bad control" problem (Angrist and Pischke, 2008) and, in the worst case, remove the variation of interest.[37]

3.7 Data

As there is currently no long panel of annual U.S. migration data available that has information on different worker types, as well as sufficient sample size to accurately capture the network of city-level migration, I will limit the reduced-form analysis to the CZ-by-year level – implicitly treating workers as homogeneous. This section documents the main data sources to implement the reduced-form network IV approach described above. All data is crosswalked to 1990 commuting zones, aggregating county-level data using a crosswalk provided by Autor and Dorn (2013).[38]

Migration flows. I compute migration flows between commuting zones from county-to-county migration flows provided by the Internal Revenue Service (IRS) for 1990-2017. The IRS uses changes in the zip codes on individual income tax returns that were filed for the previous tax years to infer moves. Based on this methodology, the IRS is able to assemble a data set of the annual movement of tax returns and the number of exemptions – which correspond closely to the number of people – across counties, covering $> 90\%$ of the U.S. population.[39] This mobility data should correspond closely to the actual movement of people across counties - with small caveats (Gross, 2003).[40] I use this data to compute CZ-to-CZ flows and the share of people not moving in each year, omitting flows in and out of New Orleans from any totals.[41] The baseline in-migration

[37] I thank Cecile Gaubert for raising this issue with regard to an earlier draft.

[38] Available online at https://www.ddorn.net/data.htm.

[39] For instance, for the fiscal year 2014, the IRS received 149 million income tax returns - which correspond to tax households - on which 291 million personal exemptions were claimed (Source: https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-returns), which compare to ca. 319 million people in the U.S. in 2014.

[40] The small number of people who do not have to file income tax returns is not captured in this data, and in that number low-income people and the elderly tend to be overrepresented. At the same time, a small number of tax returns that are filed late - after September of the filing year - are not captured. The latter returns are usually granted extensions due to their complexity, which correlates with high income, so the mobility data may under-represent very high income returns as well. Moreover, changes in marital status may lead to a failure to match returns, as for joint filers only the primary taxpayer's social security number is captured, so that after a divorce, for instance, one of the former couple's social security numbers will not be associated with a corresponding county of residence for tax purposes for the prior year and thus gets omitted from the data set. It is also important to note that the IRS does not name counties if the pairwise flow consists of less than 10 tax returns, so small migration flows are censored. However, these flows are contained in the totals.

[41] The reason for excluding flows to and from New Orleans here and in all other analyses is that the city's loss of the majority of its population as a result of Hurricane Katrina in 2005 represents a large outlier that would distort all migration-related analyses and is mostly unrelated to the deliberate equilibrium location choices that I focus on in this paper.

weights $\phi^{i\leftarrow k}_{'90\text{-}'95}$ are calculated as the average inflow from city k to i in each year 1990-1995, as a share of the average population in city i in those years. Similarly, the share of people staying in a city $\mu^{ii}_{'90\text{-}'95}$ is the average number of people (tax exemptions) not moving away from city i as a share of the previous year's population – also averaged over 1990-1995.

For the construction of the city-to-city migration network, I only include continental U.S. commuting zones, i.e. not including flows to and from Hawaii, Alaska, Puerto Rico, or other U.S. territories. For migration totals, for instance to calculate gross inflows, outflows and net migration from each city, I take total migration to *any* destination and subtract foreign migration to obtain a measure of total domestic migration flows, albeit not restricted to the continental U.S., which cannot be separately identified in total flow data.

I also construct Covid-era migration measures from Census estimates of county-level population changes from March 2020 to July 2021, which I aggregate to measures of commuting zone population growth during the first 16 months of the pandemic.

House price growth. An annual panel of house prices is obtained from the Federal Housing Finance Authority (FHFA) at the county level for 1990-2017. Different counties have data starting in different years, but by 1990 house price indices are available for most of the counties in the U.S. The data are repeat-sales indices, so while the levels are not comparable cross-sectionally, they can be used to construct the house price growth terms relevant for the reduced-form analysis.

I first compute county-level log changes in house prices and then aggregate these growth rates to commuting zones as weighted averages, with weights derived from county populations in 2000. Moreover, to avoid bias from small counties entering the sample in later years as FHFA coverage improves, I fill in the county-level growth rates with state-level house price growth for CZ-years where less than 10% of the CZ population are in counties with available house price data. This imputation is a minor issue as CZs with more than 50% of their population being covered by FHFA reporting constitute 95% of the continental U.S. population by 1990 and 99% by the year 2000.[42]

The resulting balanced panel of 1990-2017 house price growth for all CZs is the dependent variable used in the reduced form regressions.

For one of the applications of the migration spillover approach, I additionally use monthly house price data for 2019-2021 in the form of Zillow's housing value index for all homes, which is obtained as a seasonally adjusted series at the county level from Zillow's website.[43]

Housing construction permits. New housing construction permits by commuting zone are constructed from county-level counts of permits available from the Census Building Permits Survey for 1990-2018, which I aggregate to 1990 CZs. Some counties do not report permit numbers. Thus, in cases where not all component counties of a CZ are reporting, I scale the reported numbers in proportion to the reporting counties' share of the CZ population to make up for the missing share, using county populations in 2000 as weights. To minimize the impact of any scaling, only CZ-years with counties reporting that correspond to at least 80% of the CZ population are included in the analysis.[44]

Wages. I use IRS tax data on total salary income by county to compute average salary income per capita for each commuting zone as a measure of average wages.

[42]In unreported regressions, I have also verified that the main results of the reduced-form analysis are robust to omitting all CZs that did not have more than 50% FHFA reporting coverage for their population for the entire 1990-2017 sample period.

[43]URL: `https://www.zillow.com/research/data/`

[44]This is a minor issue, as <8.6% of county-years are part of CZs that have <99% of their populations accounted for. In any year, at least 717 CZs have at least 80% of their total population fully covered by permit reporting.

Mortgage lending. In order to measure the effect of migration on mortgage market activity, I use loan-level data on mortgages provided under the Home Mortgage Disclosure Act (HMDA). For each year and borrower county, I retain all the loans that are originated for home purchases (excl. refinancing and other loan purposes) and aggregate total lending volume in dollars, as well as the count of new originated purchase loans to the CZ level. Moreover, for the construction of credit shocks (see details in Section 4.2 for details) I also use data on the volume of applications for purchase mortgages by county in the HMDA data.

Vacancy data. I obtain vacancy rates from the Census Housing Vacancy Survey and compute changes in average vacancy rates for homeowners and renters between a baseline period (1991-1997) and the 2000-2007 and 2012-2017 boom periods in my analysis. This comparison can be computed for 98 CZs, which have data for all time periods. I group these cities into those with above- and below-median housing supply constraints based on their share of land unavailable for construction (Saiz, 2010; Lutz and Sand, 2019).

4 House Price Spillover Effects

In this section, I first present the baseline results on house price spillovers between cities. Second, I show that the expected effect pattern holds for net migration to be an important transmission channel, and that other city-level variables behave in line with housing demand increasing as a result of migration. Third, I demonstrate that migration spillovers are quantitatively important by estimating their ability to explain the cross-sectional variation in house prices in the run-up to the housing boom of the 2000s, and by showing that migration links are good predictors of house price correlations between city pairs.

4.1 Migration spillover effects on house prices: baseline estimates

Recall that I am estimating equations of the form

$$p_{it} = \alpha_i + \alpha_t + \eta^{nw} p_{it}^{nw} + \beta' \Gamma_{it} + \xi_{it}^{P},$$

where Γ_{it} represents various control variables capturing characteristics of the focal city or its migration network. Moreover, in the baseline specifications I will only allow for CZs to be included in one another's migration network, if they are at least 150 miles apart.

Network IV first stage. Before proceeding to the causal effect estimation, I verify that the network house price shocks (wage shocks interacted with supply constraints) affect network house price changes in the expected manner – the "first stage" of the estimation. That is, we are interested in the effect of the instrument bp_{it}^{nw} on the migration-weighted index of actual house price changes in each city's network p_{it}^{nw} across the years 1991-2017, after residualizing all variables with regard to the full set of control variables. The corresponding first-stage coefficient estimate is shown in column 5 of Table 4. Note that the relationship is strongly positive and significant.[45] The first-stage F-statistics for the instrument are reported at the bottom of the main IV results in Table 4, and they are large in all specifications based on conventional thresholds. Note that this is the case even though we are orthogonalizing the instrument with regard to direct effects of the wage shocks used in its construction on the focal city itself. This means that there is sufficient variation in these shocks between cities that we can identify significant variation in house prices in

[45]The same relationship is also shown visually in Appendix Figure A3 as a binned scatter plot of the residuals of the instrument and of the network house price changes.

the other cities due to their exposure to the national shocks which is uncorrelated with the direct effect of the same national shocks on the focal city.

Network IV baseline estimates. Table 4 shows the network spillover effects obtained by estimating equation 9 with IV-GMM, using with the network house price shocks to instrument for network house price changes. Column 1 uses the minimal set of control variables, including only year and CZ fixed effects. Column 2 adds in the full set of control variables described in the previous section, which control for concerns around non-spillover effects on house prices and orthogonalize the network spillover instrument with regard to the direct effect of national shocks on local house prices. The IV coefficients are positive, statistically significant and sizeable in these specifications. For comparison, column 4 shows the OLS parameter estimate of the network spillovers, which is about half as large as the IV estimate in column 2. This negative bias in the OLS could for instance be explained by network house price increases being in part caused by positive shocks to other cities' amenities, which will lead to smaller ouflows of migrants to the focal city, inducing a downward bias in the estimated house price effect, which the IV corrects.

To interpret the effect sizes, note that the network house price term scales house price changes in other cities by migration exposure weights $\psi^{ij}_{`90\text{-}`95}$, which represent the degree to which the local population in i is affected by migration flows from other cities and the sum of which may vary across locations. If a city's population was fully replaced every year by migrants from other locations, these exposure weights for that city would sum to 100% whereas a city with minimal migration flows from other locations would have exposure weights that sum to close to zero. In the estimation sample at hand, the average city has migration exposures that sum to 2.4%, with an interquartile range of 0.3%-3.5%. Focusing on a city with average migration exposure, the estimate in column 2 means that a 10 ppt shock to house prices in all the other cities would result in a 10 ppt $* 0.024 * 11.37 \approx 2.7$ ppt increase in house prices in the focal city. That is, the short-run passthrough for the average city of house price shocks everywhere else is about 27%.[46]

Dynamic house price effects. The estimated time pattern of OLS and IV estimates of period t shock effects on the level of house prices is shown in Panels a and b of Figure 2. The graphs show the cumulative effect of period t network house price changes on focal city house prices at different horizons h, where the $h = 1$ estimates correspond to contemporaneous price effects shown previously in columns 2 (IV) and 4 (OLS) of Table 4. As the graph shows, the cumulative IV spillover effect increases for two periods past the initial impact, and then declines somewhat before settling at a medium-term effect in year 5 that is about 50% higher than the immediate impact. The 5th-year effect of a spillover shock in the is 16.8 and only marginally statistically significant as longer-run effects are more noisily estimated.[47] This means that an exogenous house price change of average size 10 log points in city i's migration network causes a $10 * 0.024 * 16.8 \approx 4.0$ log point increase in the house prices of a city with average migration exposure. This highlights that effects operating through a migration channel may have a delayed impact as frictions prevent workers from moving instantaneously. However, the shape of the dynamic effects graph suggests that most of the long-run spillover effect has been realized 1-2 years after the initial shock.

Heterogeneity by migration exposure. The effective size of the passthrough from shocks in other cities to house price changes in the focal city depends on its migration exposure: as all the migration network terms weight shocks in other cities by the migration weights, cities with

[46]Note that these reduced-form effects do not necessarily capture the full general equilibrium effect of shocks, as a shock that spills over to city i could then have secondary spillover effects to cities that are exposed to migration as a result of city i house prices.

[47]See column 1 of Appendix Table A3, which shows the numerical estimates for the long-run effects plotted in Figure 2.

greater migration exposure will have a larger change in their weighted network house prices and a larger spillover effect for the same estimated coefficient. This heterogeneity can be substantial (see Appendix Table A4): the short-run and 5-year effects of a 10 log point shock in all the other cities on focal city house prices can range from 0.3 and 0.5 log points at the 25th percentile of migration exposure to 4.0 and 5.9 log points at the 75th percentile.

Heterogeneity in spillover effects. The effect of the spillover shocks on each city might also differ based on the degree to which the focal city has elastic housing supply (Glaeser et al., 2008). To test whether there is heterogeneity in this regard, I allow the network spillover effect to vary by city in the form

$$\eta_i^{nw} = \eta_0^{nw} + \eta^{land} x_i^{land},$$

where x_i^{land} is the city share of land unavailable for construction, which proxies for geographic constraints to housing supply (Saiz, 2010; Lutz and Sand, 2019). The coefficient on the interaction term should be positive if the housing supply function in response to spillover shocks is steeper for cities that are more land-constrained.

The results are shown in column 3 of Table 4. Here, I include the same control variables as in the baseline specification, with the exception of the focal city wage shock interaction with land constraints, which is a "bad control" in this specification that would remove most of the heterogeneity in responsiveness to shocks that we are interested in. [48] The coefficient on the constraint interaction term is positive and significant at a 1% level. That is, cities that are more constrained experience stronger house price responses to network spillovers. To interpret the magnitudes of this heterogeneity, note that it implies that, for a city with average migration exposure, a one standard deviation change in the unavailable land share of ~ 22 ppt (see summary statistics in Appendix Table A1) would increase the short-run spillover elasticity by about $0.22 * 8.87 * 0.024 \approx 0.05$, or about 26% higher relative to the effect on an unconstrained city. Similarly, the 5-year effects for an average migration exposure city can vary over the interquartile range of land constraints from 4.1 to 5.1 log points (see Appendix Table A4).

The house price effects over time when allowing for heterogeneity are shown in panels (c) and (d) of Figure 2. The results show that constraints significantly increase the relative house price response even more in the medium-run than in the short-run. In contrast, unconstrained cities seem to be able to accommodate some of the migration inflows in the medium-term and by the 5th year have given back some of the house price increases from the first few years after the network shock. This is in line with the intuition that, while both constrained and unconstrained cities experience congested construction and housing markets in the short run, land constraints additionally limit cities' ability to adapt in the long run by expanding their housing stock.

4.2 Robustness checks

Alternative house price shocks. One concern with the network IV design used in the baseline specification may be that using shift-share shocks based on national wage trends and supply constraints as the shock to house prices in other cities leaves open the –arguably remote – possibility of an unobserved house price shock structure *not* driven by migration flows that somehow aligns

[48] That is, I cannot precisely identify the differential effect of network spillovers, while holding direct wage shocks and the differential impact by land availability of direct wage shocks constant – and it would be difficult to interpret such an estimate. In fact, when I control for this interaction, only the uninteracted effect coefficient of network shocks remains significant, while the interaction is estimated not to have an additional effect, or a weakly negative effect.

with the focal city being migration-linked to cities experiencing house price changes driven by these Bartik shocks but is uncorrelated with its own direct exposure to these Bartik shocks and not captured by CZ fixed effects or regional trends. Moreover, an important feature of the migration network approach is that it can be applied to study the spillovers of a variety of different location-specific shocks. To eliminate the afore-mentioned concerns and show the versatility of this network identification approach, I therefore repeat the baseline analysis with a different instrument for house price shocks in other cities.

For this alternative instrument, I construct credit shocks following Loutskina and Strahan (2015). These shocks are shift-share instruments based on national variation in conforming loan limits for mortgages that are eligible for securitization by the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac. When more loans become eligible for GSE securitization, they are effectively subsidized. Therefore, the local share of loan applications near the CLL in the previous year can be used as a proxy for exposure to the loosening of credit that results from increases in the CLL at the national level. I use the HMDA loan application data for each county to construct the $CreditShock_{it}$ as the product of the national % growth in the CLL from year $t-1$ to year t and the local % of loan applications in year $t-1$ that were within 5% of the CLL that applied in year $t-1$. Unfortunately, the national CLL does not vary between 2006 and 2016, so this method limits the sample to the 1991-2006 period.

Analogous to the baseline approach based on Bartik shocks to house prices, I construct the migration network exposure-weighted average of these credit shocks in other cities as an instrument for house price changes in migration-linked cities and control for any direct effects of the credit shocks on the focal city. That is, this instrument identifies the spillover effects off variation in house prices in other cities caused by mortgage credit shocks that are orthogonal to direct credit shocks to city i. [49]

The results of the estimation using this version of the network instrument are shown in columns 1 and 2 of Table 6. Both the average spillover effect estimate in column 1 and the effects allowing for supply heterogeneity in column 2 are smaller than the comparable estimates in columns 2 and 3 of Table 4 obtained using the Bartik wage shock network instrument. However, they are close enough in magnitude that I cannot reject the hypothesis that the parameters estimated using the two approaches are the same.[50] The fact the finding of sizeable and significant spillover effects does not rely on the particular instrument used to identify network shocks both supports the robustness of the finding and shows how this approach could be applied in a variety of settings to estimate geographic spillovers.[51]

The role of migration distance. One concern with the baseline specification may be that I am restricting the migration network for each city to not include other cities that are less than 150 miles away. As the size of migration flows is in part driven by distance (see the estimates in Table 2), this may preclude a substantial share of migration flows from being considered in these estimates. Moreover, as distance has a differential effect on different education groups,[52] the migration network constructed from longer-distance flows may select for the effect of particular

[49]That is, in addition to the full control variables included in the baseline regressions I now include $CreditShock_{it}$ as an additional control variable in the regressions

[50]Formally, the z-scores comparing the coefficients in the two tables computed as $Z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}}$ (Paternoster, Brame, Mazerolle, and Piquero, 1998) are 1.2 for the average effect, 0.57 for the intercept with heterogeneity, and 0.85 for the supply heterogeneity coefficient, corresponding to p-values of 0.12, 0.28, and 0.20.

[51]Data on migration links between commuting zones can be readily obtained from the author by request via email.

[52]Compare Columns 2 and 3 in Table 2.

workers' migration.

To analyze the role of expanding the migration networks to include short-distance flows, I repeat the baseline effect estimation of column 2 in Table 4 for migration networks that include flows at *all* distances - expanding the network to allow for short-distance links. The results of this exercise are shown in columns 3 and 4 of Table 6. The spillover effects for short-distance migration networks continue to be significant as more cities are included, but decline in magnitude. As would be expected, even though the sample size increases as the migration network is expanded to include nearby cities, the strength of the instrument reflected in the 1st-Stage F-statistic is smaller when all cities are included, perhaps because the orthogonalization with regard to local shocks in the CZ of interest also purges some of the network shock variation coming from nearby cities. These findings are aligned with the original rationale for excluding short-distance migration: spillover effects between neighboring cities may operate through a variety of channels that do not require migration, as their labor markets and housing markets may partially overlap. Therefore, the effect of interest which operates through migration spillovers is more precisely identified at a distance that ensures that workers would have to move residences in order to respond to economic shocks in other cities.

City size. While the baseline estimation treats each commuting zone equally as an observation, one may wonder whether the spillover effects are only driven by smaller cities for which shifts in migration flows might be more likely to congest local housing markets. To test this idea, I re-estimate the spillover parameters, weighting each commuting zone observation by the log of its population. The results, in columns 5 and 6 of Table 6, show that the average effect with these weights is almost identical to the baseline estimate, albeit less precisely estimated, while the results allowing for heterogeneity show that the effect is now almost entirely driven by the significant spillover effect on supply-constrained cities. This suggests that, for larger cities, housing supply constraints play a bigger role in whether migration spillovers affect local house prices.

Conversely, one might be concerned that large cities have more migration links to distant cities and are therefore more likely to have housing markets that are affected by long-distance spillovers through migration flows. To test this idea, the estimates in columns 7 and 8 of Table 6 exclude the twenty largest commuting zones from the data – and I find that the parameter estimates are very similar to the baseline. That is, the baseline results are very robust to excluding the largest U.S. metropolises from the sample.

5 Testing the Spillover Mechanism

If migration flows and their impact on housing demand are the mechanism through which house price changes propagate, we should be able to observe spillover effects on other characteristics of city i in line with the house price effects documented in the previous section. This section documents that house price spillovers are accompanied by large effects on migration and credit variables in affected cities that support the proposed spillover mechanism.

5.1 Spillover effects on population

The theory in section 3.1 predicts that the spillover effects operate through net migration flows out of cities becoming expensive as people choose to substitute towards migration-linked cities that are more affordable. I can test this prediction of how the mechanism operates by estimating the cumulative network spillover effect on population in the focal city as a result of net migration.

This approach is analogous to the way in which we estimated the cumulative house price effects, but with city i's cumulative net migration rate in the IRS data as the dependent variable.[53]

The estimates of the contemporaneous effect of network house prices on population changes as a result of net migration are shown in column 1 of Table 5. In line with the model prediction that the network house price effects are transmitted through migration, I find that the network house price shock effect on focal city population is positive and significant. The coefficient estimate for the contemporaneous effect on net migration is 3.57, which implies that a 10 ppt increase in house price growth in a city's migration network would drive a contemporaneous increase in the net migration rate for a city with average migration exposure of $10 * 0.024 * 3.57 \approx 0.9$ ppt. Note that the significant positive effects on migration in the predicted direction also strengthen the case for the exclusion restriction of the network IV approach. Many concerns about unobserved joint variation in house prices across migration-linked cities imply that the focal city is becoming expensive at the same time as other cities for reasons unrelated to migration flows. However, if that were the case there would be little incentive for people to move *towards* the focal city when its house prices increase. [54] The fact that we observe people moving towards the cities becoming expensive provides additional evidence that the causal chain likely runs from migration flows to house prices.

Long-run effects on population. The dynamic effects on population over time (cumulative sums of log population changes from net domestic migration) are shown in panel (a) of Figure 3. The graph shows that a 10 log point shock to network house prices in period t leads to an increase in local population by $10 * 0.024 * 9.1 \approx 2.2$ log points 4 years after the shock for a city with average migration exposure.[55] This means that the long-run population effect more than doubles the contemporaneous effect shown in Table 5. Again, this effect is predicted to vary with a city's migration exposure, with the interquartile range corresponding to the 5th-year effect on population ranging from 0.3 to 3.2 log points. The pattern of effects over time in Figure 3 (panel A) shows that most of this increase has been realized 1-2 years after the initial shock, with smaller increases thereafter, in line with the time pattern of house price effects.

Reasons for moving. As additional qualitative evidence of why workers move between cities, I consider evidence from the Current Population Survey in Appendix E.4, which explores the stated reasons for moving among migrants in the U.S. I find that, in general, housing reasons dominate among within-county migrants, while housing is stated with similar frequency as family or employment among moving reasons for inter-county migrants. Moreover, migrants moving for housing reasons are especially prevalent during the housing boom of the 2000s, and migration for housing reasons is an important driver of the variation in inter-county migration over time – in line with the migration spillover mechanism proposed in this paper.

5.2 Benchmarking the implied housing supply elasticity

Implied housing supply elasticity. How reasonable are the estimated responses of population relative to the house price effects of the migration network shocks? If we assume that net migration

[53]Note that migration flows are sufficient but not necessary for the spatial location choice mechanism to operate as suggested. In theory, if all cities were perfectly supply-constrained, any additional demand to migrate there would raise house prices until the excess demand by migrants was eliminated. That way, house price changes could propagate in equilibrium even with only modest or no migration flows being observed.

[54]Even when the price increase is caused by a change in amenities, the migration effect of a simultaneous change in amenities in the migration network and the focal city would be ambiguous.

[55]See column 3 of Appendix Table A3 for the source of the coefficient point estimate of 9.1.

is the main transmission channel of house price growth between cities in the same migration network, we can compare the effect estimates for net migration and house price growth to get a rough estimate of the implied elasticity of house price growth with regard to population changes, i.e. the implied slope of the inverse housing supply curve. Comparing the contemporaneous effects, I obtain implied elasticities of house price growth with regard to population growth of about $11.37/3.57 \approx 3.2$ in the short run,[56] and $16.77/9.09 \approx 1.8$ by the 5th year[57] The smaller house price elasticity in the medium term corresponds intuitively to the idea that supply curves are generally more elastic with longer horizons, as new housing construction happens increase with a delay but over time is able to accommodate some of the increase in housing demand (see the evidence on construction effects in section 5.3). To benchmark these estimates, we can compare them to historical relationships between migration and house prices and other estimates of supply elasticities in the real estate literature.

Historical migration effects on house prices. To put the implied price elasticity in context, I first compare the magnitudes to the cross-sectional relationship between migration and house price growth during recent housing cycles (see Appendix Figure A6). Comparing the change in average net migration rates and house price growth for different periods to their baseline values in 1991-1999 a 1 log point increase in the net migration rate was on average associated with increases in house price growth of 4.1 log points during the 2000-2007 boom, 2.1 log points during the bust 2008-2012, 1.9 log points during the 2012-2017 boom, and 2.8 log points during the 2000-2017 period as a whole. While these raw correlations are not directly comparable to my causal effect estimates, they suggest that the implied house price elasticity with regard to net migration of my spillover IV estimates is very close to historical relationships between these variables.

