Author's personal copy

Journal of Urban Economics 70 (2011) 1–14

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Demand substitution across US cities: Observable similarity and home price correlation

DeForest McDuff *

Princeton University & Quant Economics, Inc., Industrial Relations Section, Firestone Library, Princeton University, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:
Received 5 May 2009
Revised 15 February 2011
Available online 5 March 2011

Jel classification:
R

Keywords:
Housing demand
Substitution
Migration
Home price correlation

ABSTRACT

This paper studies demand substitution in the context of US cities. Demand substitution occurs when individuals on the margin between certain city pairs affect demand patterns in the aggregate, causing certain cities to be better substitutes than others. Using a discrete model of city choice, I derive two predictions for migration flows and test them empirically using city-to-city migration data from the US Census. I show that cities which are similar on a variety of observable measures have higher levels of gross migration flows in the steady state and higher net migration flows in response to labor demand shocks. Finally, I propose pairwise correlation in metropolitan home prices as a price-based measure of substitutability and show that it contains substantial predictive power for migration flows relative to observable similarity.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Traditional urban spatial equilibrium models require that identical households and firms be indifferent across locations.[1] Implicit in these models is the assumption that households sort into locations based on varying preferences for city amenities, much in the same way that individuals choose employment in the face of compensating wage differentials based on varying preferences for job attributes. Thus far, little attention has been devoted to studying demand substitution across locations and its importance for understanding intercity migration.

This paper studies demand substitution across US cities. I utilize methods from discrete choice literature to model intercity equilibrium in the context of varying individual preferences for city amenities. From this equilibrium, I derive two key predictions for migration data and test candidate measures of substitutability empirically. Understanding how and why cities are connected in demand helps to explain important features of housing markets not accounted for traditional urban equilibrium models.

Table 1 and Fig. 1 present an example on which an empirical understanding of demand substitution sheds some light. Table 1 displays observable similarity between Allentown-Bethlehem, PA and four nearby cities: Scranton, Harrisburg, Philadelphia, and New York. Measured by population, industrial composition, and housing supply elasticity, Allentown is more similar to Scranton and Harrisburg than Philadelphia and New York. Yet Fig. 1 illustrates how Allentown's housing market more closely resembles Philadelphia's and New York's. Indeed, Allentown attracts a far greater percentage of outward migration from New York and Philadelphia than do Scranton and Harrisburg.[2] In principle, these correlated price movements could reflect supply-side factors, but more plausibly