Other supply elasticity estimates in the literature. Comparing these inverse supply elasticity estimates to other findings in the literature, they are towards the upper end of the range of long-run inverse housing supply elasticities found in the seminal studies by Saiz (2010) and Diamond (2016). However, their analyses are not directly comparable as (1) I estimate 1-5 year inverse housing supply elasticities where they use decadal intervals, and (2) my sample includes the large increase in house price volatility during the boom-bust period of the late 2000s and early 2010s. Both of these differences should lead the inverse housing supply elasticities to be higher than those in these older studies, suggesting that my findings are quite consistent with their estimates. In fact, my inverse supply elasticity estimates are towards the *lower* end of those implied by the more sluggish short-run supply responses found by Gorback and Keys (2020) for U.S. cities over the recent decade. Overall, the implied strength of the transmission mechanism through which migration flows translate into house price shock spillovers seems of a plausible magnitude when compared to recent cross-sectional patterns and other findings in the literature.

5.3 Mortgage credit and construction effects

To highlight the key channels for how extensive margin increases in population could affect the housing market, this section estimates spillover effects on local mortgage lending and construction. I show that mortgage credit increases as a result of migration spillover shocks and that the dynamics of construction rationalize the observed pattern of house price increases.

Mortgage lending. To measure the effects on mortgage markets, I compute the number and dollar volume of purchase mortgages originated annually in each CZ in HMDA data. Columns 3

[56] See column 2 of Table 4 and column 1 of Table 5.
[57] See columns 1 and 3 of Table A3.

and 4 of Table 5 show the effects of the network house price shock on the log of housing purchase mortgages originated, as well as log purchase mortgage lending volume.

The results show that migration network spillovers significantly increase local lending activity, with a ten log point increase in network house prices leading to a $10 * 0.024 * 1.9 * 100 \approx 46$ log point increase in the number of purchase loans originated in a city with average migration exposure. Similarly, mortgage lending volume in USD increases more than proportionally by $10 * 0.024 * 2.32 * 100 \approx 56$ log points in response to a 10 log point network house price shock, suggesting that average loan size is going up. In part this can be attributed to the fact that house prices are rising as a result of the network spillovers, but it is also consistent with other explanations, for example that migrating out-of-town buyers overpay as a result of being less well-informed (Chinco and Mayer, 2016).

Is this a plausible effect size? We can benchmark this effect by comparing it to the spillover effects on migration. In my HMDA sample, an average city and year see 1.1 mortgages originated per 100 residents. This means that the short-run increase in population of 0.9% that is estimated to result from a 10 log point network shock, multiplied by a U.S. homeownership rate of 67% in 2020, and divided by an average household size around 2.5,[58] should result in new loans corresponding to about $0.9 * 0.67/2.5 = 0.24$ loans per 100 residents, or a 22% increase over existing loan volume.[59] Of course, lending to existing residents would also be expected to increase as a result of the impact of spillover shocks on house prices and within-city moving activity, and the total effect on the level of population in the medium-run is more than twice as large as the short-run estimate used in this calculation. Taking into account these possible variations, the mortgage origination effect size appears plausible.

The dynamic effects on housing credit variables also behave in line with the migration and house price dynamics, with Panels (c) and (d) of Figure 3 showing the effect on new purchase mortgage originations and mortgage lending volume at different year horizons. The credit effects persist for several years but decline in magnitude over time, in line with the migration flows being larger in the years right after the shock. The effect on lending volume diminishes less by the 5th year than the effect on loan originations, consistent with the fact that house prices remain elevated, likely increasing loan sizes persistently.

Construction. Additional evidence on the changes in housing demand in response to network house price shocks comes from effects on the construction sector, measured by changes in construction permits issued. Column 2 of Table 5 and panel (b) of Figure 3 show that spillover shocks significantly increase construction. The short-run effect on annual housing unit permits of a 10 log point network house price change is about 56 log points ($\approx 75\%$) for a city with average migration exposure, and 51 log points ($\approx 67\%$) after 5 years .[60]

While this might seem large, it is important to put it in perspective on a per capita basis: over the 1990-2017 period, the Census reported an annual average of around 0.46 housing permits per 100 U.S. residents.[61] So, if net migration tends to increase the local population by ∼2.2 ppt in

[58] This analysis assumes that the additional population has average characteristics and that all migrating homeowners take out new mortgages. Average characteristics are from Census data at `https://www.census.gov/housing/hvs/files/currenthvspress.pdf` and `https://www.census.gov/data/tables/time-series/demo/families/households.html`.

[59] This may be an underestimate of the impact, as movers in the Current Population Survey for 1990-2017 are less likely to be married than non-movers (30% vs. 42%) and therefore might have a smaller family size.(Source: Author's calculation from CPS March samples obtained from IPUMS.)

[60] Computed as $10 * 2.35 * 0.024 \approx 56$ and $10 * 2.14 * 0.024 \approx 51$.

[61] Calculated by dividing total units permitted 1990-2017 from `https://www.census.gov/construction/bps/`,

an average exposure city by year 5 as a result of a 10 log point network house price shock, and applying an average household size of 2.5, this implies a need for $2.2/2.5 = 0.88$ additional new units per 100 residents, or $0.88/0.46 = 191\%$ of average permitting volume. Spreading this additional construction volume over 5 years, for example, the average city would have to increase annual construction by 38% to accommodate the increase in population – which accounts for more than half of the estimated increase in construction even without allowing for additional construction for existing local residents that might result from the house price increases and moves triggered by the inflows.

The relatively large increases in construction required–relative to normal building volume– for even modestly-sized population changes also rationalize why I find large house price effects of migration spillovers even in less geographically constrained cities: if a 2.2 ppt increase in population corresponds to a $> 38\%$ increase in construction volume over 5 years, it does not seem unreasonable for house prices to increase by 4.0 log points as the construction sector moves up its supply curve.

Housing vacancies during housing booms. The claim that congestion in construction markets plays an important role in understanding the propagation of house price increases across cities through migration contrasts with some narratives about how "overbuilding" was an important feature of the housing boom of the 2000s.[62] To provide additional evidence that congestion in construction markets plays a role in understanding the cross-section of housing booms, I therefore look at the relationship between vacancies and house price growth.

If migration congests housing markets, vacancy declines should be associated with a greater house price increase – and more so in more supply-constrained markets because they take longer to produce enough housing to accommodate the population increase. In contrast, under a narrative where overbuilding occurs during a boom and eventually attracts in-migration, vacancies should be positively associated with house price growth – and more so in constrained cities, as building is less responsive to a given house price increase.

To test these predictions, I compute the change in vacancy rates and house price growth for housing boom periods relative to a baseline period, and compare more or less land-constrained cities.[63] The graphs of the relationship between vacancies and house price growth are shown in Figure 4 for owner-occupied housing (rental housing results are shown in Appendix Figure A8).

Vacancies are systematically lower in cities with higher house price growth in both periods - and this relationship is statistically significantly more negative in constrained cities. That is, the congestion effect of in-migration is worse if housing supply is harder to come by. This means that the evidence from vacancies is more consistent with migration causing house price growth through a shortage of housing supply than a speculative boom that leads to overbuilding which then entices migrants.[64]

---

by total U.S. population from `https://fred.stlouisfed.org/series/POPTOTUSA647NWDB`.

[62]As an example of this view, see e.g. Haughwout, Peach, Sporn, and Tracy (2012) (p. 70): "*While it is now clear that too much housing was built in the United States in the boom phase, identifying how much and where overbuilding occurred remain important issues...Our results suggest that 3 to 3.5 million excess housing units were produced during the boom.* "

[63]The baseline period is 1991-1997 and the boom periods are 2000-2007 and 2012-2017. This comparison can be computed for 98 CZs, which have data for all time periods. I group these cities into those with above- and below-median housing supply constraints based on their share of land unavailable for construction (Saiz, 2010; Lutz and Sand, 2019).

[64]Note that I am not saying that those highlighting the national increase in vacancies during the housing boom (e.g. Haughwout et al. (2012)) are incorrect. It simply turns out that, in the cross-section, the highest vacancy rates were not in the cities where the boom was happening, making it unlikely that speculative overbuilding was driving the vacancies - a fact that is obscured when looking at national averages. See Sumner and Erdmann (2020)

Time-to-build delays. Another piece of evidence highlighting the significance of short-run congestion in construction markets is that time-to-build delays for residential housing construction increase during housing booms. This empirical pattern is shown in Appendix Figure A9. Using Census Survey of Construction data to estimate average time-to-build from start to completion of construction projects completed in a given year for different regions, I find that all U.S. regions show increasing construction delays during the 2000s housing boom, falling delays during the Great Recession, and then increasing delays again during the recent 2012-2017 run-up in house prices.[65] The existence of such time-to-build delays in housing construction provides an additional explanation for why housing supply does not immediately accommodate the migration spillovers on housing demand.

6 Applications: Importance of migration spillovers for house price dynamics

The migration spillover mechanism presented in this paper is general, in the sense that *any* shock to an origin city's housing markets can propagate through migration flows to other cities. This section considers the implications of this insight for understanding house price dynamics in three applications: (1) Estimating the effect of a credit supply shock on house prices. (2) Explaining geographic differences in house price betas. (3) Predicting the across-city housing market effects of the Covid pandemic.

All three applications show that taking into account spillovers through migration exposure improves the explanatory power of the standard approach–and that spillovers through migration channels can easily be incorporated into many applications in finance, as well as urban and regional economics.[66]

6.1 Credit supply shocks and house prices

To get an estimate of how important migration spillovers are for understanding the effects of local shocks to house prices, I first demonstrate their use in the case of quantifying the effects of a shock to house prices from the finance literature. I again use the credit shocks to house prices already discussed in section 4.2. The implementation used here is based on Loutskina and Strahan (2015) but similar identification approaches based on changes in conforming loan limit have been used in other recent studies of housing markets (e.g. Adelino et al. (2015); Greenwald and Guren (2021)).

The credit shocks, which result from national variation in conforming loan limits for GSE mortgages interacting with the degree to which these limits are binding in different locations, generate house price shocks at a city level. Assuming that these shocks are orthogonal to other city characteristics, equation 5 in the location choice model suggests that the resulting variation in city i house prices would be

$$p_{it} = \gamma^{cs} CreditShock_{it} + \eta^{nw}\gamma^{cs} \sum_{j:j\neq i} \psi^{ij} CreditShock_{jt} + \xi^{P}_{it} \tag{13}$$

for a similar argument.

[65] In a structural model calibration, Oh and Yoon (2020) show that these time-to-build increases at the height of the 2000s housing boom can be attributed to construction bottlenecks, whereas the continuation of delays at the beginning of the bust arises due to investor uncertainty about the housing market.

[66] Measures of migration exposure are easily computed based on the methodology in this paper, and will be made available by the author via email or link on his website.

where γ^{cs} is the direct effect of the shock on local house prices. Note that this expression implies, however, that there is additionally an indirect effect that results from a city being linked to other cities that are experiencing the same type of shock. To the degree that these migration exposures to shocks in other cities are correlated with direct exposure to the credit shocks, this may introduce omitted variable bias into any estimation of the effects of the credit shocks that does not take the geographic spillovers into account.

In Table 7, panel A, I estimate the degree to which taking into account the migration spillovers matters: column 1 first regresses house price growth in a 1991-2006 panel of commuting zones on control variables consisting of CZ and year fixed effects, as well as the fraction of GSE loan applications near the conforming loan limit, to establish the benchmark R-squared. In column 2, I then add the direct credit shocks and find a significant positive effect on house prices.[67] Comparing the two columns, the direct credit shocks incrementally explain about 2.5% of the variation in house price growth.

In column 3, I then test the importance of geographic spillovers: adding in the migration exposure to credit shocks,[68] increases the explanatory power of the regression by 1.1 ppt (or 44% of the direct effect's contribution) and the indirect effect through migration exposure has a significant positive effect on house price growth. Moreover, the estimated coefficient for the direct effect is reduced by a quarter, perhaps due to the migration exposure term correcting the omitted variable bias mentioned above. This shows that accounting for migration exposure significantly changes the estimates of the effect of credit shocks on house price growth – suggesting that such spillovers should be taken into account when estimating the effect of shocks that affect many locations at once.

While the estimates in column 2 distinguish the direct and indirect channels for credit shock effects, in practical applications it may be useful to be able to estimate *the* effect of credit shocks as a single parameter. Equation 13 shows that we can use our previous estimates of the migration spillover parameter η^{nw} to construct a model-implied total effect of credit shocks on city i that combines direct and indirect channels:

$$CreditShock_{it}^{comb} = CreditShock_{it} + \widehat{\eta^{nw}} \sum_{j:j \neq i} \psi^{ij} CreditShock_{jt},$$

where I use the spillover effect $\widehat{\eta^{nw}}$ estimated in Table 4, column 2. Under the assumption that the estimated spillover parameter is correct and that the model is correctly specified, the effect of this combined credit shock on house price growth gives us an estimate of γ^{cs}. Estimating this coefficient in column 4 of Table 7, panel A, I find that the estimated combined credit shock effect is smaller than the baseline effect and closer to the one controlling flexibly for migration exposure. Moreover, it is more precisely estimated than the original credit shock effect that did not take migration exposure into account. Using this model-informed combined credit shock exposure term therefore seems preferable to only using local direct shocks when estimating the size of the effect of mortgage credit on house price growth.

[67] Note that the coefficient is comparable in magnitude to that found in the original study by Loutskina and Strahan (2015), but does not match it exactly, as I am using a slightly longer sample (starting in 1991 rather than 1994), and different geographies (CZs rather than CBSAs).

[68] To be precise, this is defined analogous to the migration exposure expression in equation 7 as $\sum_{j:j \neq i} \psi^{ij} CreditShock_{jt}$, where all cities j are at least 150 mi. away from city i.

6.2 Explaining geographic differences in house price betas

Returning to one of the motivating questions for this paper, how do the migration network spillovers help us to explain the variation in housing market betas shown in in Figure 1?[69] For any given national shock to housing markets, e.g. a shift in housing demand as a result of changes in interest rates (Himmelberg et al., 2005; Glaeser et al., 2012), a city may experience a large local effect on house prices because of a greater local exposure or responsiveness to the shock, e.g. because of local supply constraints.[70] However, in addition, some cities may experience large effects from a national shock because of their migration exposure to the large direct effect of the shock on *other* cities. Thus, cities that are more exposed to migration spillovers from supply-constrained locations, for example, may have a higher beta with regard to the national house price cycle than similar cities that are not exposed to the housing markets which are experiencing large direct effects from national shocks.

To quantify this explanation for differences in housing market cyclicality, I first regress local housing market betas for 1990-2017 on local land constraints as a proxy for the responsiveness of local house prices to national housing demand shocks in column 2 of Table 7, panel B, controlling for log city population. Relative to the baseline of including only the control variable in column 1, the land constraints incrementally explain 16.9 ppt of the variation in betas across commuting zones.

To explore whether knowing a housing market's migration exposure helps us better understand its cyclicality, column 3 adds the long-distance migration exposure to land constraints in other cities as an explanatory variable.[71] This migration exposure has a significant effect on betas – cities with greater exposure to supply constraints elsewhere have higher betas – and adds 7.5 ppt in explanatory power (or 44% of the direct effect's contribution). Analogous to the combined model-consistent shock in Section 6.1, I also construct a term that combines direct and indirect exposure to land constraints - and the effects of this term on betas is shown in column 4. [72] Again, the combined term is very precisely estimated and yields a land constraint effect on betas that lies between the naive direct effect estimate and the version that controls flexibly for migration spillovers. This application shows that incorporating migration exposures can help explain a substantial share of the variation in housing market cyclicality across cities.

6.3 Covid effects across housing markets

One of the largest dislocations in housing markets in recent years was the shock that resulted from the effects of the Covid pandemic. As a result of pandemic-related shifts in housing demand, we saw, for example, a change in within-city rent and price gradients with distance to the city center (Gupta, Mittal, Peeters, and Van Nieuwerburgh, 2021), and inter-city migration induced by the pandemic (Haslag and Weagley, 2021). One important determinant of pandemic-related changes in housing demand was the initial density of a neighborhood, which was associated with large declines in housing demand (see, e.g. Liu and Su (2021); Ramani and Bloom (2021)). As populations shift across cities in response to these Covid-related shocks, can the migration exposure methodology proposed in this paper give us additional insights into *where* we would expect the housing demand to move as it shifts away from dense cities?

[69] See Appendix D.1 for details on the beta estimation

[70] See, e.g. Saiz (2010); Paciorek (2013); Diamond (2016).

[71] Defined as $\sum_{j:j\neq i} \psi^{ij} x_j^{land}$, where all cities j are at least 150 mi. away from city i.

[72] This is constructed as $Land_{it}^{comb} = x_i^{land} + \widehat{\eta^{nw}} \sum_{j:j\neq i} \psi^{ij} x_j^{land}$,

Conceptually, the density effect on house prices differs from the house price shocks considered above in that it represents something closer to a negative shock to amenities (e.g. a change in the availability of local food and drink establishments due to lock-downs, or a perceived health risk of living in high density areas), which results in *lower* house prices and migration away from the affected areas. We would therefore expect the direct effect of the shock to have the opposite sign of the indirect effect coming from migration spillovers caused by it.

In column 2 of panel C of Table 7, I first estimate the reduced-form relationship between commuting zone population density[73] and house price growth in the CZ over the first 20 months of the pandemic, controlling for house price growth in the 12 months before the pandemic, as well as log city population and the log of city house price levels. The estimated effect of density on house price growth across cities is robustly negative, with an increase in density by one standard deviation (∼1.7K people per sq. mi.) resulting in a decrease of Covid-era house price growth by ~ 2.6 ppt.[74]

Where do people move as they leave dense areas? Based on the model presented in this paper, the migration spillover mechanism should operate similarly for amenity shocks as it does for house price shocks: cities that have greater migration exposure to the affected locations should be the destination for a larger share of the Covid-era housing demand changes. That is, while a city's own density is negative for local house prices, being highly *connected* to dense cities through long-distance migration links should be positive for local house prices. Column 3 of panel C of Table 7 confirms this intuition: A city's long-distance migration exposure to density in other cities has a significant positive association with house price growth during the pandemic (and Appendix Table A5 shows that this aligns with population changes in the expected direction). Moreover, comparing the explanatory power across columns 2 and 3 relative to the control variables benchmark in column 1, density alone incrementally explains 3.8 ppt of the variation in Covid house price growth, and a city's migration exposure adds an additional 1.3 ppt (or 34% of the direct effect).

7 Conclusion

In this paper, I have documented that house price dynamics in one city can have a causal effect on house prices in another city because of the migration connection that these cities share. These migration spillovers play an important role in propagating economic shocks across cities and not taking them into account when estimating the impact of geographically dispersed shock, such as changes in credit supply, on local housing markets can result in misleading conclusions.

However, the potential use of this methodology in a U.S. context is not limited to spillover effects on housing markets: the same model could be used to explore the network effects through migration channels of many varieties of shocks, such as trade shocks or natural disasters, and other local outcomes. I aim to expand the range of economic shocks and outcomes studied in future work. For instance, it remains to be explored in more detail what we can learn about local mortgage markets and the construction sector if we consider the role of migration flows that result from shocks in distant locations.

[73] Computed as a population-weighted average across CZ zip codes of of zip code population in 1000s divided by zip code area in square miles.

[74] Note that this contrasts with the across-city results of Ramani and Bloom (2021) who find a weak *positive* relationship between density and house price growth across cities. I replicate their finding if I omit the controls for city size and city house price *levels*, but including either of these controls is sufficient for recovering the significant negative density effect shown here.

Moreover, the analysis in this paper has focused on the importance of *domestic* inter-city migration flows in the U.S. However, immigration to the U.S. from other countries has often been a more salient political issue than domestic flows. An important direction for future research is to extend the analysis in this paper to incorporate the role of international migration, which can play an important role by either complementing or counteracting the effect of domestic moves on housing markets.

References

Adelino, M., A. Schoar, and F. Severino (2015): "Credit supply and house prices: evidence from mortgage market segmentation," Tech. rep., National Bureau of Economic Research.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Allen, T. and D. Donaldson (2018): "The geography of path dependence," *Unpublished manuscript.*

Altonji, J. G. and D. Card (1989): "The effects of immigration on the labor market outcomes of natives," Tech. rep., National Bureau of Economic Research.

Angrist, J. D. and J.-S. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Autor, D. H. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review*, 103, 1553–97.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018a): "The economic effects of social networks: Evidence from the housing market," *Journal of Political Economy*, 126, 2224–2276.

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

Bartel, A. P. (1989): "Where do the new US immigrants live?" *Journal of Labor Economics*, 7, 371–391.

Bartelme, D. (2018): "Trade Costs and Economic Geography: Evidence from the US," .

Bartik, T. J. (1991): "Boon or Boondoggle? The debate over state and local economic development policies," .

Baum-Snow, N., D. A. Hartley, and K. O. Lee (2019): "The long-run effects of neighborhood change on incumbent families," .

Beaudry, P., D. A. Green, and B. M. Sand (2014): "Spatial equilibrium with unemployment and wage bargaining: Theory and estimation," *Journal of Urban Economics*, 79, 2–19.

Borusyak, K., P. Hull, and X. Jaravel (2020): "Quasi-experimental shift-share research designs," Tech. rep., National Bureau of Economic Research.

Boustan, L. P. (2010): "Was postwar suburbanization white flight? Evidence from the black migration," *The Quarterly Journal of Economics*, 125, 417–443.

Caliendo, L., M. Dvorkin, and F. Parro (2019): "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 87, 741–835.

Card, D. (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19, 22–64.

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996): "Migration with endogenous moving costs," *The American Economic Review*, 909–930.

Chinco, A. and C. Mayer (2016): "Misinformed speculators and mispricing in the housing market," *The Review of Financial Studies*, 29, 486–522.

Chodorow-Reich, G., A. M. Guren, and T. J. McQuade (2021): "The 2000s housing cycle with 2020 hindsight: A neo-kindlebergerian view," Tech. rep., National Bureau of Economic Research.

Chodorow-Reich, G. and J. Wieland (2020): "Secular Labor Reallocation and Business Cycles," *Journal of Political Economy.*

Coate, P. and K. Mangum (2019): "Fast Locations and Slowing Labor Mobility," .

Cotter, J., S. Gabriel, and R. Roll (2015): "Can housing risk be diversified? A cautionary tale from the housing boom and bust," *The Review of Financial Studies*, 28, 913–936.

Davidoff, T. et al. (2016): "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated With Many Demand Factors," *Critical Finance Review*, 5, 177–206.

DeFusco, A., W. Ding, F. Ferreira, and J. Gyourko (2018): "The role of price spillovers in the American housing boom," *Journal of Urban Economics*, 108, 72–84.

DeFusco, A. A., W. Ding, F. V. Ferreira, and J. Gyourko (2015): "The role of contagion in the american housing boom," Tech. rep., Working Paper.

Derenoncourt, E. (2019): "Can you move to opportunity? Evidence from the Great Migration," .

Diamond, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *The American Economic Review*, 106, 479–524.

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

Eckert, F., S. Ganapati, and C. Walsh (2019): "Skilled tradable services: The transformation of US high-skill labor markets," *Available at SSRN 3439118*.

Favara, G. and J. Imbs (2015): "Credit supply and the price of housing," *The American Economic Review*, 105, 958–992.

Fishback, P. V., W. C. Horrace, and S. Kantor (2006): "The impact of New Deal expenditures on mobility during the Great Depression," *Explorations in Economic History*, 43, 179–222.

Glaeser, E. L., J. D. Gottlieb, and J. Gyourko (2012): "Can cheap credit explain the housing boom?" in *Housing and the financial crisis*, University of Chicago Press, 301–359.

Glaeser, E. L., J. Gyourko, and A. Saiz (2008): "Housing supply and housing bubbles," *Journal of urban Economics*, 64, 198–217.

Glaeser, E. L., J. Kolko, and A. Saiz (2001): "Consumer city," *Journal of economic geography*, 1, 27–50.

Glaeser, E. L. and J. M. Shapiro (2001): *City growth and the 2000 census: Which places grew, and why*, Brookings Institution, Center on Urban and Metropolitan Policy Washington, DC.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2018): "Bartik instruments: What, when, why, and how," Tech. rep., National Bureau of Economic Research.

Gorback, C. S. and B. J. Keys (2020): "Global Capital and Local Assets: House Prices, Quantities, and Elasticities," Tech. rep., National Bureau of Economic Research.

Greenwald, D. L. and A. Guren (2021): "Do credit conditions move house prices?" Tech. rep., National Bureau of Economic Research.

Gross, E. (2003): "US population migration data: Strengths and limitations," *Internal Revenue Service Statistics of Income Division, Washington, DC. http://www. irs. gov/pub/irs-soi/99gross_update. doc.*

Guerrieri, V., D. Hartley, and E. Hurst (2013): "Endogenous gentrification and housing price dynamics," *Journal of Public Economics*, 100.

Gupta, A., V. Mittal, J. Peeters, and S. Van Nieuwerburgh (2021): "Flattening the curve: pandemic-induced revaluation of urban real estate," *Journal of Financial Economics*.

Gyourko, J., C. Mayer, and T. Sinai (2013): "Superstar cities," *American Economic Journal: Economic Policy*, 5, 167–99.

Haslag, P. H. and D. Weagley (2021): "From LA to Boise: How migration has changed during the COVID-19 pandemic," *Available at SSRN 3808326*.

Haughwout, A., R. W. Peach, J. Sporn, and J. Tracy (2012): "The supply side of the housing boom and bust of the 2000s," in *Housing and the financial crisis*, University of Chicago Press, 69–104.

Himmelberg, C., C. Mayer, and T. Sinai (2005): "Assessing high house prices: Bubbles, fundamentals and misperceptions," *Journal of Economic Perspectives*, 19, 67–92.

Howard, G. (2020): "The migration accelerator: Labor mobility, housing, and demand," *American Economic Journal: Macroeconomics*, 12, 147–79.

Jaeger, D. A., J. Ruist, and J. Stuhler (2018): "Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration," Tech. rep., Working Paper.

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American economic review*, 95, 161–182.

Kaplan, G., K. Mitman, and G. L. Violante (2020): "The Housing Boom and Bust: Model Meets Evidence," *Journal of Political Economy*, 128.

Kotova, N. and A. L. Zhang (2021): "Liquidity in residential real estate markets," Tech. rep., Working paper.

Landvoigt, T. (2017): "Housing demand during the boom: The role of expectations and credit constraints," *The Review of Financial Studies*, 30, 1865–1902.

Liu, S. and Y. Su (2021): "The impact of the COVID-19 pandemic on the demand for density: Evidence from the US housing market," *Economics letters*, 207, 110010.

López-Salido, D., J. C. Stein, and E. Zakrajšek (2017): "Credit-market sentiment and the business cycle," *The Quarterly Journal of Economics*, 132, 1373–1426.

Loutskina, E. and P. E. Strahan (2015): "Financial integration, housing, and economic volatility," *Journal of Financial Economics*, 115, 25–41.

Lutz, C. and B. Sand (2019): "Highly disaggregated land unavailability," *Available at SSRN 3478900*.

Mian, A., K. Rao, and A. Sufi (2013): "Household balance sheets, consumption, and the economic slump," *The Quarterly Journal of Economics*, 128, 1687–1726.

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

Mian, A., A. Sufi, and E. Verner (2019): "How does credit supply expansion affect the real economy? the productive capacity and household demand channels," *The Journal of Finance*.

Molloy, R., C. L. Smith, and A. Wozniak (2011): "Internal migration in the United States," *Journal of Economic perspectives*, 25, 173–96.

Moretti, E. (2013): "Real wage inequality," *American Economic Journal: Applied Economics*, 5, 65–103.

Oh, H. and C. Yoon (2020): "Time to build and the real-options channel of residential investment," *Journal of Financial Economics*, 135, 255–269.

Paciorek, A. (2013): "Supply constraints and housing market dynamics," *Journal of Urban Economics*, 77, 11–26.

Paternoster, R., R. Brame, P. Mazerolle, and A. Piquero (1998): "Using the correct statistical test for the equality of regression coefficients," *Criminology*, 36, 859–866.

Ramani, A. and N. Bloom (2021): "The Donut effect of COVID-19 on cities," Tech. rep., National Bureau of Economic Research.

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

Saks, R. E. (2008): "Job creation and housing construction: Constraints on metropolitan area employment growth," *Journal of Urban Economics*, 64, 178–195.

Severen, C. (2021): "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," Tech. rep.

Sinai, T. and N. Souleles (2013): "Can owning a home hedge the risk of moving?" *American Economic Journal: Economic Policy*, 5, 282–312.

Stock, J. H. and M. W. Watson (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *The Economic Journal*, 128, 917–948.

Sturm, D., S. Heblich, and S. Redding (2020): "The making of the modern metropolis: evidence from London," *Quarterly Journal of Economics*.

Sumner, S. and K. Erdmann (2020): "Housing Policy, Monetary Policy, and the Great Recession," *Mercatus Research Series*.

Table 1: U.S CZ-to-CZ domestic migration links 1990-2010

	# of links		*% of links that overlap after...*			
Migration distance	**Avg.**	**Median**	**1 year**	**2 years**	**5 years**	**10 years**
Full Network	30	11	83	82	82	80
Distance $>$ 50 **mil**	26	6	75	74	72	70
Distance $>$ 150 **mil**	21	2	69	67	65	62

Note: Table uses continental U.S. CZ migration data constructed from IRS flows for 1990-2010 that had at least one inflow link to another CZ in 1990-2010, which includes 717 CZs. The IRS does not record flows corresponding to $<$ 10 tax returns, and data for years past 2010 are omitted due to a methodology change in IRS gross flows post-2010. Link persistence is computed as the % of links in a given year that are still links after a certain # of years for each CZ-year, averaged over CZs and years. Networks based on migration distances (rows 2 and 3) exclude any CZ that contains counties with a centroid that is less than the stated distance away from the centroid of any county in the CZ for which inflow links are counted.

Figures and Tables

8 Tables

Table 2: Migration cost determinants

Dependent var.:	Log migration between cities		
Period:	1990-2017	2005-2017	
Education group:	All	College	No college
	(1)	(2)	(3)
Log distance	-1.228***	-0.682***	-0.948***
	(0.059)	(0.055)	(0.033)
Different region	-1.126***	-0.342	-0.802**
	(0.350)	(0.212)	(0.323)
Diff. region× Log dist.	0.137**	0.021	0.093
	(0.059)	(0.037)	(0.058)
Different state	-2.632***	-1.849***	-2.744***
	(0.463)	(0.406)	(0.440)
Diff. state× Log dist.	0.372***	0.187**	0.340***
	(0.096)	(0.084)	(0.086)
Nontrad. Christ. Share diff.	-0.658***	-0.389***	-0.282***
	(0.136)	(0.075)	(0.064)
Ethnicity shares vector dist.	-1.015***	-0.510***	-0.604***
	(0.170)	(0.090)	(0.098)
Water surface	-1.253***	-0.642***	-0.656***
	(0.176)	(0.131)	(0.137)
Jan. Temp. diff.	0.022***	0.006***	0.010***
	(0.004)	(0.002)	(0.003)
Industry shares vector dist.	-3.523***	-1.291***	-1.070***
	(0.302)	(0.187)	(0.181)
Observations	13,390,146	6,686,743	6,686,743
Year FE	✓	✓	✓
Origin City FE	✓	✓	✓
Destination City FE	✓	✓	✓

Heteroskedasticity-robust standard errors clustered at the origin CZ level in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Analysis includes migration flows between all continental U.S. CZs, excl. New Orleans, leading to a total of 721 CZs. Migration flows by education group 2005-2017 were imputed from ACS and IRS data as described in the text. Industry share distance is computed as Euclidean distance between vectors of 2-digit NAICS employment shares in 2000. Ethnic distance is the Euclidean distance in CZ ethnicity shares in 2000. Estimated using Poisson Pseudo-Maximum Likelihood with fixed effects for origin and destination cities.

Table 3: Large cities with highest and lowest migration exposure

Rank	*Commuting Zone*	*Migration exposure (in %)*	*House price beta '90-'17*	*Population (in '000s)*
1	Colorado Springs, CO	14.68	0.63	526
2	Pensacola, FL	10.87	1.45	533
3	Las Vegas, NV	9.73	2.68	1,338
4	Virginia Beach, VA	9.30	1.21	959
5	Fayetteville, NC	9.13	0.25	548
6	El Paso, TX	8.90	0.56	766
7	Tucson, AZ	8.70	1.56	808
8	Cape Coral, FL	8.35	2.53	578
9	Jacksonville, FL	8.32	1.63	1,047
10	Albuquerque, NM	8.20	0.74	568
⋮	⋮	⋮	⋮	⋮
94	Erie, PA	1.18	0.21	554
95	Rockford, IL	0.94	0.59	563
96	Youngstown, OH	0.87	0.37	690
97	Allentown, PA	0.82	0.95	571
98	Harrisburg, PA	0.82	0.62	975
99	Canton, OH	0.75	0.42	632
100	Racine, WI	0.73	0.78	518
101	Reading, PA	0.57	0.60	1,007
102	Scranton, PA	0.40	0.57	722
103	Appleton, WI	0.40	0.43	502

Table shows the ten large commuting zones (IRS population >500K) with the highest and lowest migration exposure. See Section 3.2 for the definition and measurement of migration exposure. The third column shows the house price beta for each CZ (see Appendix D.1 for details on the beta estimation), and the fourth column its population as of the year 2000, as measured in IRS data.

Table 4: Reduced-form IV spillover estimation - baseline results

Dependent variable:	Log House price growth$_{it}$				Δ Network House Prices$_{it}$
Specification:	IV (1)	IV All controls (2)	IV Supply Heterogeneity (3)	OLS All controls (4)	First Stage (5)
Δ NW House Prices$_{it}$	4.50***	11.37***	7.50*	6.03***	
	(0.68)	(3.60)	(4.41)	(0.70)	
Δ NW House Prices$_{it}$ $\times$ Supply$_i$			8.87***		
			(3.25)		
NW House Price Instrument$_{it}$					2.21***
					(0.19)
Observations	16,767	16,767	16,767	16,767	16,767
1st-stage F-stat.	3,999	134	51		
Year FE	✓	✓	✓	✓	✓
CZ FE	✓	✓	✓	✓	✓
Regional trend FE		✓	✓	✓	✓
Migration Access$_{i,t-1}$		✓	✓	✓	✓
NW Wage Shock$_{it}$		✓	✓	✓	✓
Wage Shock$_{it}$		✓	✓	✓	✓
Wage Shock$_{it}$ $\times$ Unavail. Land$_i$		✓		✓	✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Includes data from 621 CZs for 1991-2017. Network measures include only CZs with > 150 mi. distance from the focal city in the migration network. See section 3 for details on the variable construction. Supply constraints are measured by the share of land unavailable for construction in the CZ.

Table 5: Additional outcomes: IV estimates

Dependent variable:	Net Migration	Log Hous. Permits	Log Mtg. Loans	Log Mtg. Lending ($)
	(1)	(2)	(3)	(4)
Δ NW House Prices$_{it}$	3.57***	2.35***	1.90***	2.32***
	(1.01)	(0.74)	(0.46)	(0.55)
Observations	16,767	16,362	16,738	16,738
Year FE	✓	✓	✓	✓
CZ FE	✓	✓	✓	✓
Regional trend FEs	✓	✓	✓	✓
Migration Access$_{i,t-1}$	✓	✓	✓	✓
NW Wage Shock$_{it}$	✓	✓	✓	✓
Wage shock$_{it}$	✓	✓	✓	✓
Wage shock$_{it}$ $\times$ Uavail. Land$_i$	✓	✓	✓	✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Includes data from 621 CZs for 1991-2017. Migration network by distance includes all CZs for which the center of no component county is no closer than 150 miles to the center of any focal city component county. Network house prices are the weighted sum of other CZs' house price changes. See section 3 for details on the variable construction.

Table 6: IV Spillover effects - robustness checks

Dependent variable:	Log House price growth$_{it}$							
Specification:	Alternative IV		Incl. short dist.		Log pop. weights		Excl. largest CZs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ NW House Prices$_{it}$	6.81***	4.82***	4.17***	4.56***	11.29*	0.67	10.92***	7.41*
	(1.24)	(1.60)	(1.18)	(1.10)	(6.66)	(7.63)	(3.59)	(4.51)
Δ NW HP$_{it}$ $\times$ Supply$_i$		5.14*		2.80***		20.04***		8.46***
		(2.95)		(1.00)		(5.40)		(3.26)
Observations	9,929	9,929	19,170	19,170	16,146	16,146	16,227	16,227
1st-stage F-stat.	1,587	516	67	70	131	56	130	48
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
CZ FE	✓	✓	✓	✓	✓	✓	✓	✓
Regional trend FE	✓	✓	✓	✓	✓	✓	✓	✓
Migration Access$_{i,t-1}$	✓	✓	✓	✓	✓	✓	✓	✓
NW Wage Shock$_{it}$	✓	✓	✓	✓	✓	✓	✓	✓
Wage Shock$_{it}$	✓	✓	✓	✓	✓	✓	✓	✓
Wage Shock$_{it}$ $\times$ Supply$_i$	✓	✓	✓		✓		✓	
Credit Shock$_{it}$	✓	✓						

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Includes data from 621 CZs for 1991-2017 in columns 3-8, and 1991-2006 in columns 1-2. Network measures include only CZs with $>$ 150 mi. distance from the focal city in the migration network. See section 3 for details on the variable construction and section 4 for a description of the specifications.

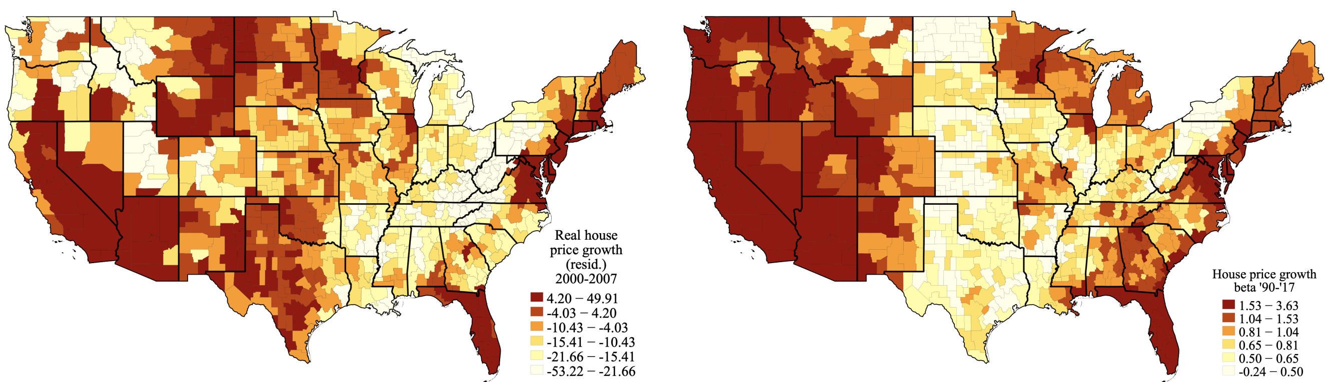
Table 7: Migration spillover applications

	(1)	(2)	(3)	(4)
Panel A: Credit Shocks				
Dependent variable:	Log House price growth$_{it}$			
Credit shock$_{it}$		10.23***	7.49***	
		(0.93)	(0.96)	
Mig. exposure to credit shock			120.36***	
			(18.03)	
Comb. direct & mig. exposure effect of credit shocks				8.65***
				(0.65)
Observations	9,929	9,929	9,929	9,929
Adj. R-squared	0.212	0.237	0.248	0.248
Controls	✓	✓	✓	✓
Panel B: House Price Betas				
Dependent variable:	1990-2017 House Price Beta$_i$			
Land Constraint$_{it}$		0.92***	0.74***	
		(0.09)	(0.08)	
Mig. exposure to land constraints			15.27***	
			(2.29)	
Comb. direct & mig. exposure effect of land constraints				0.88***
				(0.07)
Observations	621	621	621	621
Adj. R-squared	0.156	0.325	0.400	0.391
Controls	✓	✓	✓	✓
Panel C: Covid and Density				
Dependent variable:	Mar. ‘20-Nov. ‘21 House price growth$_i$			
Pop. density		-1.50***	-1.43***	
		(0.41)	(0.43)	
Mig. exposure to density			14.42***	
			(3.52)	
Observations	602	602	602	
Adj. R-squared	0.260	0.298	0.311	
Controls	✓	✓	✓	

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Migration network by distance includes all CZs for which the center of no component county is closer than the stated number of miles to the center of any focal city component county. See section 6 for details on the variables and specifications.

9 Figures

Figure 1: House price growth and betas. Panel (a) shows a map of commuting zone average (cumulative) real house price growth over 2000-2007, residualized with regard to the effect of annual time trends, local land constraints, and the interaction between them in a population-weighted regression. Panel (b) shows house price growth "betas" for each CZ for 1990-2017 (see Appendix D.1 for details on their estimation).

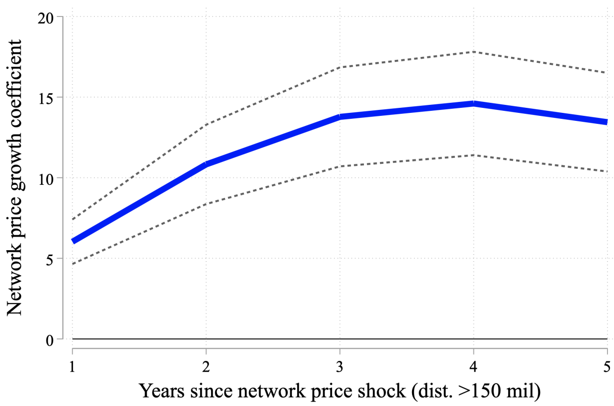


(a) Resid. real house price growth 2000-2007 **(b)** House price growth beta 1990-2017

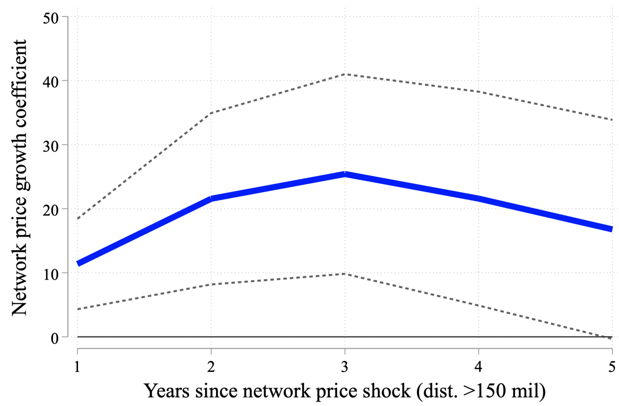
Figure 2: Dynamic spillover effects on house prices: Graphs show OLS and IV local projection coefficients corresponding to the effect $\tilde{\eta}_h^{nw}$ of period t network house price growth on period $t-1+h$ outcome variables of the form

$$\sum_{k=1}^{h} \Delta \ln P_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \eta_h^{land} p_{it}^{nw} x_i^{land} + \beta_{cum}' \Gamma_{it} + \tilde{\xi}_{i,t-1+h}^{P,cum}.$$

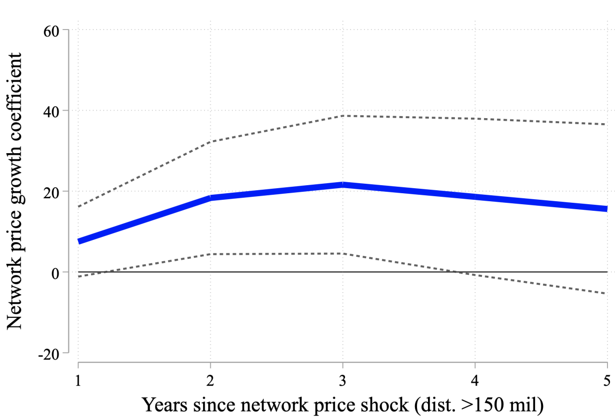
Baseline control variables Γ_{it} in all regressions include year & CZ FEs, regional trend FEs, migration access, network wage shocks and local Bartik wage shocks. Instruments for network house price growth in IV regressions consist of the migration-weighted network average of migration-weighted productivity shocks interacted with other cities' land share unavailable for construction. Regressions in panels a and b do not contain the interaction term multiplying the η_h^{land} parameter and also control for an interaction between local land share unavailable for construction and the wage shocks. Regressions in panels c and d show the estimates of η_h^{nw} and η_h^{land}, respectively, from a regression including the land supply constraint interaction term, and use additional instruments consisting of the baseline network instrument interacted with the unavailable land share in city i. Estimation uses data for 1991-2017 and only includes cities at $>$ 150 mi. distance in migration networks. Dashed lines show 95% CI based on std. errors clustered at the CZ level.



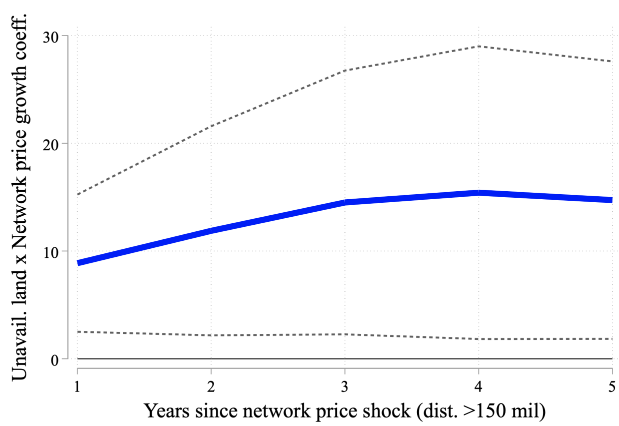
(a) OLS: baseline spillover effect η_h^{nw}



(b) IV: baseline spillover effect η_h^{nw}



(c) IV w. heterogeneity: η_h^{nw}

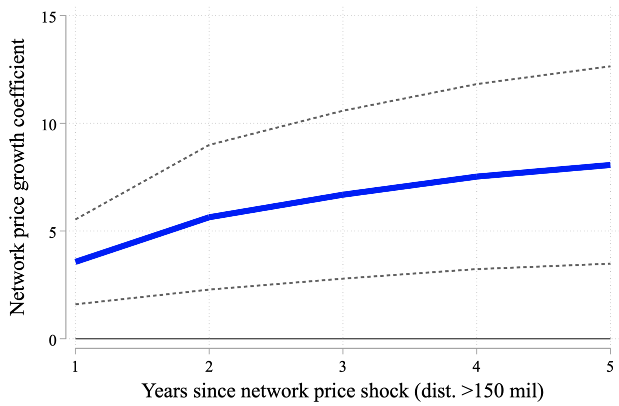


(d) IV w. heterogeneity: η_h^{land}

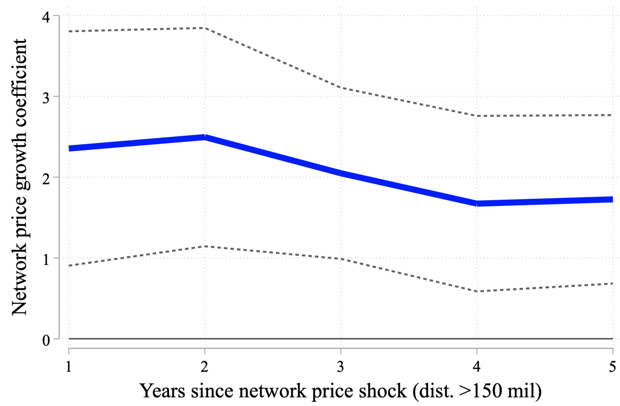
Figure 3: Dynamic IV spillover effects on other outcomes: Graphs show local IV projection coefficients corresponding to the effect of period t network house price growth with period $t-1+h$ outcome variables of the form

$$Y_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \beta' \Gamma_{it} + \tilde{\xi}_{i,t-1+h}^{P}.$$

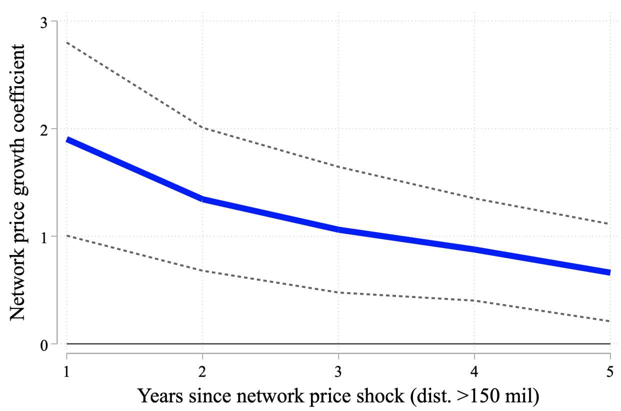
Control variables Γ_{it} in all regressions include year & CZ FEs, regional trend FEs, migration access, network wage shocks, local Bartik wage shocks, and an interaction between local land share unavailable for construction and the wage shocks. Instruments for network house price growth in IV regressions consist of the migration-weighted network average of migration-weighted productivity shocks interacted with other cities' land share unavailable for construction. Estimation uses data for 1991-2017 and only includes cities at $>$ 150 mi. distance in migration networks. Dashed lines show 95% CI based on std. errors clustered at the CZ level.



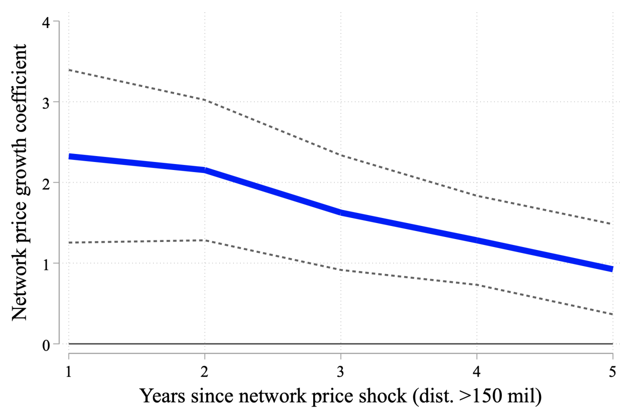
(a) Cumul. Pop Impact of Net Mig. (log pts.)



(b) Log Housing Unit Permits

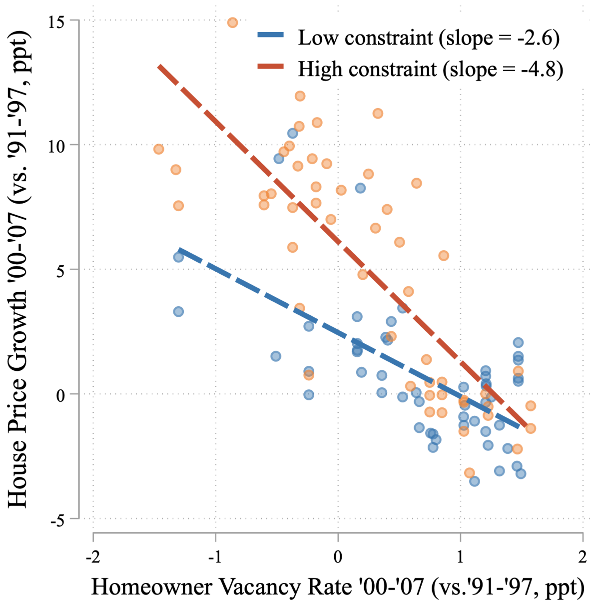


(c) Log Mortgage Purchase Loans

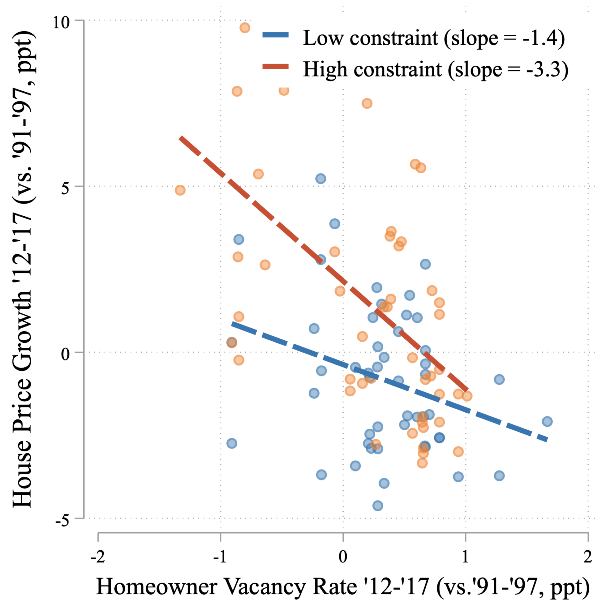


(d) Log Mortgage Lending Volume (in USD)

Figure 4: House price growth and owner-occupied vacancy rates. The graphs in panels (a) and (b) plot CZ average annual house price growth during the housing booms of 2000-2007 and 2012-2017 over the change in average vacancy rates for owner-occupied housing. All averages are computed as differences relative to the 1991-1997 average. The graphs and fitted lines include 104 CZs for 2000-2007 and 98 CZs for 2012-2017. The corresponding rental housing results are shown in Appendix Figure A8.

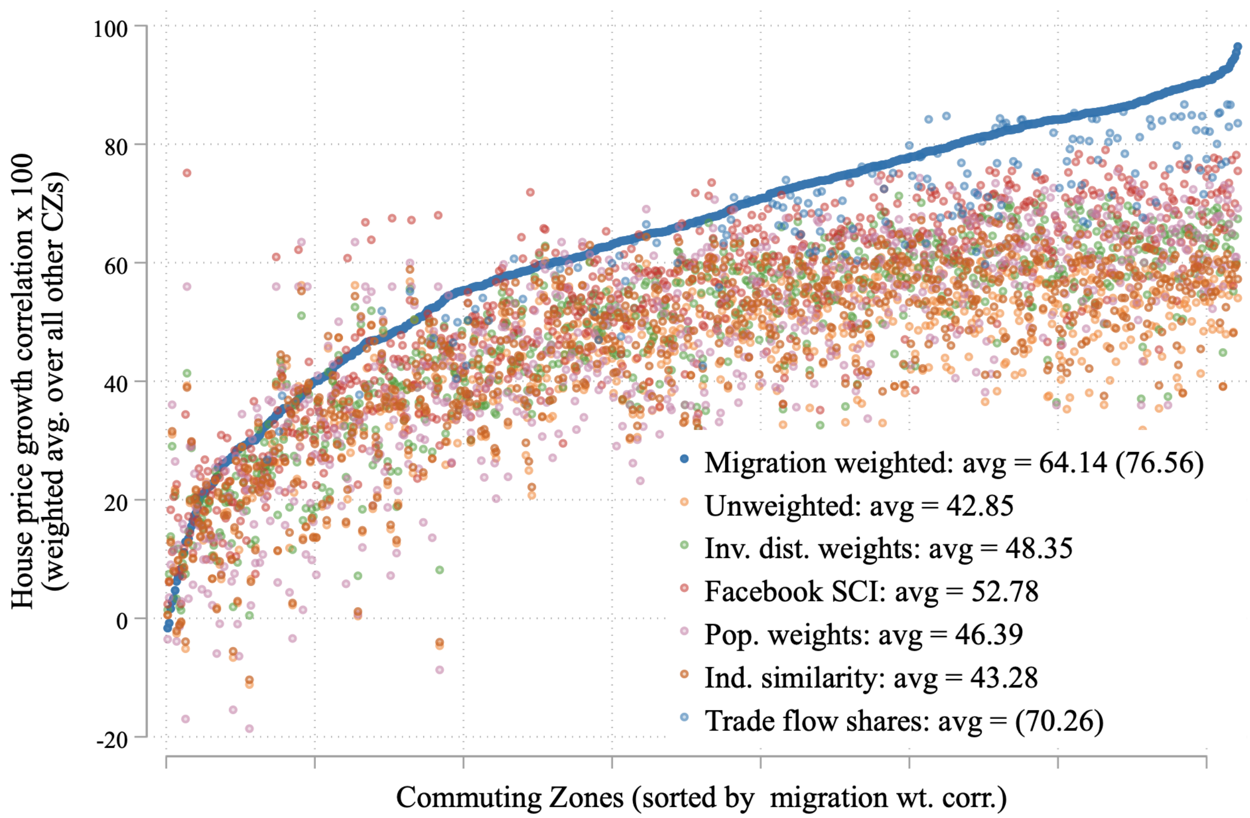


(a) Owner-occupied: 2000-2007



(b) Owner-occupied: 2012-2017

Figure 5: Expected weighted house price correlations between CZs The graph sorts CZs by their migration-weighted house price correlation, and then plots for each CZ the expected house price growth correlation 1990-2017 with all other CZs, weighting each CZ using the weights stated in the legend, which are (1) migration outflow shares (2) equal weights (3) inverse distance in miles (4) Facebook social connectedness weights, and (5) population weights (6) Trade flow weights. A higher value therefore indicates that the link captured by those weights is more strongly associated with a co-movement in house prices. The legend also shows the expected correlation using the weight, with numbers in parentheses being calculated only for the smaller sample of CZs where trade-flow-weighted expectations are available.



ONLINE APPENDIX

A Appendix Figures

Figure A1: Migration inflow network examples. Shading in the map indicates % of continental U.S. origin inflows to the city coming from the other CZs over 2000-2007 period in IRS migration data.

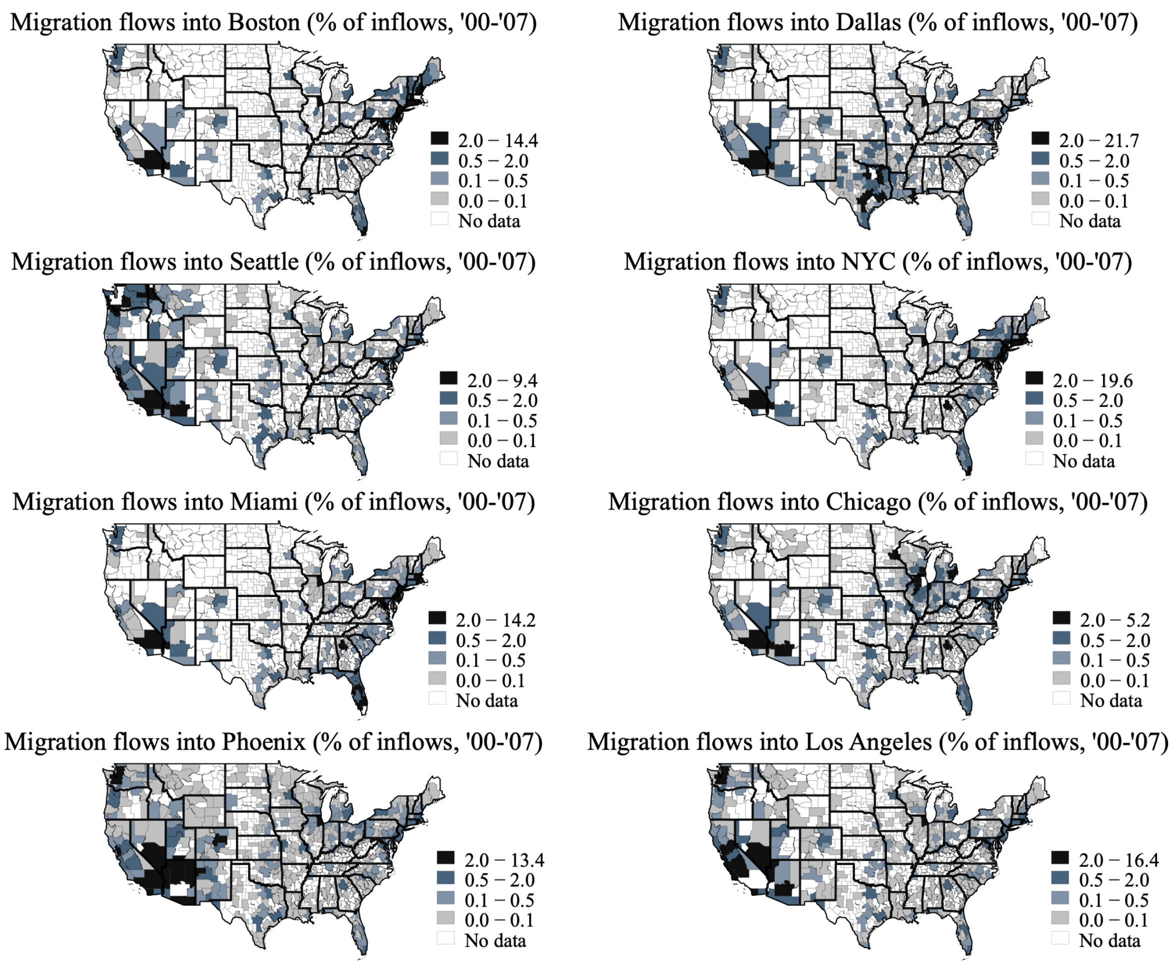


Figure A2: Ilustration of network IV identification.

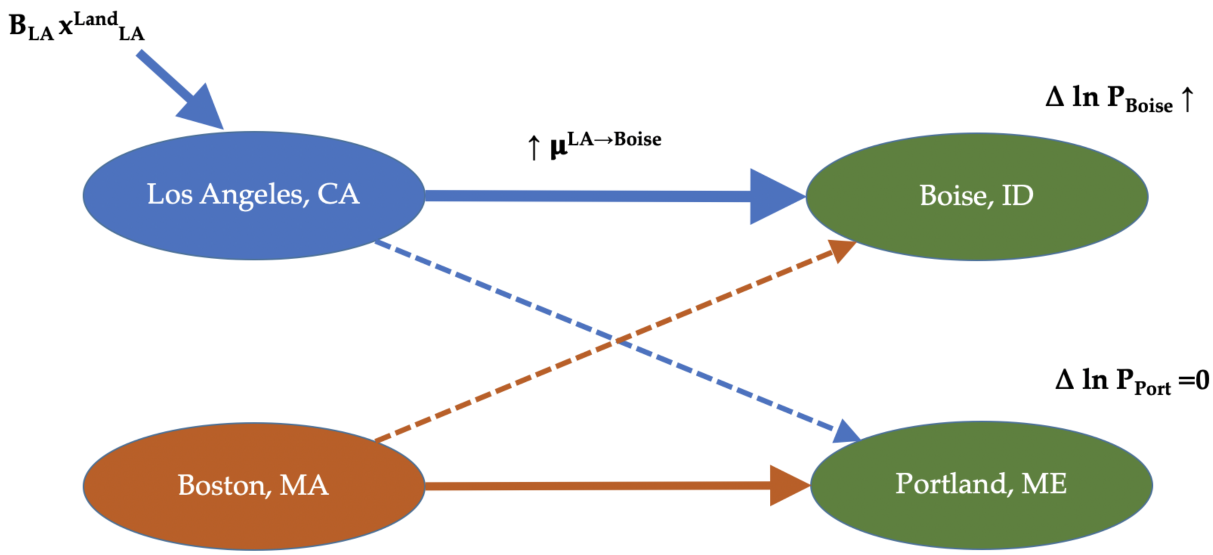


Figure A3: First stage: network instrument and network house price growth. Graph shows binned scatter plot of the first-stage relationship between network house price growth and the network house price growth instrument. See Section 3.2 for the definition of the network instrument. All variables are residualized with regard to city and year FEs, regional trends, migration access, local Bartik shock, local Bartik shock $\times$ land unavailability interaction, and network wage shocks. Data pools years 1991-2017 for 621 CZs and includes the migration network with > 150 mi. distance from each city.

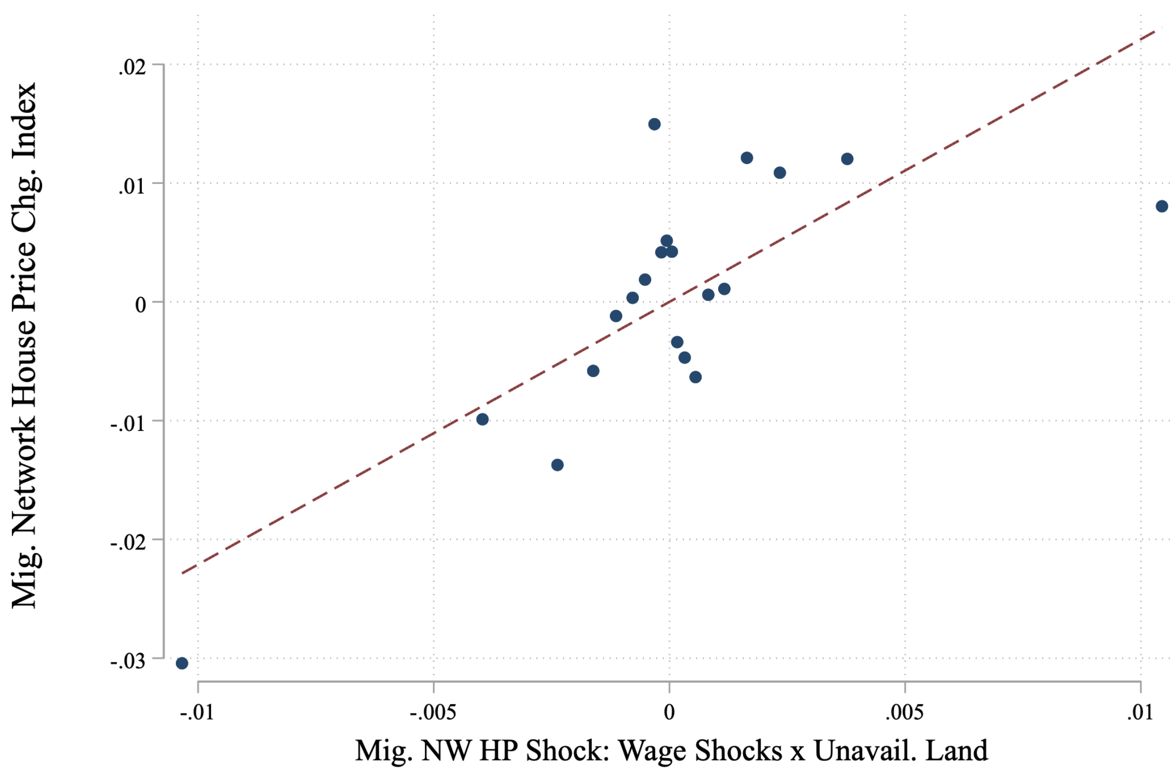
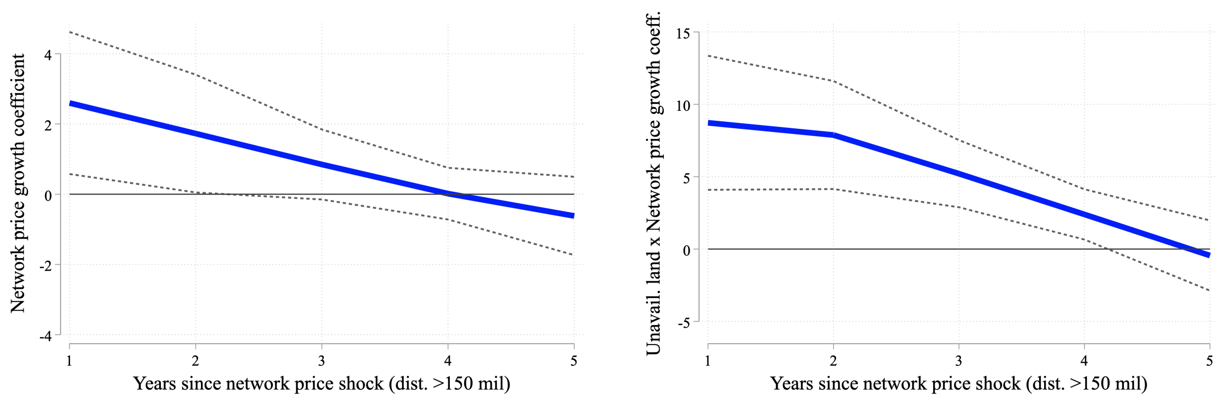


Figure A4: OLS dynamic spillover effects with supply heterogeneity. Graphs show local OLS projection coefficients corresponding to the effects η_h^{nw} and η_h^{land} of period t network house price growth on period $t-1+h$ outcome variables in regressions of the form

$$\sum_{k=1}^{h} \Delta \ln P_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \eta_h^{land} p_{it}^{nw} x_i^{land} + \beta_{cum}' \Gamma_{it} + \tilde{\xi}_{i,t-1+h}^{P,cum}.$$

Baseline control variables Γ_{it} in include year & CZ FEs, regional trend FEs, migration access, network wage shocks and local Bartik wage shocks. The panels a and b show the estimates of η_h^{nw} and η_h^{land}, respectively, from a regression including the land supply constraint interaction term. Estimation uses data for 1991-2017 and only includes cities at > 150 mi. distance in migration networks. Dashed lines show 95% CI based on std. errors clustered at the CZ level.

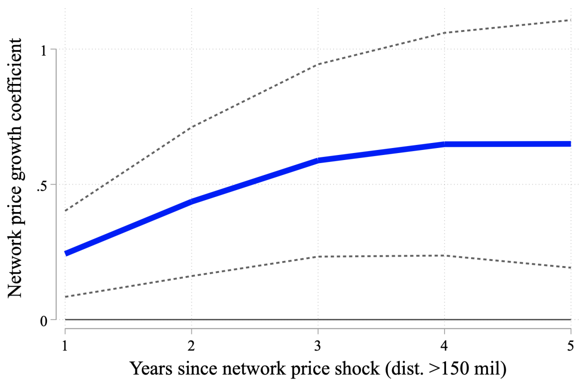


(a) OLS w. heterogeneity: η_h^{nw} **(b)** OLS w. heterogeneity: η_h^{land}

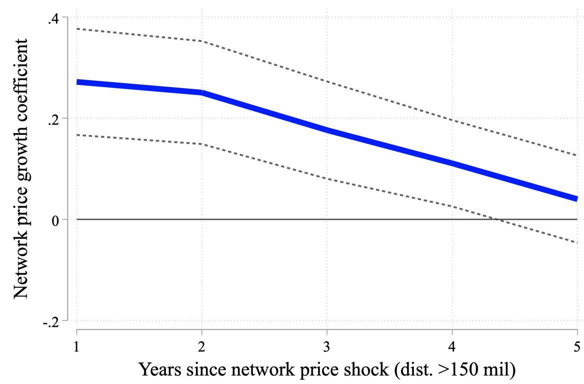
Figure A5: **Dynamic effects: other outcomes - OLS** Graphs show local OLS coefficients corresponding to the effect of period t network house price growth with period $t-1+h$ outcome variables of the form

$$Y_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \beta' \Gamma_{it} + \tilde{\xi}_{i,t-1+h}^P.$$

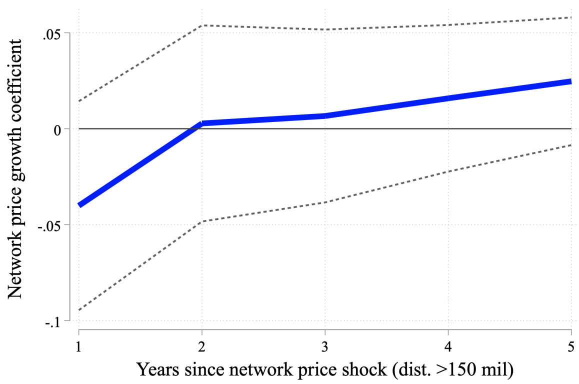
Control variables Γ_{it} in all regressions include year & CZ FEs, regional trend FEs, migration access, network wage shocks, local Bartik wage shocks, and an interaction between local land share unavailable for construction and the wage shocks. Estimation uses data for 1991-2017 and only includes cities at > 150 mi. distance in migration networks. Dashed lines show 95% CI based on std. errors clustered at the CZ level.



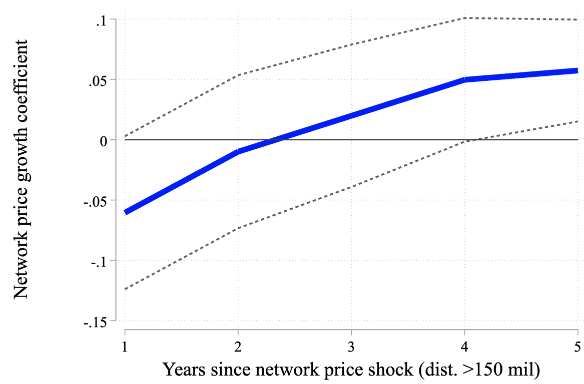
(a) OLS: Cumul. Pop Impact of Net Mig. (log pts)



(b) OLS: Log Housing Unit Permits



(c) OLS: Log Mortgage Purchase Loans

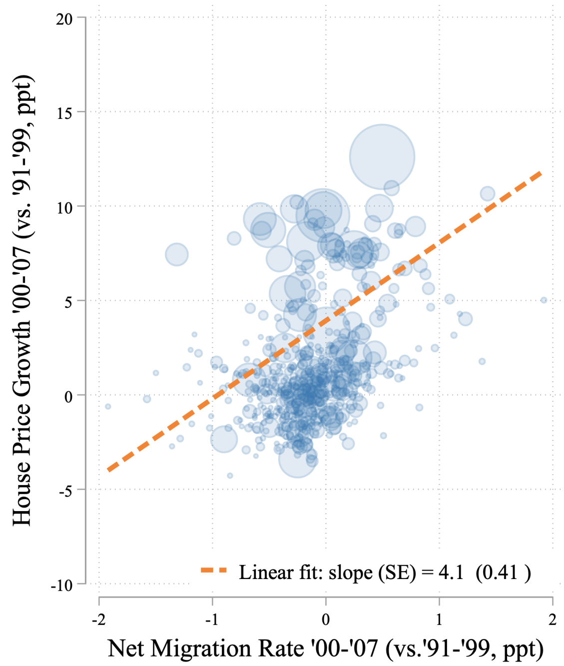


(d) OLS: Log Mortgage Lending Volume (in USD)

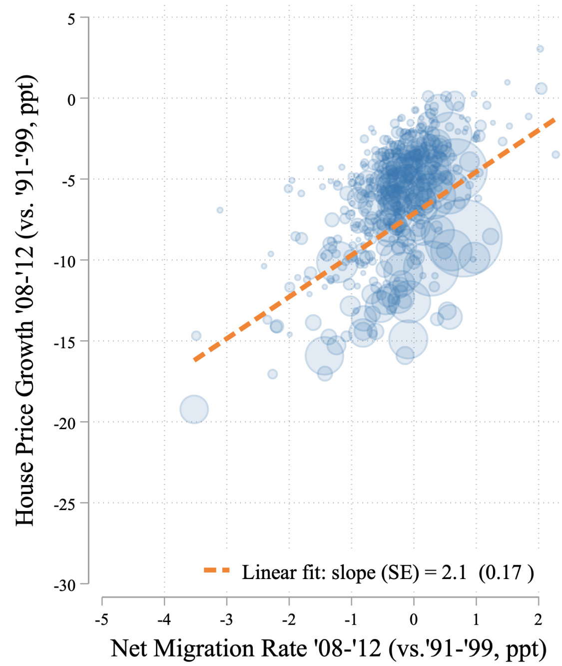
Figure A6: House price growth and net migration. The graph plots CZ average annual house price growth and average net migration rates during different time periods. All period averages are net of their averages during 1991-1999. For example, the fitted line for the 2000-2007 boom period corresponds to the regression

$$\overline{\Delta \ln P}_{\text{Boom},i} - \overline{\Delta \ln P}_{\text{`91-`99},i} = \alpha + \beta(\overline{NetMig\%}_{\text{Boom},i} - \overline{NetMig\%}_{\text{`91-`97},i}) + \epsilon_i$$

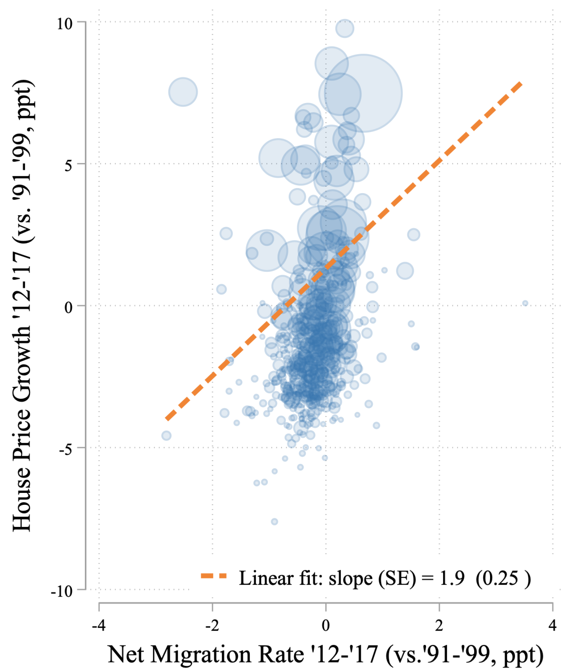
Graph and fitted line include all 559 CZs with population > 30K, but estimated line slopes represent CZs of all population sizes, weighted by their population and excluding New Orleans.



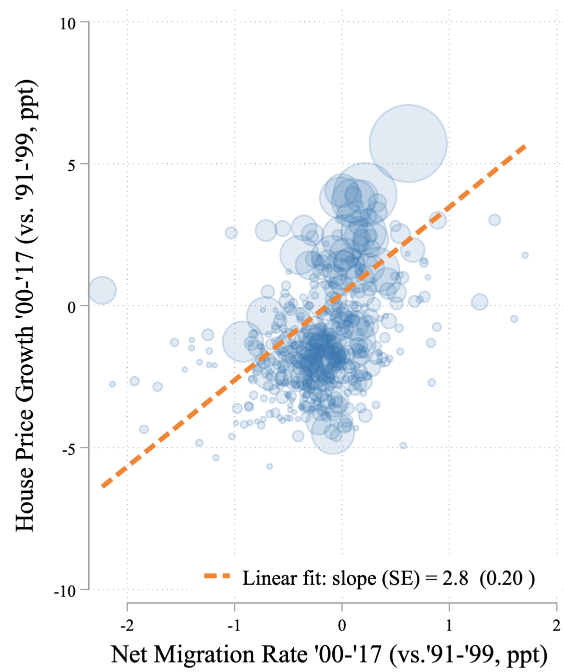
(a) Boom: 2000-2007



(b) Bust: 2008-2012

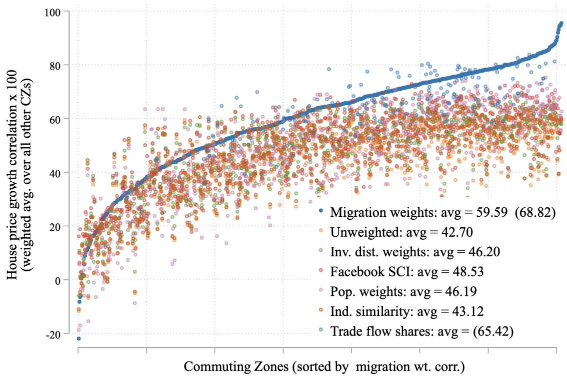


(c) Boom: 2012-2017

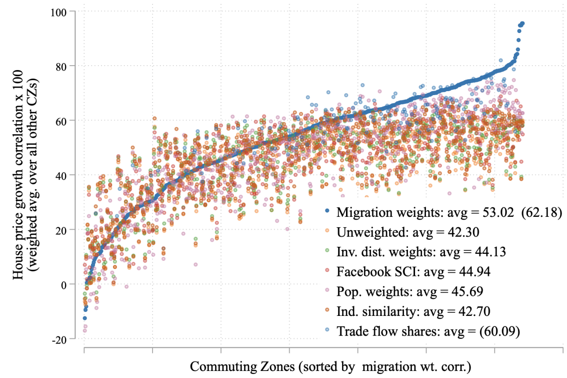


(d) 2000-2017

Figure A7: Expected weighted house price correlations between CZs The graph sorts CZs by their migration-weighted house price correlation, and then plots for each CZ the expected house price growth correlation 1990-2017 with all other CZs, weighting each CZ using the weights stated in the legend, which are (1) migration outflow shares (2) equal weights (3) inverse distance in miles (4) Facebook social connectedness weights, and (5) population weights (6) Trade flow weights. A higher value therefore indicates that the link captured by those weights is more strongly associated with a co-movement in house prices. The legend also shows the expected correlation using the weight, with numbers in parentheses being calculated only for the smaller sample of CZs where trade-flow-weighted expectations are available. Each panel excludes CZs at less than the stated distance from the computation of a CZ's average housing correlation with each weight.

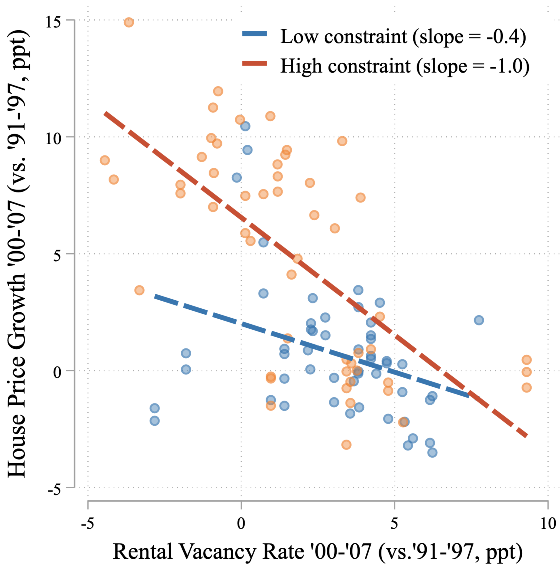


(a) Distance > 50 mi.

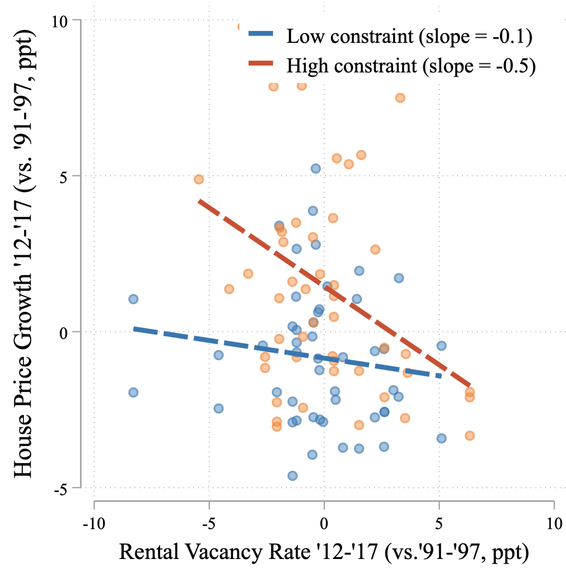


(b) Distance > 150 mi.

Figure A8: House price growth and vacancy rates: rental housing. The graphs plot CZ average annual house price growth during the housing booms of 2000-2007 and 2012-2017 over the change in average vacancy rates for rental housing. All averages are computed as differences relative to the 1991-1997 average. The graphs and fitted lines include 104 CZs for 2000-2007 and 98 CZs for 2012-2017.

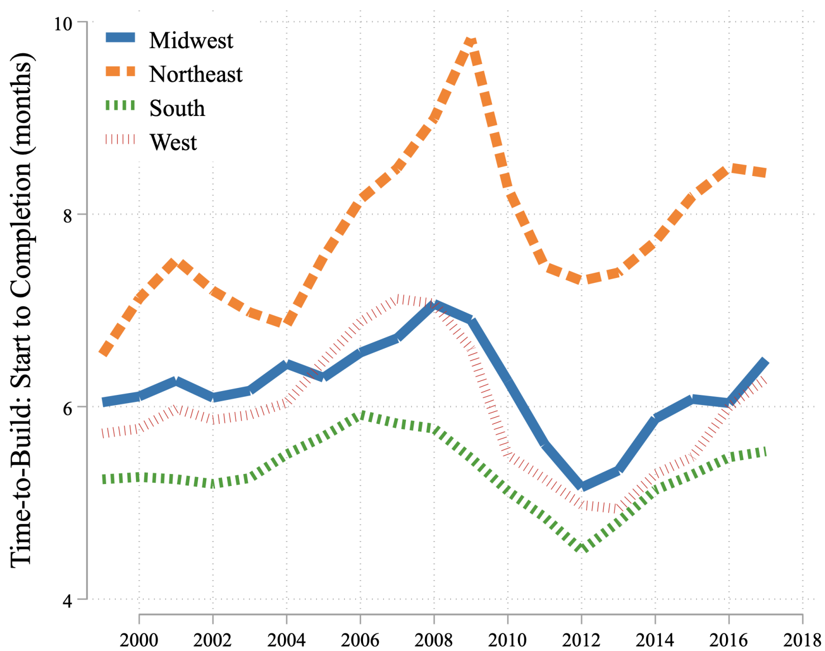


(a) Rental: 2000-2007



(b) Rental: 2012-2017

Figure A9: Time-to-build delays. The graph shows time-to-build delays by completion year estimated from construction start-to-completion time microdata from the Census Survey of Construction. Region averages are estimated from regressions of individual project delays on region-by-year fixed effects. Estimates are broken out by U.S Census region.



(a) Time-to-build delays by region

B Appendix Tables

Table A1: Summary statistics. The table below shows summary statistics for the key variables used in the estimation for the continental U.S. CZs (under the 1990 boundary definition) of which there are 721 (excl. New Orleans), and for the years 1990-2017.

Variable	Mean	SD	Median	Min.	Max.	Obs.
thisstat1						
Population ('000s, IRS)	345.82	948.80	89.95	1	17360	20,188
Log chg. in house prices (FHFA)	2.92	4.50	2.97	-41	44	20,188
Net domestic migration (% of pop., IRS)	0.05	6.09	-0.04	-50	851	20,188
Gross domestic in-migration (% of pop., IRS)	4.70	6.26	4.37	0	858	20,188
Gross domestic out-migration (% of pop., IRS)	4.65	1.64	4.37	0	52	20,188
Bartik wage shift-share shock (1990 wts., 3-dig.)	3.08	1.27	3.07	-11	19	19,467
Construction permits ('000 housing units)	1.86	5.05	0.28	0	90	19,681
Purch. mortgage originations ('000s, annual)	5.35	16.11	0.79	0	429	20,055
Purch. mortgage orig. vol. (Mil. USD, annual)	922.42	3818.25	77.95	0	137882	20,055
Unavailable land share (%, Lutz & Sand, 2019)	26.01	21.67	19.88	0	92	19,170
Network house price chg. exposure	0.32	0.47	0.32	-4	4	19,170
Network productivity shock	0.33	0.18	0.30	-0	1	19,170
Network land share× productivity shock	0.08	0.07	0.06	-0	1	19,170

Table A2: Reduced-form house price spillover effects - by network distance

Dependent variable:	CZ Log House price growth$_{it}$			
Network min. distance	0 mi.	50 mi.	100 mi.	150 mi.
	(1)	(2)	(3)	(4)
Panel A: OLS				
Δ NW House Prices$_{it}$	6.34***	4.98***	5.88***	6.03***
	(0.36)	(0.47)	(0.59)	(0.70)
Panel B: IV				
Δ NW House Prices$_{it}$	4.17***	5.06***	7.99***	11.37***
	(1.18)	(1.25)	(1.98)	(3.60)
Observations	19,170	18,603	17,631	16,767
1st-stage F-stat.	67	53	26	134
Year FE	✓	✓	✓	✓
CZ FE	✓	✓	✓	✓
Regional trend FEs	✓	✓	✓	✓
Migration Access$_{i,t-1}$	✓	✓	✓	✓
NW Wage Shock$_{it}$	✓	✓	✓	✓
Wage shock$_{it}$	✓	✓	✓	✓
Wage shock$_{it}$ $\times$ Unavail. Land$_i$	✓	✓	✓	✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Includes data from 621-709 CZs for 1991-2017. Migration network by distance includes all CZs for which the center of no component county is closer than the stated number of miles to the center of any focal city component county. See section 3 for details on the variable construction.

Table A3: Reduced-form house price spillovers: IV long-run effects

Dependent variable:	Δ Log HP$_{i,t+4}$	Δ Log HP$_{i,t+4}$	Δ Log Pop.$_{i,t \text{ to } t+4}$	Log Permits$_{i,t+4}$	Log Mtg. Loans$_{i,t+4}$	Log Mtg. Lend.$_{i,t+4}$
	(1)	(2)	(3)	(4)	(5)	(6)
Δ NW House Prices$_{it}$	16.77*	15.59	8.07***	1.73***	0.66***	0.92***
	(8.72)	(10.69)	(2.34)	(0.53)	(0.23)	(0.28)
Δ NW House Prices$_{it}$ $\times$ Supply$_i$		14.73**				
		(6.57)				
Observations	14,283	14,283	14,283	13,927	14,283	14,283
Year FE	✓	✓	✓	✓	✓	✓
CZ FE	✓	✓	✓	✓	✓	✓
Regional trend FEs	✓	✓	✓	✓	✓	✓
Migration Access$_{i,t-1}$	✓	✓	✓	✓	✓	✓
NW Wage Shock$_{it}$	✓	✓	✓	✓	✓	✓
Wage shock$_{it}$	✓	✓	✓	✓	✓	✓
Wage shock$_{it}$ $\times$ Unavail. Land$_i$	✓		✓	✓	✓	✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Includes data from 621 CZs for 1991-2017. Migration network by distance includes all CZs for which the center of no component county is closer than 150 miles to the center of any focal city component county. See section 3 for details on the variable construction. All estimates shown for house price and population are cumulative total effects of a period t shock on period $t+4$ log house prices and population (cumulated domestic net migration). Permit, loan, and lending effects are long-run effects on the annual *flow*.

Table A4: Spillover effect size heterogeneity by migration exposure and land constraints

Land constraint (%)		Migration exposure (%)				
		Measure	25th Pctl	Median	Mean	75th Pctl
Measure	*Value*	*Value*	0.3	1.4	2.4	3.5
Panel A: Short-run effects		*Parameters:*				
		Tab. 4, Col. 2	0.3	1.6	2.7	4.0
25th Pctl	10.4	Tab. 4, Col. 3	0.3	1.2	2.0	2.9
50th	22.6	Tab. 4, Col. 3	0.3	1.3	2.3	3.3
Mean	27.5	Tab. 4, Col. 3	0.3	1.4	2.4	3.5
75th Pctl	38.6	Tab. 4, Col. 3	0.3	1.5	2.6	3.8
Panel B: Year 5 effects		*Parameters:*				
		Tab. A3, Col. 1	0.5	2.3	4.0	5.9
25th Pctl	10.4	Tab. A3, Col. 2	0.5	2.4	4.1	6.0
50th	22.6	Tab. A3, Col. 2	0.6	2.6	4.5	6.6
Mean	27.5	Tab. A3, Col. 2	0.6	2.7	4.7	6.9
75th Pctl	38.6	Tab. A3, Col. 2	0.6	3.0	5.1	7.4

Estimates in Panel A are computed by applying the parameter estimates in Table 4, while Panel B uses estimates in Table A3. Land constraints are measured as % of land unavailable for construction and are obtained from Lutz and Sand (2019). For the definition of migration exposure see Section 3.2.

Table A5: Migration spillover applications: additional results

Dependent variable:	March '20 - July '21 CZ Population growth$_i$		
Pop. density		-0.22***	-0.22***
		(0.06)	(0.07)
Mig. exposure to density			1.52***
			(0.46)
Observations	602	602	602
Adj. R-squared	0.233	0.299	0.311
Controls	✓	✓	✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Migration network by distance includes all CZs for which the center of no component county is closer than the stated number of miles to the center of any focal city component county. See section 6 for details on the variables and specifications.

C Derivations

C.1 Derivation of the estimating equation from the static location choice model

In this section, I show how to derive the functional form for the reduced form analysis from the static model of inter-city migration introduced in section 3.1.

Starting from expression 2, first take the log of the migration choice probabilities and totally differentiate both sides to obtain

$$\begin{aligned}\Delta \ln \mu_t^{ki} =& \theta^{-1} a_i + \theta^{-1} w_{it} - \theta^{-1} p_{it} \\ & - \sum_j \mu_t^{kj} \left(\theta^{-1} a_j + \theta^{-1} w_{jt} - \theta^{-1} \alpha p_{jt} \right).\end{aligned}$$

Here, and going forward, lowercase letters denote log changes in the corresponding upper case variables. That is, any increase in the probability of migrating from city k to city i either comes from a change in the expected utility of the destination or a deterioration in the other available options for city k residents.

Next, note that equilibrium requires that the current local population in each city is the sum of all the city populations in the previous period that decided to move that city (or stay there):

$$\begin{aligned} & L_{it} = \sum_k \mu_t^{ki} L_{k,t-1} \\ \iff \quad & l_{it} = \sum_k \phi_{t-1}^{i \leftarrow k} \left(\Delta \ln \mu_t^{ki} + l_{k,t-1} \right)\end{aligned}$$

where $\phi_{t-1}^{i \leftarrow k} = \mu_{t-1}^{ki} L_{k,t-2} / \sum_j \mu_{t-1}^{ji} L_{j,t-2}$ is the share of city i's population coming from city k in the previous period, which represents city i's exposure to changes in other cities.

Now, we can substitute for the log change in migration shares from above to write

$$l_{it} = \Delta \mathcal{M}_{i,t-1} + \theta^{-1} \left[(a_{it} - a_{it}^{nw}) + (w_{it} - w_{it}^{nw}) - \alpha (p_{it} - p_{it}^{nw}) \right] \tag{14}$$

where I have defined the change in "migration access" $\Delta \mathcal{M}_{i,t-1} = \sum_k \phi_{t-1}^{i \leftarrow k} l_{k,t-1}$, and the notation $x_{it}^{nw} = \sum_k \phi_{t-1}^{i \leftarrow k} \sum_j \mu_{t-1}^{kj} x_{jt}$ denotes a migration-weighted sum of the log change in characteristic X over city i's entire migration network. Intuitively, this captures the degree to which each other city's options have changed (weighting those changes by the importance of those options to the other city) and how much that affects the focal city through the other city being a contributor to migration flows into it.

Substituting equation 4 into the log change version of equation 3, and rearranging, we can then write local house price growth as

$$p_{it} = \zeta \Delta \mathcal{M}_{i,t-1} + \alpha \zeta \theta^{-1} p_{it}^{nw} + \zeta \theta^{-1} (a_{it} - a_{it}^{nw}) + \zeta \theta^{-1} (w_{it} - w_{it}^{nw}) + \tilde{\varepsilon}_i^H t \tag{15}$$

where $\zeta = \frac{\eta^H}{1 + \alpha \eta^H \theta^{-1}}$ and $\tilde{\varepsilon}_i^H t = \Delta \varepsilon_i^H t / (1 + \alpha \eta_i^H \theta^{-1})$. In the baseline estimation, I ignore the potential effect of heterogeneity in inverse housing supply elasticity across cities However, in extensions, I will also present results that explicitly allows coefficients to vary with housing supply elasticity across cities.

Then, the baseline estimating equation is

$$p_{it} = \eta^{nw} p_{it}^{nw} + \xi_{it}^p, \tag{16}$$

which corresponds to equation 5 in the body of the paper, where $\eta^{nw} = \alpha \zeta \theta^{-1}$ is the network spillover effect, and $\xi_{it}^p = \zeta \Delta \mathcal{M}_{i,t-1} + \zeta \theta^{-1} (a_{it} - a_{it}^{nw}) + \zeta \theta^{-1} (w_{it} - w_{it}^{nw}) + \tilde{\varepsilon}_i^H t$.

C.2 Network IV exclusion restriction derivation

Shock-level reformulation of network IV. Following Borusyak, Hull, and Jaravel (2020), I can formulate the exclusion restriction at the level of industry shocks. For simplicity, I focus on just one instrument

$z_{it} = bp^{nw}_{it}$, but the analysis generalizes to the case of multiple instruments. The Frisch-Waugh Lovell Theorem implies that the spillover coefficient estimate in a panel of length T with N cities can be written as the second-stage coefficient of a residualized IV regression of the form

$$\hat{\eta}^{nw} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} z_{it} y_{it}^{\perp}}{\sum_{t=1}^{T}\sum_{i=1}^{N} z_{it} x_{it}^{\perp}}$$

where $y_{it} = p_{it}$ (the log change in city i house prices) for the main house price spillover estimation, and $x_{it} = p^{nw}_{it}$. The notation $\nu_{it}^{\perp}$ denotes the residual from a projection of ν_{it} on the vector of control variables Γ_{it} detailed in Section 3.6. Combining the definition of the network instrument in Equation 10 and the shift-share shock definition, we can write

$$\begin{aligned} z_{it} &= \sum_{j:j\neq i} \psi^{ij}_{\text{`}90\text{-`}95} x_j^{land} \sum_{\iota=1}^{N_{ind}} \tilde{\omega}_{\iota,j,\text{`}90} \Delta \ln W^{US}_{\iota,-j,t} \\ &= \sum_{\iota=1}^{N_{ind}} s_{i\iota} g_{\iota,t} \end{aligned}$$

where $g_{\iota t} \approx \Delta \ln W^{US}_{\iota,-j,t}\ \forall j$, ignoring the fact that the industry shock is constructed as a leave-one-out variable and therefore varies slightly between cities in a finite sample.[75] Here, I have combined the industry exposure, migration network structure, and land constraints, into a weight $s_{i\iota} = \sum_{j:j\neq i} \psi^{ij}_{\text{`}90\text{-`}95} \tilde{\omega}_{\iota,j,\text{`}90} x_j^{land}$ that summarizes the migration network exposure of city i to industry ι. We can then write

$$\hat{\eta}^{nw} = \frac{\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_{\iota} g_{\iota t} \bar{y}_{\iota t}^{\perp}}{\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_{\iota} g_{\iota t} \bar{x}_{\iota t}^{\perp}},$$

where the notation $\bar{\nu}_t = \frac{\frac{1}{N}\sum_{i=1}^{N} s_{i\iota}\nu_{it}}{\frac{1}{N}\sum_{i=1}^{N} s_{i\iota}}$ represents a weighted average across cities, with weights given by cities' relative exposure to industry ι through their migration network.[76] Moreover, $s_\iota = \frac{1}{N}\sum_{i=1}^{N} s_{i\iota}$ is the average city migration exposure to industry ι wage growth. This expression shows that we could estimate the same IV spillover coefficient in an industry-level regression of industry exposure weighted city house price growth on industry exposure weighted network house price changes, using national industry growth trends as the instrument.

This rewriting of the estimator at the industry level provides a perspective on the identifying variation underlying the network IV estimate: We can think of the spillover effect estimation as identifying the spillover effect from the degree to which the covariance in industry wage growth shocks with house price growth in the cities most exposed to the industry (the numerator) is higher in the cities that are "treated" in the form of having house price changes in their migration network that vary with industry wage shocks interacting with land constraints in other cities (the denominator). For instance, if cities (e.g. Boise, ID) with migration connections to land-constrained technology hubs (such as San Francisco) see higher house price growth precisely when tech hub house prices rise (e.g. in San Francisco) as a result of national trends in knowledge industry wages, then this variation allows us to infer a causal positive spillover effect.

Network IV exclusion restriction. This rewriting of the network IV estimator in the form of industry-level shocks then allows me to formulate the exclusion restriction of the network approach as follows:

Proposition 1 *If the network instrument bp^{nw}_{it} is relevant and mild regularity conditions hold (the variance matrix of control variables has full rank, and the covariance matrices of instruments and residuals with*

[75] This approximation is asymptotically correct as the number of cities gets large, and simplifies the notation and intuition substantially here. The main insight is robust to using the leave-one-out version.

[76] To see how I arrived at this expression, note that the numerator can be written as $\sum_{t=1}^{T}\sum_{i=1}^{N} z_{it} y_{it}^{\perp} = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{\iota=1}^{N_{ind}} s_{i\iota} g_{\iota,t} y_{it}^{\perp} = \sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_{\iota} g_{\iota,t} \left(\frac{\sum_{i=1}^{N} s_{i\iota} y_{it}^{\perp}}{\sum_{i=1}^{N} s_{i\iota}}\right) = \sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_{\iota} g_{\iota t} \bar{y}_{\iota t}^{\perp}$. Rewriting the denominator in a similar way gives the expression shown.

controls are bounded and exist), then the IV estimate of the spillover effect $\hat{\eta}^{nw}$ is consistent if and only if

$$\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_\iota g_{\iota t}\bar{\xi}_{\iota t}^{P,\perp} \to^p 0.$$

The proof is analogous to that for Proposition 2 in Borusyak et al. (2020). Here, $\bar{\xi}_{\iota,t}^{P,\perp}$ is the error in the house price growth regression in Equation 9, residualized with regard to the control variables Γ_{it}, and averaged over cities, weighting them by their migration network exposure to industry shocks.

This condition shows that, for the network IV estimate to be consistent, industry wage shocks cannot be systematically higher for those industries that have a systematically larger migration network impact on cities that are experiencing large unobserved house price shocks, conditional on control variables.

As Borusyak et al. (2020) argue, this identification allows for a city's migration network to be endogenously determined – it only requires the national industry trends over time to be exogenous in the sense defined in Proposition 1. This would be invalidated, if, for example, cities that experience more migration flows from land-constrained cities that specialize in the booming tech industry are also systematically experiencing greater idiosyncratic house price movements in a way that is not captured by their own exposure to knowledge industries or any other included control variables.

This approach also provides insights regarding the concern that the network house price instruments are correlated with focal city i industry shocks if industry structure is correlated across cities that share migration links. Note that this concern is supported by the significant coefficient on industry structure in the migration cost determinants analyzed in Section 2.2: Table 2 showed that migration costs appear to be lower among cities with similar industry structures, making them more likely to have strong migration links.

I can explicitly express this concern in the industry shock formulation of the exclusion restriction: if the direct shift-share shock B_{it} (and its interaction with city i land constraints) is not included in the control variables, and industry wage shocks affect city i house prices, then the house price residual $\xi_{it}^{P,\perp}$ would be a function of $B_{it} = \sum_\iota \tilde{\omega}_{\iota,i,`90}^{\perp} g_{\iota,t}$. Using the definition of $\bar{\xi}_{\iota,t}^{P,\perp}$, we can write

$$\bar{\xi}_{\iota,t}^{P,\perp} = \frac{\sum_{i=1}^{N} s_{i\iota}\xi_{it}^{P,\perp}}{\sum_{i=1}^{N} s_{i\iota}} = \frac{\sum_{i=1}^{N}\sum_{j:j\neq i}^{N}\psi_{`90\text{-}`95}^{ij}\tilde{\omega}_{\iota j,`90}\xi_{it}^{P,\perp}}{\sum_{i=1}^{N} s_{i\iota}} = \frac{\sum_{i=1}^{N}\sum_{j:j\neq i}^{N}\psi_{`90\text{-}`95}^{ij}\tilde{\omega}_{\iota j,`90}(\sum_\iota \tilde{\omega}_{\iota i,`90}^{\perp} g_{\iota,t} + \epsilon_{it})}{\sum_{i=1}^{N} s_{i\iota}}.$$

Note that this expression contains $\sum_{i=1}^{N}\sum_{j:j\neq i}^{N}\psi_{`90\text{-}`95}^{ij}\tilde{\omega}_{\iota j,`90}\tilde{\omega}_{\iota i,`90}^{\perp}$, which is the average migration-weighted covariance in (demeaned) industry employment shares of a city with other cities. If a city is more likely to receive migrants from other cities that are likely to have large industry employment shares in the same industries, then this covariance would be positive, and the exclusion restriction would not hold. By controlling for B_{it} directly in the regressions, I eliminate the systematic component related to common industry exposures (the first term in parentheses) from this bias as a source of potential violations to the exclusion restriction.

Note that it is not an issue that the direct effects of the Bartik industry shocks on the focal city that are included as control variables might be an imperfect proxy for unobserved (and potentially correlated) actual industry shocks that are affecting both a focal city and its migration network. If there is measurement error in the local industry exposure or if there are unobserved local productivity shocks that affect both the focal city and its migration network, these shocks will not be picked up by the IV estimation if they are not systematically correlated with the variation induced by the Bartik shocks.

Dynamic effect exclusion restriction. I also estimate IV forecasting regressions that correspond to local projections with external instruments (Jordà, 2005; Stock and Watson, 2018) of the form

$$p_{i,t-1+h} = \alpha_i + \alpha_t + \eta_h^{nw} p_{it}^{nw} + \beta'\Gamma_{it} + \xi_{i,t-1+h}^{P}, \tag{17}$$

where the vector Γ_{it} includes the same additional control variables as the static model (see Section 3.6), and I am again instrumenting for the time t network price growth shock using network labor demand shocks. The coefficients η_h^{nw} now represent the impulse response in period $t-1+h$ of the shock. That is, the contemporaneous impacts correspond to $h = 1$, the impact on the dependent variable in the year after the shock is $h = 2$, and so on.

As Stock and Watson (2018) note, to interpret $\tilde{\eta}_h^{nw}$ as causal dynamic effects we need the period t network instruments to not just be uncorrelated with contemporaneous price growth shocks $\xi_{i,t}^P$, but also all leads and lags of the shock. That is, the exclusion restriction becomes

$$\sum_{t=1}^{T}\sum_{\iota=1}^{N_{ind}} s_\iota g_{\iota t}\bar{\xi}_{\iota,t-1+h}^{P,\perp} \to^p 0. \ \forall h.$$

This condition requires that industry wage shocks are not predictable given the exposure weighted pattern of city house price growth residuals in *any* period. This is a stronger condition than for the contemporaneous effect. It would be violated, for instance, if house prices rise in anticipation of a technology boom in other cities that share migration links.

However, this restriction again only needs to hold conditional on the included control variables. I control, for example, for contemporaneous industry wage shocks in the focal city and its migration network, as well as regional trends and commuting zone fixed effects. This means that, to the degree that the economic dynamics anticipated by past house price run-ups in the focal city are captured by these contemporaneous control variables, industry shocks will be exogenous *conditional* on including these covariates.

Supply constraints as instruments. Note also that the Davidoff et al. (2016) critique of supply constraint instruments is not applicable to the network instruments constructed here: my main analyses identify effects off variation in Bartik shift-share shocks *over time,* and the supply constraints only matter in so far that they form part of the exposure term in the shift-share instrument construction. As Borusyak et al. (2020) show, it is not necessary for these exposure terms (and therefore the supply constraints) to be exogenous with regard to local unobservable shocks, as long as the shifter in the form of a national industry trend is exogenous. Moreover, I control directly for the interaction of the focal city's shift-share shock with local supply constraints (see Section 3.4) – and the identifying variation is only coming from the residual effect of *other* cities' supply constraints interacting with shift-share shocks. For these reasons, the Davidoff et al. (2016) critique of supply constraints used in IV settings is not relevant in my setting.

D Data appendix

D.1 House price beta estimation

House price growth "betas" for each commuting zone are computed as the slope coefficient in a regression of each CZ's annual house price growth series 1990-2017 from Federal Housing Finance Authority data on the series of leave-one-out population-weighted average national house price growth in each year.

That is, I construct for each commuting zone and year the leave-one-out national average of house price growth as

$$p_{it}^{loo,US} = \sum_{j:j\neq i} \frac{L_{jt}}{L_{US,t}} \cdot p_{jt},$$

where p_{jt} is the log change in the FHFA house price index in commuting zone j in year t, and L_{jt} and $L_{US,t}$ are IRS population estimates for commuting zone j and the full sample of US commuting zones in the data, respectively.

Then, I restimate the following regression separately for each commuting zone i in a 1990-2017 panel:

$$p_{it} = \alpha_i + \beta_i p_{it}^{loo,US} + \epsilon_{it},$$

where the estimates $\hat{\beta}_i$ represent the house price beta estimates for the 1990-2017 period for each of the 741 commuting zones.

D.2 Migration network weight construction

This section discusses how the fixed migration exposure weights are estimated. In Section 3, I show that the migration exposure of city i to city j house prices is measured by

$$\psi^{ij} = \sum_{k \in N} \underbrace{\phi^{i \leftarrow k}}_{\substack{\text{In-migration} \\ \text{share for } i}} \underbrace{\mu^{k \rightarrow j}}_{\substack{\text{Out-migration} \\ \text{share for } k}} .$$

I use these migration exposure weights to construct the migration network house price changes, as well as the network instruments and any control variables measured at the migration network level.

One interpretation of this measure is that economic shocks to other cities can have effects on city i because of their *direct* migration connection, or due to *indirect* links as they share migration connections to another city in city i's migration network. This is analogous to the way that industry productivity shocks can indirectly affect other industries through shared local labor markets in network models in macroeconomics (Acemoglu, Akcigit, and Kerr, 2016).

There are several empirical issues to consider in constructing these migration weights: On the one hand, with a finite number of observations, actual year-to-year migration flows between cities are only a noisy measure of the true latent out- and in-migration probabilities. Therefore, μ_{t-1}^{ik} and $\phi_{t-1}^{i \leftarrow k}$ measured in a particular year might not capture true latent probabilities for city pairs with limited realizations in the data (Dingel and Tintelnot, 2020). In order to make these estimates of migration network connections more precise we would therefore want to average over several years.[77]

On the other hand, if migration is to some degree endogenous – as is one of the key arguments of this paper – then the migration network weights of other cities would fluctuate over time if they are measured on an ongoing basis. This would make any estimated effects difficult to interpret as they would combine changes in weights with changes in network prices. Moreover, if the effects of network price changes are auto-correlated, then migration network weights would anticipate future network price changes to some degree, leading to bias if price changes are not exogenous (Jaeger, Ruist, and Stuhler, 2018).

As an empirical compromise that addresses these concerns, I therefore hold the migration network weights fixed across years at baseline period values, measured using average migration flows over 1990-1995. The migration during this period, besides being the first years of migration data in the IRS migration flows sample, also has the advantage of preceding the dramatic house price boom-and-bust cycle that started in the mid-1990s, and is therefore unlikely to be caused by the house price dynamics that are of the greatest policy interest over the last three decades. Besides, holding the migration weights fixed at a baseline period level has a strong precedent in the literature using historical migration shift-share instruments (Altonji and Card, 1989; Boustan, 2010; Howard, 2020; Derenoncourt, 2019).

D.3 Bartik shift-share shock construction

This section summarizes the construction of the Bartik (1991)-style shift-share instrument used in the network IV estimation.

The fundamental idea behind these shift-share shocks is that, under some assumptions, one can obtain an instrument for exogenous local wage changes by using the interaction between local exposure to national industry trends with the size of those trends. Identification of the effect of local wage changes in such a setting follows either from the exogeneity of the cross-sectional variation in exposure (Goldsmith-Pinkham, Sorkin, and Swift, 2018), or the exogeneity of national industry trends with regard to the exposure patterns (Borusyak et al., 2020).

I construct log wage growth instruments by combining an exposure term in the form of the local wage bill share $\tilde{\omega}_{\iota,i,t_0}$ of workers in industry ι in a baseline period t_0 with shifters consisting of national wage

[77] Another approach to smoothing them statistically is shown in the construction of the migration-by-education-group data set in Appendix D.7.

growth $\Delta \ln W^{US}_{\iota,-i,t}$ in each industry ι in the form

$$B_{it,t_0} = \sum_{\iota} \tilde{\omega}_{\iota,i,t_0} \Delta \ln W^{US}_{\iota,-i,t}.$$

The industry averages of log wage growth are computed as leave-one-out measures to avoid mechanical correlation between the national trend estimate and city i wages (Borusyak et al., 2020). The source of industry data for the shift-share shocks is the Quarterly Census of Employment and Wages (QCEW).

Baseline period. In order to minimize bias from endogeneity in the local industry exposure – which might result from auto-correlated national industry shocks (Goldsmith-Pinkham et al., 2018; Jaeger et al., 2018) – I fix industry exposure shares at the baseline level in some period t_0 for each analysis. This choice trades off efficiency of the instrument to prevent bias. In particular, the farther away the baseline period is from the years in the analysis, the less likely the baseline year exposure shares are to reflect the effect of national trends, because employment shares may have changed over time. The earliest period for which I have industry-level employment data – and also the first year of the panel used in my analysis – is 1990. The analysis thus uses 1990 industry employment shares for the full 1990-2017 panel of shift-share instruments.

Industry aggregation level. The other data choice to be made in constructing the shift-share instruments is the level of industry aggregation to use in defining exposure and estimating wage trends. The trade-off involved in this decision is that using a more detailed industry definition for shift-share instruments (e.g. using 2-digit instead of 3-digit NAICS industries) will be less reflective of actual exogenous local wage shocks over time if local industry structure changes more between narrow categories than broad sectors, or if there are spillovers. For example, if the local presence of Credit Intermediation (NAICS 522) firms is linked to the local growth of Funds & Trusts (NAICS 525), then a shock to one of these narrow industries might also affect the other one. Moreover, firms or employees might shift between these industries. However, narrowly defined shift-share exposure would miss these effects. In addition, the QCEW data is more likely to suppress employment counts in narrow industries (for anonymity reasons) than in broader sectors, such that the estimate of actual industry structure might be noisier at the more detailed level. On the other hand, a broader industry definition might introduce noise due to the fact that growth in a particular 3-digit subsector at the national level might not affect a different subsector that is contained within the same 2-digit sector code. Aggregating across these subsectors might therefore mistakenly infer shocks to cities' local industries where there are none. To better capture local differences in shock exposure, I use the more detailed 3-digit level aggregation.

Geographic aggregation. This paper uses the full set of continental U.S. commuting zones based on 1990 boundaries. That is, I compute shift-share instruments at the level of this geographic unit after aggregating any county-level data to the commuting zone level.

Identifying variation. If there is heterogeneity in the causal effect of interest, the shift-share instruments, like any IV approach, will identify a local average treatment effect that represents a weighted average of the causal effects at different units. As is well known, the weights will depend on the degree to which the shift-share instruments represent larger shocks to some markets than others (Angrist, Graddy, and Imbens, 2000). Therefore, it is of interest to understand the underlying identifying variation in the shifters used in the construction of these instruments. Two margins of variation are particularly relevant for my analysis: geography and industry.

First, differences in wage trends across industries are driving the variation over time and space in the shift-share shocks in different CZs. These industry wage trends are shown for aggregated industry supersectors in Appendix Figure A10. The graph shows that more high-skilled white collar sectors, such as information, finance, and business services have done comparatively well in recent decades, while manufacturing and construction, for instance, have fared less well in term of wages.

Second, the cross-sectional differences in city exposure to wage shocks, either directly or through their migration network come from the fact that cities vary in their industry structure. For instance, larger cities have benefitted more from the technological changes favoring skilled services industries (Eckert, Ganapati, and Walsh, 2019). Figure A11 illustrates this identifying variation by plotting the difference in wage trends by tercile of city size. As the graph shows, wage growth has been higher in larger cities than in smaller cities, with larger cities experiencing double the wage growth of small cities during 1990-2005.

Overall, these graphs illustrate that the shift-share shocks based on differences in location exposure to

Figure A10: Wage trends by sector. The graph plots average annual wage growth by NAICS super sector. The data shown are average wages from the Quarterly Census of Employment and Wages, converted into real wages by adjusting for the urban consumer CPI, and indexed to 1990 values.

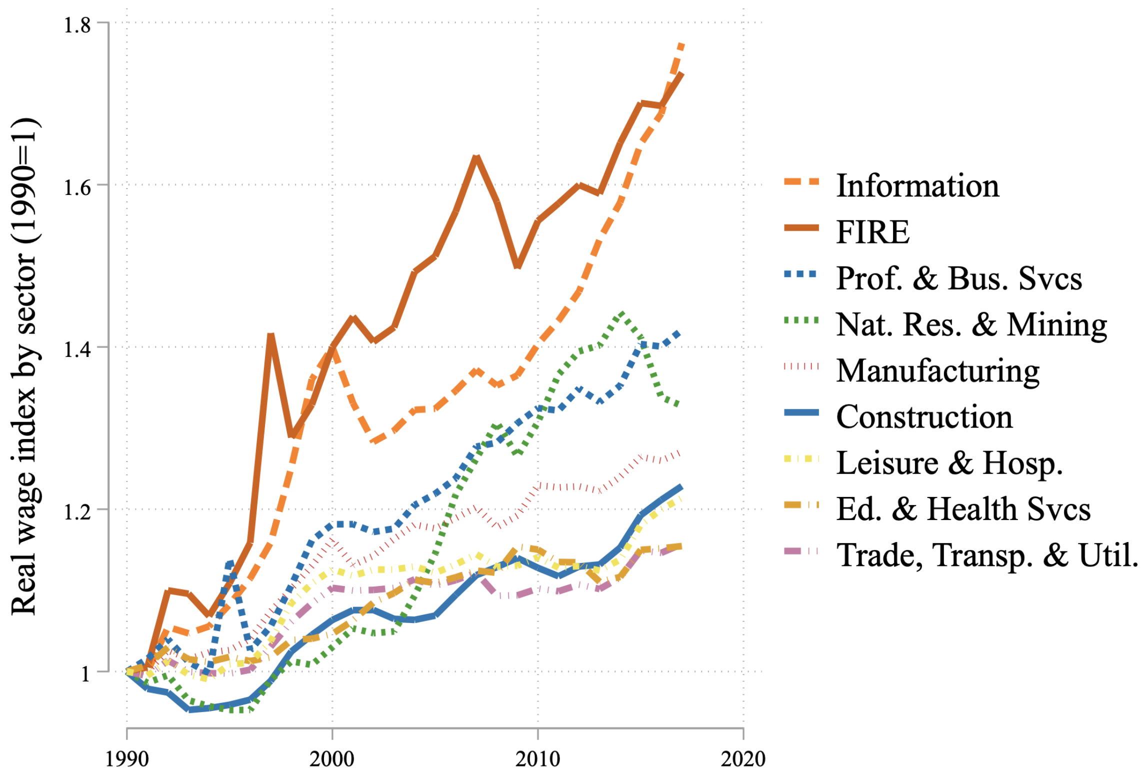
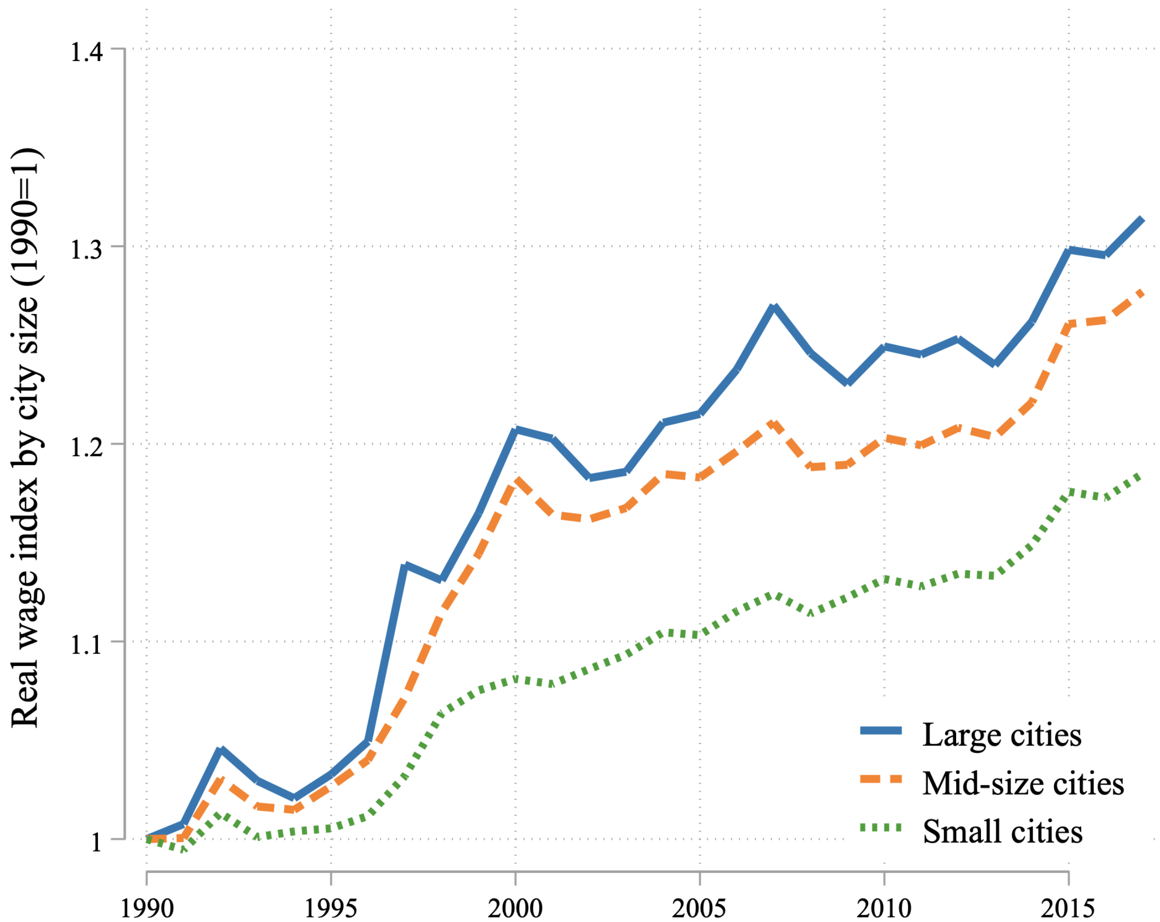


Figure A11: Wage trends by city size. The graphs plot average annual wage growth by different city and education groups. The data shown are average wages from the Quarterly Census of Employment and Wages, converted into real wages by adjusting for the urban consumer CPI, and indexed to 1990 values. The subgroups in the graph are terciles of Commuting Zone population size.



national wage trends can generate substantial exogenous cross-sectional and time variation in wages, which I use to identify the parameters of interest in the network analysis.

Constructing house price instruments from wage shift shares. To construct shift-share shocks that can be used as instruments for local changes in house prices, I follow a large empirical literature (see, e.g. Diamond (2016)) in interacting wage shift share shocks with a proxy for exogenous housing supply constraints. As the measure of constraints, I use local land unavailability for construction x_i^{land} from Lutz and Sand (2019).[78] This measure captures geographic constraints to marginal housing construction, which would be expected to increase the slope of the housing supply curve, and thereby increase the responsiveness of house prices to the wage shocks. I aggregate county-level measures of land constraint into commuting zone averages, weighting the county-level measure by the county population to reflect the likelihood that a city's residents are constrained by the geography.

[78]These are comparable to the Saiz (2010) land availability measures commonly used in the literature. Lutz and Sand (2019) build on his methodology to expand the number of covered cities and, among other things, improve the measurement of land availability for overlapping city areas and coastal locations.

D.4 Amenities index constructions

I follow Diamond (2016) and Almagro and Domínguez-Iino (2020) in measuring the availability of local amenities by using counts of local establishments in particular industries from the County Business Patterns data. These establishments are chosen based on their importance for serving tourists and residents for leisure activities, and therefore serve as proxies of how attractive amenities in a location are. Both Diamond (2016) and Almagro and Domínguez-Iino (2020) provide evidence that different demographic groups may value amenities differently. Supporting evidence that changes in local establishments are associated with changes in demographics is provided by Glaeser, Kim, and Luca (2018) who note that a change in the number of certain establishments such as wine bars can predict an increase in the average education level of a neighborhood.

Informed by these papers, I use establishment counts in the following NAICS codes as proxies for local amenities:

- Food Services and Drinking Places (NAICS 722)
- Grocery store industry (NAICS 4451)
- Motion picture theaters (NAICS 512131)
- Dry Cleaners (NAICS 8123)
- Clothing and Accessory Stores (NAICS 448)
- Museums, historical sites, and similar institutions (NAICS 7121)
- Sports teams (NAICS 71211)
- Scenic and Sightseeing Transportation (NAICS 487)

I use the crosswalks for different NAICS classifications from Eckert, Fort, Schott, and Yang (2020) together with the county-level annual CBP data for 1989-2016 to create a panel of establishments by county by NAICS 2012 code. Then, I aggregate the county-level data to 1990 commuting zones using the David Dorn crosswalk.

Moreover, for use in the regressions I scale the number of establishments in these categories by the IRS return-based population estimate for each CZ in order to express amenities in the form of a density of establishments per 1,000 residents.

The amenities index for each city is then computed as the first principal component $Amen_{it}^{1st}$ of the counts of establishments per 1,000 residents in the industries enumerated above that provide cultural or consumption services.

D.5 Trade flow link construction

One alternative measure of city-to-city links that I consider is the value of trade flows between them. This section describes how I construct a measure of the trade flow link between cities.

Trade flow data source. As the basis for the computation of trade flows between areas, I use the Commodity Flow Survey (CFS) Public Use Microdata for 2012 provided by the U.S. Census Bureau. This data is provided at the level of CFS areas, which are aggregations of counties, and captures the value and weight of individual shipments between CFS areas.

Computing CFS area trade links. First, I aggregate the microdata into origin-destination CFS area pair shipment values by summing across individual items, adjusting for a given weight factor that captures the representativeness of each line item. Then, for each origin CFS area, I compute the share of total trade value sent to each other CFS area. This is the measure of the strength of trade links between areas, which I then crosswalk to commuting zones.

Area crosswalk to commuting zones. I use a crosswalk from CFS areas to counties and from counties to 1990 commuting zones to probabilistically map CFS areas to commuting zones. CFS areas tend to be larger than commuting zones: the continental U.S. data contains 70 CFS areas, which I map into 169 commuting zones, and only 8 CZs map into more than one CFS area. I assume that CZs that are fully contained within a CFS area inherit all of the trade links of that CFS area. That is, their links with other CFS areas and the CZs contained within them are the same as for the CFS area that they are a part of. Where a CZ contains counties that are part of different CFS areas, I take a weighted average of the links of those CFS areas with other areas, weighting them by the share of the respective counties in the CZ's total population in the year 2000. The result of this crosswalk is a CZ-to-CZ measure of the *relative* strength of trade links with other CZs from the perspective of a CZ as the origin of trade flows.

D.6 Quality-adjusted house price index

In order to be able to compare house prices across cities, it is necessary to adjust them for differences in the quality of the housing stock. I follow Albouy and Ehrlich (2018) in constructing a quality-adjusted index of commuting zone-level house prices for the year 2000. House price data are obtained from the United States Census Integrated Public-Use Microdata Series (IPUMS), from Ruggles, Flood, Goeken, Grover, Meyer, Pacas, and Sobek (2020).

House price indices for each commuting zone j are calculated from a 5% sample from year 2000 Census. The sample is restricted to owner-occupied units. I regress the logarithm of house value $\ln P_{ij}$ for each household i on hedonic control variables X_{ij}, and indicator variables for each CZ cell. The regression specification is

$$\ln P_{ij} = \beta' X_{ij} + \psi_j + \epsilon_i,$$

where the estimated CZ fixed effects ψ_j are then treated as the commuting zone-level house price premia or discounts.

The housing characteristics X_{ij} included in the regression are:

- 10 indicators of number of units in the building
- 9 indicators for when the building was built (by decades or 5-year spans)
- 9 indicators for the number of rooms, 6 indicators for the number of bedrooms, as well as indicators for number of rooms interacted with number of bedrooms
- 2 indicators for complete plumbing and kitchen facilities

I follow Albouy & Ehrlich (2018) in estimating the regression in two stages: First, the regression is run weighting by census-housing weights, adjusted by the weight of the PUMA in the commuting zone (following David Dorn's crosswalks). A new value-adjusted weight is calculated by multiplying the CZ-adjusted census-housing weights by the predicted house value from this first regression using housing characteristics alone, but omitting CZ differences. A second regression is run using these new weights on the housing characteristics, along with the CZ indicators. The housing-price indices are obtained from the CZ fixed effect variables estimated in this second regression.

In order to compare house prices across cities, I then compute the predicted price of a "standard" home in each city, where "standard" is defined as using the mode of each variable in the hedonic housing characteristics. The baseline price is thus calculated for a single-family detached home, built in the 1990s, with 3 bedrooms and 6 rooms overall, and with complete kitchen and plumbing facilities. To this baseline home value is added the commuting zone house price indicator to compute quality-adjusted prices in each commuting zone.

This index is the converted into a panel under the assumption that quality differences are constant at their year 2000 values by constructing CZ-level values for other years using the FHFA repeat-sales index. That is, I apply the local house price growth in the FHFA to the year 2000 house price index to construct a 1990-2017 panel of house prices that has been adjusted for year 2000 quality differences.

D.7 Migration flow by education group

In order to be able to estimate location choice parameters separately by education group, I need a data set of annual migration probabilities between city pairs for both college- and non-college-educated workers. Unfortunately, no single data set provides this information for U.S. cities over a long time horizon.

Given enough data, we could simply calculate the empirical choice probabilities for each state of interest where the state space consists of worker types, origins, destinations, and years. Unfortunately, the IRS data used in the reduced form estimation does not allow for distinguishing flows of different groups of workers. The best publicly available migration data at an annual frequency for the U.S. that also contains migrant characteristics, is from the American Community Survey (ACS) for 2005-2017, with sample sizes in the single-digit millions for most years. However, even in the most basic setup, the state variables consist of all permutations of high and low worker skill with potential origin continental U.S. commuting zones and destination commuting zones, so up to $2 \times 512 \times 512 = 524,288$ different states in each year. As a result, the raw estimates of CZ-to-CZ migration flows by education are noisy because some cells are necessarily estimated with a small sample size.[79] To improve the precision of my estimates, I therefore apply statistical techniques for data smoothing and imputation that allow me to make use of additional information contained in the ACS data, combine information across units and years, and incorporate information from the IRS migration data.

Therefore, in order to overcome the limitations of individual data sets, I decompose the migration probabilities by education group into components that can be reliably estimated and employ statistical tools to mitigate weaknesses in individual data sets, for instance by combining information from multiple sources.

In particular, I decompose the share of people in education group s in city i moving to city k as follows:

$$\mu_{st}^{ik} = \underbrace{\frac{m_{st}^{ik}}{\sum_s m_{st}^{ik}}}_{\text{Skill } s \text{ share in } i \to k \text{ flows}} \cdot \underbrace{\mu_{out,t}^{ik}}_{\text{Share of } i \to k \text{ in all out-mig.}} \cdot \underbrace{(1-\mu_t^{ii})}_{\text{Out-mig. rate from } i} \cdot \underbrace{\frac{L_{i,t-1}}{L_{is,t-1}}}_{1\ /\ \text{Pop. share of type } s} \tag{18}$$

where $m_{st}^{ik} = \mu_{st}^{ik} L_{is,t-1}$ are the number of migrants of type s moving from i to k in period t. As noted above, I estimate these components separately, combining data from both the ACS and the IRS.

The construction of my data set of flows by education between commuting zones proceeds in several steps: (1) Imputing CZ-to-CZ total migration flows $\mu_{out,t}^{ik}$ from IRS data. (2) Predicting education shares for flows. (3) Smoothing gross migration rates $(1-\mu_t^{ii})$ to account for IRS data issues. (4) Combining the components of migration rates into predicted pairwise migration probabilities.

Imputing CZ-to-CZ flows from IRS migration data. In general, I use the aggregate data from the Internal Revenue Service Statistics of Income (IRS SOI)[80] based on the near-universe of tax returns as a measure of the relative size of overall migration flows between locations. The IRS counts moves in the form of changing adresses on tax returns and also totals up exemptions claimed on moving tax returns – which I will use as a proxy for the number of people moving.

The IRS data comes in the form of county-to-county flows, as well as total flows in and out of each county and flows to and from some aggregate regions. A peculiar feature of the IRS data is that it only records county-to-county flows by name when the flows add up to at least 10 tax returns for the pair. Smaller "unnamed" flows are, however, included in the out- and inflow totals as well as in more aggregate flow totals by census region.

In order to avoid county pairs dropping in and out of the data if they become too small and to capture the fact that true migration probabilities are unlikely to be precisely zero (Dingel and Tintelnot, 2020), I impute small flow probabilities by allocating unnamed outflows from each county using a hierarchical "empirical Bayes" approach. I assume that the likelihood of flows between regions and to cities within regions can be described as a multinomial distribution, the parameters of which come from a Dirichlet distribution. Then, I use observed aggregate information on regional flows, and observed named flows as empirical estimates of the prior for flow shares going to each city. These priors are then used to allocate unnamed flows to cities. Last, allocated unnamed flows and named flows are smoothed using again a Multinomial-Dirichlet model to estimate non-zero latent probabilities of pairwise migration between cities.

[79]While the effective number of states is substantially smaller as the migration flow matrices are sparse, a more elaborate state space, for instance taking into account migrant age or state of birth, would expand the number of states to estimate even further. Moreover as migrants only represent a small share of the surveyed individuals in each year, the effective sample size for migration purposes is much smaller.

[80]Available at URL: `https://www.irs.gov/statistics/soi-tax-stats-migration-data`

Allocating unnamed outflows to destination regions. To allocate the unnamed outflows N_c^{out} from each county c, I follow the following process: First, I subtract any foreign flows from the unnamed flows to obtain domestic unnamed flows. Second, I initially assign these unnamed flows to the regional totals listed as going to the same state or out-of-state to one of the four Census regions (Northeast, Midwest, South, West). Any remainder is assigned equally to each of these 5 categories. Third, I smooth the allocation of unnamed flows by assuming that the distribution across the 5 destination categories follows a multinomial distribution, with probabilities θ_{ic}^{reg} of an unnamed flow from c going to a destination in each of these categories i. That is, I assume that the probabilities θ_{ic}^{reg} of an unnamed flow going to a destination in each of these categories have a joint density

$$p(\theta_{1c}^{reg}, \ldots, \theta_{5c}^{reg} | \boldsymbol{\alpha}) \propto \prod_{i \in \{1,5\}} (\theta_{ic}^{reg})^{\alpha_i - 1},$$

where $\sum_{i \in N} \theta_{ic} = 1$ and the parameters α_i characterize the prior distribution. I choose a non-informative prior of $\alpha_i = 1 \, \forall i$ (a uniform distribution over all possible values of θ_{ic}^{reg}), and treat the allocated unnamed migration flows to each region as data $\boldsymbol{y}$ that is used to update the parameter estimate (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin, 2013). The marginal means of the posterior distribution will be given by

$$E[\theta_{ic}^{reg} | \boldsymbol{y}] = \frac{n_{c \to i} + \alpha_i}{N_c^{out} + \sum_i \alpha_i},$$

where the $n_{c \to i}$ are the unnamed flows initially allocated to region i. The posterior estimate of the unnamed flows to each region is then given by $\hat{n}_{c \to i}^{reg} = E[\theta_{ic}^{reg} | \boldsymbol{y}] \cdot N_c^{out}$. I compute $\hat{n}_{c \to i}^{reg}$ for each county in each year.

Allocating unnamed outflows to counties within regions. Once unnamed outflows are allocated to regions, I distribute them across counties in each region. For each region, unnamed flows can only be allocated to counties that do *not* already have named outflows registered in the IRS data. Again, I assume that flows to each county within a region follow a multinomial distribution. For each of those potential destination counties within a region, I again form a non-informative prior with a uniform density over migration probabilities, i.e. $\alpha_k = 1 \, \forall k$ for the share of flows to that region that would go that county. The data used to update that prior is the share of all *named* flows to any county in that region – from other counties in the set C_{all} – that go to each county. That is, I use the data of observed named flows to update my prior for allocating unobserved flows, which can be thought of as an approximation to a complete hierarchical Bayesian analysis (Gelman et al., 2013).

To be precise, the Dirichlet posterior mean of the share θ_{kic}^{cty} of outflows from c going to county k, conditional on going to region i is

$$E[\theta_{kic}^{cty} | \boldsymbol{y}] = \frac{\sum_{j \in C} n_{j \to k} + \alpha_k}{\sum_{k \in C_{ic}^{un}} \sum_{j \in C_{all}} n_{j \to k} + \sum_{j \in C_{ic}^{un}} \alpha_j},$$

where C_i^{un} is the set of counties that are potential unnamed destinations in region i from the perspective of origin county c. Then, final imputed unnamed flows from county c to county k in region i are given by $\hat{n}_{c \to k} = \hat{n}_{c \to i}^{reg} \cdot E[\theta_{kic}^{cty} | \boldsymbol{y}]$, calculated separately for each year.

Last, I apply David Dorn's crosswalks to aggregate county-to-county flows to 1990 commuting zones (CZ), combining small CZs into entities of at least 50K residents, and treating within-CZ flows as non-migrants. Then, I drop any flows from and to CZs in Hawaii and Alaska, focusing on cities in the contiguous U.S., as well as flows to and from New Orleans, which experiences highly abnormal migration flows as a result of Hurricane Katrina in 2005/2006. For the remaining CZs, I aggregate all observed or imputed outflows to the other CZs in the sample and compute the share of outflows going to each destination CZs, which gives me the conditional migration probability $\mu_{out,t}^{ik}$.

Gross migration rates. The previous imputation was for migration shares conditional on leaving the CZ. I also use the IRS data to compute the overall level of out-migration $1 - \mu_t^{ii}$ from each CZ in each year. However, the IRS changed its methodology for inferring moves from tax returns starting with the tax returns for income year 2011, generating a break in the series of gross migration rates (Molloy and Smith, 2019). After this transition, gross migration rates in the IRS data seem to exhibit a larger-than-normal year-to-year volatility that is not supported by other data sources (e.g. the CPS migration rates). However, there is no reason to believe that this change in gross migration rates affects *relative* migration shares for different destinations and differences between CZs in migration activity systematically.

To remove this extraneous volatility in gross migration in later years from the data, I impute the overall level of gross migration for each CZ in the following way: I regress aggregate population-weighted averages of gross inter-city outmigration from the IRS ($Mig_{IRS,t}$) on the equivalent measure of gross migration to the continental U.S. in the American Community Survey ($Mig_{ACS,t}$) in a regression of the form

$$Mig_{IRS,t} = \alpha + \beta Mig_{ACS,t} + \epsilon_t$$

for the period 2005-2010. I predict gross IRS migration for 2011-2017 from the observed ACS values and the predicted relationship from the regression. Then, I impute individual CZ gross migration rates by applying the observed ratio of their gross out-migration to the average in the IRS data to the new imputed average value for post-2011 data. That is, I compute

$$\widehat{1-\mu_t^{ii}} = \widehat{Mig}_{i,t} = \frac{Mig_{i,IRS,t}}{Mig_{IRS,t}}\left(\hat{\alpha} + \hat{\beta} Mig_{ACS,t}\right),$$

which is my measure of overall CZ outmigration rates. While I could, in theory, rely only on the ACS data to construct the same flows, the ACS data is meant to capture relative representation of population at the local level, but is not designed to provide an accurate count of population (and therefore migration) *levels*. The IRS data, which uses the universe of tax returns therefore seems better suited to estimate the size of total flows between cities. The ACS data is used instead to estimate the *composition* of flows and local populations.

Empirical Bayes shrinkage estimate of education flow shares. Next, I compute the share of the different education groups in flows between CZs. Because the ACS data on city-to-city flows is very noisy on an annual basis, so is the share of these flows that can be attributed to different education groups. However, due to the fact that migration by skill share needs to fulfill adding-up constraints, I can use information from other parts of the ACS data to predict pairwise flows and then apply an empirical Bayes shrinkage estimator to combine these predicted skill shares with the actually observed values.

I proceed in three steps: First, I use a version of the post-LASSO estimator of Belloni, Chernozhukov et al. (2013) to predict education shares for city-to-city flows. This involves first applying a Least Absolute Shrinkage and Selection Operator (LASSO) to select the origin and destination city characteristics that best predict the observed non-college share nc_i of migrants between two cities. This estimator chooses the coefficients on predictive characteristics by solving:

$$\hat{\beta}^{\text{Lasso}} = \arg\min_{\beta} \sum_{i}^{n} \left(nc_i - \sum_{j}^{p} x_{i,j} b_j \right)^2 + \lambda \sum_{j}^{p} \|b_j\|$$

where i indexes a CZ pair-year unit, and $x_{i,j}$ represents the candidate characteristics for non-college city pair flow share prediction, which consist of: the non-college share among all inflows into the origin city and destination city; all migrant outflows from the origin city; the non-college share of the total population in the origin and destination cities; the log of the total and non-college population levels in the origin and destination cities; and a constant. The penalty term λ is chosen by cross-validation with 10 folds, and the estimator applied to the pooled city pair data for 2005-2017 selects a non-zero coefficient for 6 out of the 10 candidate variables, omitting the total population levels, the log of the destination non-college populations, and the non-college population share in the origin city. The selected non-zero coefficients with non-zero coefficients are then used in an OLS forecasting regression, and the OLS coefficients $\hat{\beta}^{\text{OLS}}$ are used to forecast a predicted non-college share

$$\hat{nc}_i^{pred} = X_i \hat{\beta}^{\text{OLS}},$$

and compute the standard error of the non-college share forecast s_i^{pred}. To reduce outliers generated by the linear forecast, I cap predictions to be no bigger than 99%.

Second, I compute the estimated non-college shares of city-to-city flows $\hat{nc}_i^{ACS}$ in the ACS microdata. The standard error of the ACS non-college share $\hat{nc}_i^{ACS}$ in flows computed from IPUMS microdata (Ruggles et al., 2020) can be computed approximately as (Bureau, 2005–2017):

$$s_i^{ACS} = df \cdot \sqrt{\frac{99}{\hat{n}_i^{ACS}} \hat{nc}_i^{ACS}(1 - \hat{nc}_i^{ACS})},$$

where $\hat{n}_i^{ACS}$ are the total flows observed for city pair-year i in the ACS data – the denominator of $\hat{nc}_i^{ACS}$ – and df is a design factor that reflects the ACS sample design and is obtained for each year from Bureau (2005–2017).[81] Note that for non-college shares close to zero or one, I follow the guidance in Bureau (2005–2017) and substitute 2% and 98%, respectively, for the purpose of the standard error calculation.

Third, I combine the predicted value $\hat{nc}_i^{pred}$ with the noisy observed non-college share $\hat{nc}_i^{ACS}$ for each city pair in each year by taking a weighted average, with weights that account for the relative uncertainty of the raw ACS estimate and the predicted value. To combine the two estimates, we can treat the predicted value as a prior with normal distribution $\mathbb{N}(\hat{nc}_i^{pred}, s_i^{pred})$, and the ACS data as being generated by a process with a normal distribution $\mathbb{N}(\hat{nc}_i^{ACS}, s_i^{ACS})$. Then, for any symmetric loss function, the optimal Bayes posterior estimator for the non-college share is

$$\hat{nc}_i^{eb} = \left(\frac{s_i^{ACS}}{s_i^{ACS} + s_i^{pred}} \right) \hat{nc}_i^{pred} + \left(\frac{s_i^{pred}}{s_i^{ACS} + s_i^{pred}} \right) \hat{nc}_i^{ACS}.$$

Intuitively, this "empirical Bayes" estimator $\hat{nc}_i^{eb}$ adjusts the raw non-college shares by moving them towards their expected value – "shrinking" the deviation — and does so to a greater degree if the raw estimate was based on a smaller sample size, and thus has a greater standard error. This approach to noise reduction by combining two estimates with one being treated as a Bayesian prior, although it is empirically constructed, is often called "empirical Bayes shrinkage" and has had a long history of statistical applications (Morris, 1983; DuMouchel and Harris, 1983; Gelman et al., 2013). One advantage of this method is that I can impute skill shares even for migration city pairs that are not observed in the ACS data, but for which the IRS data records migration flows. For values missing in the ACS data, I assume $s_i^{ACS} \to \infty$, such that the estimator loads entirely on the predicted value $\hat{nc}_i^{pred}$ in the estimation. As a result, this method yields an estimate of the non-college share - and therefore also the college share - of the migrant flows between each city pair and for each year. I use these estimates of $\frac{m_{st}^{ik}}{\sum_s m_{st}^{ik}}$ in equation 18 to compute the city pair migration probabilities by education group.

As this approach to data cleaning is "statistical" rather than fully model-dependent, there might be a concern that the smoothing and imputation procedure removes variation of interest from the data. However, for large pairwise migration connections, the effect is minimal as the empirical observation dominates the weak prior in generating migration probability estimates. For the city pairs with zero observations where the imputation procedure and allocation of unnamed flows makes relatively large changes relative to the observed flows, we know that the "zero" flow estimate is wrong for some city pairs from the discrepancy between total and named flows. Assuming these flows to be precisely zero would therefore be equivalent to discarding this information. This issue is exacerbated when taking logs of the flows: a common approach in the applied literature to this "zero flow" problem is to discard the zero observations implicitly when taking logs, use ad hoc adjustments such as adding 1 to each observation, or use nonlinear estimators that treat the observations as precisely zero. By using information available in other parts of the data set to determine which of the censored observations are more or less likely to actually be zero, I therefore think that I am adding information relative to these alternatives.

[81] The design factors obtained from Bureau (2005–2017) for the applicable characteristic "Residence 1 Year Ago" is 3.0 for 2005–2008, 2.9 for 2009–2011, and 2.8 for 2012–2017.

E Supplementary Analyses

E.1 Descriptive analysis of drivers of city attractiveness over time

While persistent migration costs determine *which* cities are more likely to be part of the same migration network, the variation over time in the level of migrant flows into or out of a city will depend on changes in the overall attractiveness of that city. In order to get a descriptive sense of what factors correlate with migration flows, and to supplement the causal effect estimates in Section 4, this section details non-causal estimates of the drivers of bilateral flows between cities.

As a first pass at determining which city characteristics have been associated with more or less migration, I take the destination city × year fixed effects obtained from the migration gravity regression shown in equation 1 and regress them on city characteristics, a national trend α_t, and – in some specifications – a city fixed effect α_j:

$$\theta_{jt} = \alpha_t + \alpha_j + \alpha_w \ln W_{jt} + \alpha_P \ln P_{jt} + \beta' \mathbf{X_j} + \epsilon_{ijt}$$

The included time-varying characteristics are the average local wage income (from the IRS), a quality-adjusted house price index,[82] and indices of local consumption amenities (constructed as principal components of the density of establishments in leisure and consumption industries).[83] In addition, I also include time-invariant measures of the *level* of the cultural and natural amenities discussed in the previous section, and population density.

Column 1 of Table A6 shows the effect of these characteristics on the city attractiveness series obtained from the 1990-2017 aggregate migration sample. Column 2 adds a city fixed effect, which absorbs the effect of any unobserved time-invariant city characteristics. The results show that greater overall migration to a city destination is significantly correlated with higher average wages in the city. Moreover, I also find significantly positive effects of a warmer climate, and there is weak evidence that greater density, college shares and water access are associated with greater aggregate migration. Once constant unobservable city attractiveness is accounted for, house prices show the expected negative sign in Column 2. Moreover, amenities in the form of consumption establishments have a significant correlation with migration in both specifications.

These descriptive findings inform the empirical approach in later sections. In particular, this analysis suggests that one needs to be wary of bias arising from time-varying changes in amenities, which is why I design a network IV approach that generates exogenous variation in house prices that is plausibly orthogonal with regard to local amenity changes.

Heterogeneity in correlates of city attractiveness. In columns 3 to 6, I analyse the determinants of city attractiveness separately for the city-year fixed effects obtained from migration flows by education group. The estimates in columns 3 and 5 show that the main differences between the two education groups are that only college-educated workers are significantly more likely to migrate into denser cities, while the college share has a weak negative effect on migration of non-college workers. In this descriptive analysis, I only find small differences in wage and house price correlations with migration by education group. [84]

E.2 Case study: Superstar cities and migration spillovers

What do migration links look like among large cities in the U.S – and how does that affect house price dynamics? In this section, I show an example of how identifying the migration networks of a particular set of cities with particularly strong economic booms allows us to predict which other cities will experience population growth. I focus on a small set of "superstar" cities that were shown by Gyourko, Mayer, and Sinai (2013) to have had historically particularly inelastic housing supply and high housing demand, which

[82] I adjust the level of house prices for the quality of the housing stock in the year 2000, following Albouy and Ehrlich (2018), and then use the FHFA repeat-sales index to adjust house prices for changes over time. See Appendix D.6 for details.

[83] See Appendix D.4 for details

[84] Note that the full sample estimates also differ in the sample period, in addition to varying the level of aggregation.

Table A6: City attractiveness determinants

Sample:	1990-2017		2005-2017			
	All	All	Coll.	Coll.	Non-Coll.	Non-Coll.
	(1)	(2)	(3)	(4)	(5)	(6)
Log Inc. per cap.	2.212***	0.701***	4.923***	1.714***	4.541***	1.739***
	(0.298)	(0.095)	(0.486)	(0.299)	(0.478)	(0.312)
House price index (qual.-adj.)	0.699***	-0.202***	0.677***	-1.142***	0.653***	-1.215***
	(0.121)	(0.050)	(0.211)	(0.156)	(0.205)	(0.157)
Amenities index (1st PC)	-0.378***	-0.011**	-0.399**	0.030	-0.379**	0.030
	(0.088)	(0.005)	(0.164)	(0.022)	(0.158)	(0.021)
Amenities index (2nd PC)	0.291***	-0.030***	0.441***	-0.072*	0.444***	-0.079**
	(0.048)	(0.010)	(0.085)	(0.039)	(0.084)	(0.039)
Water surface share	0.762*		0.786		0.831	
	(0.390)		(0.655)		(0.634)	
Nontrad. Christ. Share	0.162		0.416		0.431*	
	(0.168)		(0.265)		(0.256)	
Jan. Temperature	0.028***		0.035***		0.036***	
	(0.003)		(0.006)		(0.006)	
College share (2000)	3.406*		0.949		-5.890*	
	(1.793)		(3.169)		(3.061)	
Population density	0.436*		0.789**		0.565	
	(0.241)		(0.399)		(0.370)	
Observations	19,208	19,208	9,308	9,308	9,308	9,308
Year FE	✓	✓	✓	✓	✓	✓
City FE		✓		✓		✓

Heteroskedasticity-robust standard errors clustered at the CZ level in parentheses: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. The table shows results from regressions where the dependent variable consists of CZ × year fixed effects from a migration gravity regression. Analysis includes all continental U.S. CZs, excl. New Orleans, for which fixed effects could be computed, leading to a total of 613 - 716 CZs in a given year. See text for description of the explanatory variables.

result in high house price elasticities with regard to population growth, which means that "lower income households [are] crowded out by higher income households"(Gyourko et al., 2013).

As a result of these characteristics, superstar cities should be very likely to originate migration flows during the housing boom 2000-2007 that lead to spillovers to destination cities. Gyourko et al. (2013) identify superstar cities by noting for the respective 20 years preceding the years 1970, 1980, 1990, and 2000, whether a metro area was above-median in the sum of house price growth and housing unit growth ("high demand"), and in the top decile of the ratio of these two variables ("inelastic"). Superstars fulfill these criteria for at least two of the four periods considered. The resulting list of 21 superstar MSAs corresponds to 14 different 1990 commuting zones, which I [illegible] in my analysis.[85]

Identifying migration spillover cities. In order to identify cities that are likely to be impacted by migration spillovers from superstar cities, I proceed as follows: First, I restrict the search to large cities[86] Second, for each superstar city, I compute the average annual share of all IRS-reported migration outflows going to each other city for the years 1990-2000. Then, for each superstar city, I retain the Top 4 large migration destinations receiving the highest 1990-2000 outflows. Any of the retained top destinations that are not superstar cities themselves, I add to the list of "superstar spillover cities", of which I identify 12.[87] Note that both the superstar and the spillover cities were defined solely based on data up to the year 2000.

The housing boom in superstar and spillover cities. To understand how the superstar and spillover cities defined *ex ante* from pre-2000 data fared subsequently during the housing boom, Figure A12 plots average house price growth and population growth during 2000-2007 for all large CZs. The graph shows that the superstar characteristics observed by Gyourko et al. (2013) in pre-2000 data are highly persistent: all the superstar cities are in the upper left corner of the graph, which means they continue to have high house price growth and a high ratio of house price growth to population growth. The "spillover" cities are characterized by high house price growth and higher population growth than the superstar cities during the boom.[88]

The graph also highlights that it is misleading to speak of "the" U.S. housing boom of the early 2000s: While house prices experienced rapid growth on average, there is substantial cross-sectional variation, with more than a third of large cities experiencing nominal house price increases of less than 5% per year. Conversely, there are clear "boom" cities with average house price growth above 10% per year, but they almost exclusively consist of superstar and migration spillover cities: of 17 "boom" CZs, 6 are superstar cities, 10 are spillover cities, and only one is neither.[89]

---

[85]These 14 superstar CZs, identified by their largest city, and the CZ code in parentheses, are: Albany, NY (18600); Newark, NJ (19600); Boston, MA 20500); Hudson Valley, NY (19300); Los Angeles, CA (38300); NYC (19400); New Haven, CT (20901); San Francisco, CA (37800); Philadelphia, PA (19700); Santa Barbara, CA (38200); Springfield, MA (20800); Pittsfield, MA (20902); Providence, RI (20401); San Jose, CA (37500).

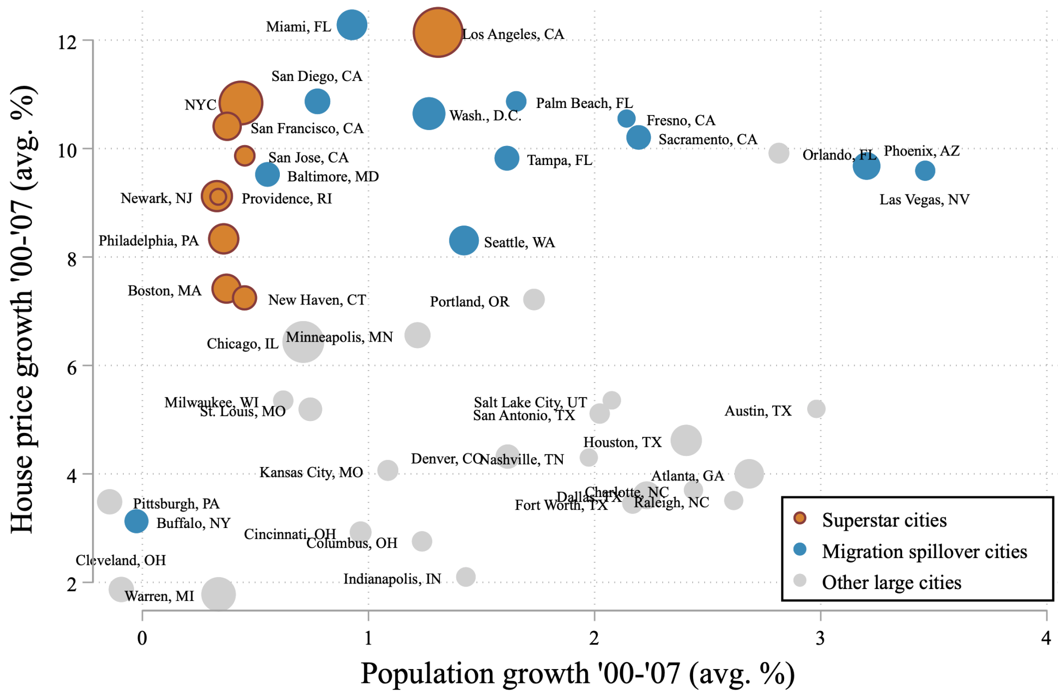
[86]Here, "large" cities are defined as having more than 0.85 M adults over the age of 21 in the year 2000 census, a cutoff chosen to ensure that they represent slightly more than 50% of U.S. adults. Of all continental U.S. CZs, 45 are "large". I exclude New Orleans as it represents a large outlier in negative population growth due to the impact of Hurricane Katrina.

[87]These "spillover cities", identified by their largest city, and the CZ code in parentheses, are: Baltimore, MD (11302); Tampa, FL (6700 Miami, FL (7000); Palm Beach, FL (7100); Wash., D.C. (11304 Buffalo, NY (18000); Phoenix, AZ (35001 Fresno, CA (37200); Sacramento, CA (37400); Las Vegas, NV (37901 San Diego, CA (38000); Seattle, WA (39400)

[88]With the sole exception of Buffalo, NY, among the spillover cities, and Los Angeles among the superstars, the two groups can be neatly separated, with the cluster of spillover cities all having higher population growth as a result of migration.

[89]That one city is Orlando, FL. Also note that, among the spillover cities, there are two outliers, which have much higher population growth (> 3% per year) than other spillover cities, but are in the lower half of house price growth in this group: Las Vegas, NV, and Phoenix, AZ. Tthese two cities are unusual in how *little* house price growth they saw during the housing boom compared to their peers. While this pair of cities is sometimes highlighted as "anomalies" (Glaeser, 2013) for experiencing high house price growth during the boom, this analysis shows that an unusually large migration shock combined with their above-median supply constraints (according to Saiz (2010)) can perhaps explain their relatively large house price response. In fact, other spillover cities, such as Washington, D.C., or Baltimore, MD, which experienced similar house

Figure A12: Population growth and house price growth. The graph plots average population growth and house price growth over 2000-2007, Data shown contains 45 CZs, consisting of all continental U.S. CZs, excl. New Orleans, with year 2000 Census population $> 0.85M$, and corresponding to $\sim 50\%$ of the continental U.S. adult population.



Overall, being a likely migration spillover destination of workers displaced from superstar cities seems to predict high population and house price growth during the housing boom, showcasing the house price propagation mechanism through migraton networks that I propose in this paper.

E.3 Migration links as predictors of inter-city house price correlation: horserace regressions

In Section 2.3, I built on the analysis in Sinai and Souleles (2013) and showed that migration links perform well as predictors for bilateral correlations in house price growth between cities when compared to other measures individually. In this section, I provide further details on the alternative link measures and show that migration links are strong predictors of house price correlation even when controlling for *all* the alternative links jointly in horserace regressions.

To put the ability of migration links to predict house price correlations in context, I consider a number of alternative measures of inter-city links: First, I include an inverse-distance weighted measure that represents the notion that house price correlations might stem from shocks that are common among cities that are geographically close to one another.

Second, I consider a social connectedness index (SCI) based on Facebook friendship links between geographic areas that was introduced in Bailey, Cao, Kuchler, Stroebel, and Wong (2018b). In a related paper, Bailey, Cao, Kuchler, and Stroebel (2018a) showed that differences in individual exposure through online social networks to house price movements in distant counties can predict differences in housing investment

price increases with much smaller demand growth and similar supply constraints, should be considered much more "anomalous" during the 2000-2007 boom than Las Vegas or Phoenix.

decisions. To measure the importance of this alternative channel, I use weights based on the SCI measure to construct a social network-weighted measure of house price correlations.[90]

Third, I construct a destination population-weighted measure of house price correlations to control for the possibility that the migration weights are simply picking up the fact that large cities have more migration links and might be driving the housing cycle of smaller cities.

Fourth, I explore the possibility that house price correlations reflect similarity in industry structure between cities. To measure industry structure differences, I compute the vector distance in 2-digit NAICS industry employment shares by city, and use the inverse of this distance to measure similarity in industry structures.

Fifth, I include an equal-weighted measure that simply reflects a city's average correlation with other cities' house prices.

We can test the predictive ability of these different city links more formally by estimating the regression model

$$corr(\Delta \ln P_i, \Delta \ln P_k) = \alpha + \beta_1 MigShare_{i \to k} + \beta_2 \ln dist_{ik} + \beta_3 SCI_{ik} + \beta_4 IndDist_{ik} \theta_i + \theta_k + \epsilon_{ik},$$

which represents a horserace between migration links, distance, social networks, and industry structure in explaining variation in house price correlations across CZs.

The results are shown in Appendix Table A7. Going from left to right, the columns add in alternative measures of inter-city links as well as origin and destination city fixed effects that capture the overall tendency of cities to be more connected (e.g. due to city size). As the table shows, migration flows have a significant positive association with the correlation in house prices. Moreover, this relationship is robust to controlling for spatial correlation due to geographic proximity as well as variation associated with social network links and industry similarity. Panel A shows the results for the full migration network whereas Panel B limits the sample to those city pairs which are at least 150 miles apart, which corresponds to the long-distance migration network used in my baseline regressions.

Even in the most stringent specifications in column 4, migration links continue to be significant predictors of house price correlations. That is, migration links contain information about city house price connections that go beyond the set of links represented by the other links included in the regressions. This predictive performance suggests migration links can *not* simply be considered as proxies for one of these other city connections.

The estimated (non-causal) coefficient in Column 4 for the long-distance network indicates that a 10 ppt greater share of migrants from city i going to city k is associated with a 6.7 ppt higher correlation in house price growth between the cities, holding constant the city's co-movement with house prices overall.

Trade flows and migration flows. In the graphs showing the individual power of different measures of links in predicting pair-wise house price correlations, I also show results using a measure of trade flow links between cities that represents industry linkages and the propagation of economic shocks through input-output networks (see Appendix Section D.5 for details on how this measure is constructed). However, this measure is only available for a limited subsample corresponding to less than 10% of the pairwise links and is therefore not included in the regressions. In unreported regressions, I find that, on that limited sample, migration links are significant in predicting house price correlations when controlling for trade flow links.

In order to further explore the degree to which migration links overlap with trade flows, consider Appendix Figure A13, where I show how gross migration flows and the value of trade flows are correlated for pairs of Commodity Flow Statistics areas (which are on average somewhat larger than CZs). This data is available for all connections between 70 CFS areas, (although the graphs drop any pairs of locations that have no migration flows or no trade flows).

As the graphs show, while there is a relatively high R-squared of 34% between migration flows and the value of trade flows between two locations in the raw data, this correlation is to a large degree driven by the fact that large cities are more likely to trade and have migration flows with *any* other city. Once I adjust for the joint population size in each location pair, the R-squared is only 12%. In conjunction with the fact

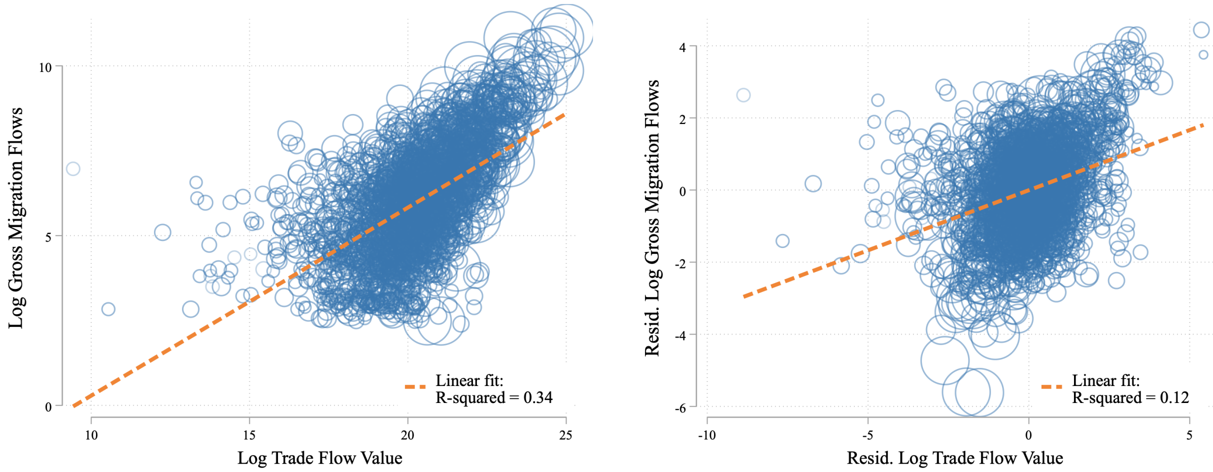
[90] The SCI index measures the relative probabilities of friendship links between counties (normalized for their respective Facebook user base) – which I aggregate to the commuting zone level, and then normalize for each CZ such that the weights for all other CZs sum to one.

Table A7: City links and house price growth correlation

Dependent variable:	House price growth correlation coeff. × 100			
	(1)	(2)	(3)	(4)
Panel A: Full migration network				
Migration outflow share	83.276***	69.339***	8.165**	10.378***
	(5.415)	(4.372)	(3.597)	(3.568)
Log distance (miles)			-5.846***	-5.405***
			(0.216)	(0.211)
Social Connectness Index			10.327	10.145
			(7.347)	(7.268)
Industry structure similarity				22.573***
				(2.446)
Observations	516,241	516,241	516,241	516,241
R-squared	0.00	0.65	0.67	0.68
Panel B: Distance $>$ 150 **mi.**				
Migration outflow share	235.132***	150.743***	72.698**	66.976**
	(80.653)	(50.749)	(29.852)	(27.499)
Log distance (miles)			-4.812***	-4.310***
			(0.242)	(0.243)
Social Connectness Index			611.968***	602.176***
			(69.872)	(70.143)
Industry structure similarity				22.413***
				(2.446)
Observations	495,554	495,554	495,554	495,554
R-squared	0.00	0.66	0.67	0.68
Origin FE		✓	✓	✓
Destination FE		✓	✓	✓

Heteroskedasticity-robust standard errors two-way clustered at the origin and destination CZ level in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Correlations in house prices growth and average outmigration shares are calculated over 1990-2017. Data contains pairs of all 721 continental U.S. CZs, excl. New Orleans, that have any known outmigration flows.

Figure A13: Trade flows and migration flows by location pair. Panel (a) shows the value of 2012 log gross migration flows (sum of flows in both directions) over the log gross value of 2012 trade flows (in both directions) for each pair of CFS areas. Panel (b) shows the same variables but residualized with regard to the log sum of populations of the location pair in the year 2000 as a proxy for city sizes. The size of location pair markers is shown proportional to the joint population size of the pair, and the graphs also show the unweighted line of best fit.



(a) Raw flows

(b) Resid. w.r.t. population size

that – as shown in Figure 5 – migration links are a better predictor of house price correlations than trade flows, this makes it unlikely that migration links are merely a proxy for trade links in driving the observed house price spillover effects.

E.4 Reasons for moving and demographics of movers

In order to understand better which demographics are driving migration patterns in the U.S., this section will use data from the Current Population Survey (CPS) to explore the characteristics of U.S. migrants.

Reasons for moving. First, I consider the reasons stated by survey respondents who moved in the last year when asked why they were moving. I group the CPS response categories into "family reasons" (*change in marital status; establishing own household; other family reasons*), "employment reasons" (*new job or job transfer; to look for work or lost job; to be closer to work/easier commute; other job-related reason*), "retirement", and "housing reasons" (*wanted to own home, not rent; wanted new or better house/ apartment; wanted better neighborhood/less crime; wanted cheaper housing; foreclosure/eviction; other housing reason*). I omit move reasons due to college attendance, climate, health, or disaster, which the CPS groups under "other reasons". For each category of reasons, I compute the share of the total population with available moving status that moved for that reason - distinguishing between all moves and moves across county boundaries.

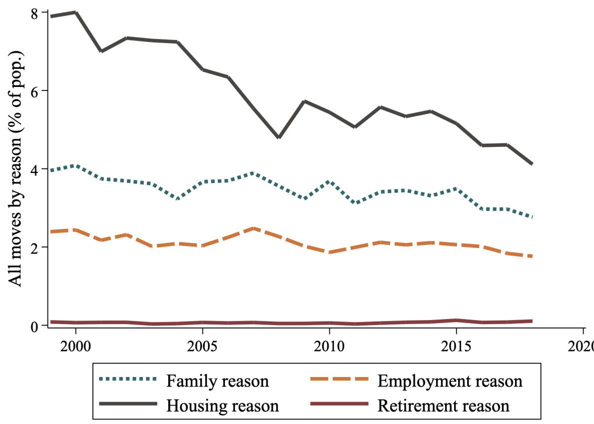
The patterns for moving reasons are shown in Appendix Figure A14, in Panels (a) and (b). As the time series show, housing reasons are the most important reason for moving by far when I include within-county moves in the tabulation in Panel (a), followed by family reasons and employment reasons. When I consider only "long-distance" moves that cross county boundaries, housing, employment and family reasons are of similar importance. In either category of moves, moves due to retirement play a negligible role in mobility. The time pattern of migration being correlated with housing booms and busts is reflected in the fact that moves for any reason are higher in the early 2000s and 2010s than in the late 2000s. However, it is important to note that this is mostly driven by the cyclicality of moves for housing reasons. This provides support for the idea that migration cycles are an important consequence and component of housing cycles rather than being driven by changes in employment opportunities that coincide with housing booms.

It is important to note the limitations of this analysis: in a spatial equilibrium setting with Rosen-Roback style preferences, like the model in this paper, residents jointly consider the effect of employment, housing, family and other amenities on their utility in deciding whether or not to change locations – so there is no clear sense in which either of these elements is *the* cause of their move. However, if we assume that respondents take the survey to be asking about which of these elements had been changing the most to occasion their change in location preference, then the prominent role of housing provides qualitative evidence of the proposed migration spillover mechanism where house price changes lead to migration in search of more affordable housing during housing booms.

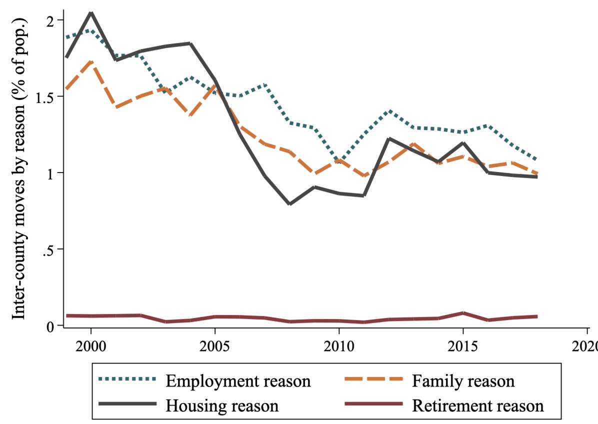
Employment status of long-distance movers. Next, I consider the employment status of movers. In line with the paper's focus on moves across cities, I focus on inter-county movers. I retain all CPS respondents who moved across county lines and have an employment status (which drops children, for instance), and are not in the armed forces. Then, I compute the number of movers in 3 non-overlapping and exhaustive categories: (1) Employed (both at work currently and not currently at work); (2) Unemployed, or not in the labor force ("NILF"), but not retired; (3) NILF and retired. The share of inter-county movers in each employment category is shown in Panel (c) of Appendix Figure A14. Around 60% of all movers are employed (when being surveyed after their move), less than 10% are retired, and the remainder are unemployed or out of the labor force but not retired. This aggregate pattern is particularly important when evaluating anecdotal evidence about particular city pairs with strong migration links. For example, it is possible that a greater number of movers from New York City to Florida are retirees than for other city pairs, but given that retirees represent a very small share of migration overall, it is unlikely that they represent the majority share of moves even for this city pair. More generally, it is unlikely that retirement location preferences play a major role in explaining the migration spillover patterns documented in this paper.

Age structure of long-distance movers. I next consider the age structure of long-distance movers, grouping CPS respondents by age, and omitting those less than 20 years old. The age shares of movers are shown in Panel (d) of Appendix Figure A14. The time series show that mobility declines precipitously with age: while the population share of those aged 40 years and older is larger than that of those under 40, their combined share of long-distance migration is less than half that of the younger group. While there is a gentle upward slope in the migration share of older groups, this can likely be explained by their increasing population share over this time period. As a result, when we are thinking about migration patterns, they are likely to be driven predominantly by the decisions of the younger working-age population.

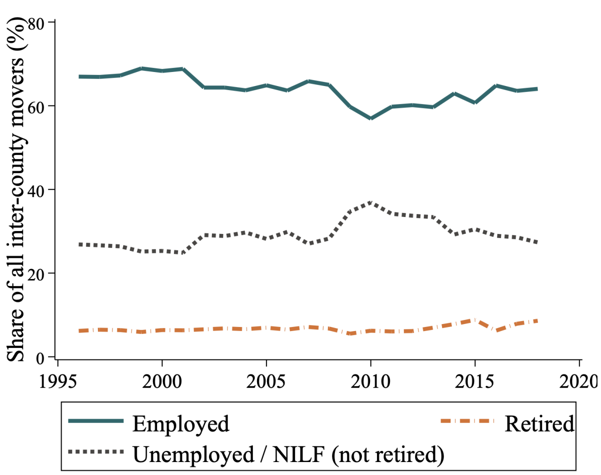
Figure A14: Reason for moving and demographics of movers (CPS). The data shown in the graphs below comes from the Current Population Survey March Supplement. Panel (a) shows the share of the U.S. population moving for the stated reason, omitting the category of "other reasons". Panel (b) considers only "long-distance" moves that cross county boundaries. Panel (c) shows the share of inter-county movers by employment status, distinguishing between the employed, retirees and the remainder of workers who are unemployed or not in the labor force. The analysis excludes workers in the armed force or where employment status is not reported. Panel (d) groups respondents by their age and plots shares of total inter-county migration for each age group.



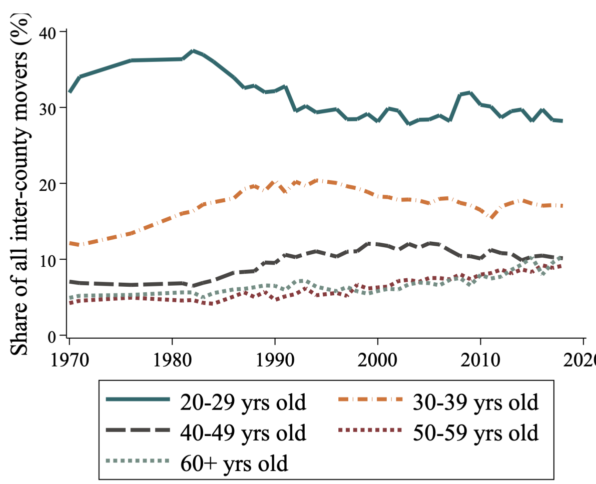
(a) Reason for moving: all movers



(b) Reason for moving: inter-county movers



(c) Inter-county mig. by employment status



(d) Inter-county mig. by age

Appendix References

Acemoglu, D., U. Akcigit, and W. Kerr (2016): "Networks and the macroeconomy: An empirical exploration," *NBER Macroeconomics Annual*, 30, 273–335.

Albouy, D. and G. Ehrlich (2018): "Housing productivity and the social cost of land-use restrictions," *Journal of Urban Economics*, 107, 101–120.

Almagro, M. and T. Domínguez-Iino (2020): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," .

Altonji, J. G. and D. Card (1989): "The effects of immigration on the labor market outcomes of natives," Tech. rep., National Bureau of Economic Research.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish," *The Review of Economic Studies*, 67, 499–527.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel (2018a): "The economic effects of social networks: Evidence from the housing market," *Journal of Political Economy*, 126, 2224–2276.

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

Bartik, T. J. (1991): "Boon or Boondoggle? The debate over state and local economic development policies," .

Belloni, A., V. Chernozhukov, et al. (2013): "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 19, 521–547.

Borusyak, K., P. Hull, and X. Jaravel (2020): "Quasi-experimental shift-share research designs," Tech. rep., National Bureau of Economic Research.

Boustan, L. P. (2010): "Was postwar suburbanization white flight? Evidence from the black migration," *The Quarterly Journal of Economics*, 125, 417–443.

Bureau, U. C. (2005–2017): "PUMS accuracy of the data," .

Davidoff, T. et al. (2016): "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated With Many Demand Factors," *Critical Finance Review*, 5, 177–206.

Derenoncourt, E. (2019): "Can you move to opportunity? Evidence from the Great Migration," .

Diamond, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *The American Economic Review*, 106, 479–524.

Dingel, J. and F. Tintelnot (2020): "Spatial Economics for Granular Settings," Tech. rep., National Bureau of Economic Research.

DuMouchel, W. H. and J. E. Harris (1983): "Bayes methods for combining the results of cancer studies in humans and other species," *Journal of the American Statistical Association*, 78, 293–308.

Eckert, F., T. C. Fort, P. K. Schott, and N. J. Yang (2020): "Imputing Missing Values in the US Census Bureau's County Business Patterns," Tech. rep., National Bureau of Economic Research.

Eckert, F., S. Ganapati, and C. Walsh (2019): "Skilled tradable services: The transformation of US high-skill labor markets," *Available at SSRN 3439118*.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013): *Bayesian data analysis*, CRC press.

Glaeser, E. L. (2013): "A nation of gamblers: Real estate speculation and American history," *American Economic Review*, 103, 1–42.

GLAESER, E. L., H. KIM, AND M. LUCA (2018): "Nowcasting gentrification: using yelp data to quantify neighborhood change," in *AEA Papers and Proceedings*, vol. 108, 77–82.

GOLDSMITH-PINKHAM, P., I. SORKIN, AND H. SWIFT (2018): "Bartik instruments: What, when, why, and how," Tech. rep., National Bureau of Economic Research.

GYOURKO, J., C. MAYER, AND T. SINAI (2013): "Superstar cities," *American Economic Journal: Economic Policy*, 5, 167–99.

HOWARD, G. (2020): "The migration accelerator: Labor mobility, housing, and demand," *American Economic Journal: Macroeconomics*, 12, 147–79.

JAEGER, D. A., J. RUIST, AND J. STUHLER (2018): "Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration," Tech. rep., Working Paper.

JORDÀ, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American economic review*, 95, 161–182.

LUTZ, C. AND B. SAND (2019): "Highly disaggregated land unavailability," *Available at SSRN 3478900*.

MOLLOY, R. AND C. SMITH (2019): "US Internal Migration: Recent Patterns and Outstanding Puzzles," in *Prepared for the Federal Reserve Bank of Boston's 63rd Economic Conference."A House Divided: Geographical Disparities in Twenty-First Century America," October*, 4–5.

MORRIS, C. N. (1983): "Parametric empirical Bayes inference: theory and applications," *Journal of the American statistical Association*, 78, 47–55.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "IPUMS USA: Version 10.0," .

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

SINAI, T. AND N. SOULELES (2013): "Can owning a home hedge the risk of moving?" *American Economic Journal: Economic Policy*, 5, 282–312.

STOCK, J. H. AND M. W. WATSON (2018): "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *The Economic Journal*, 128, 917–948.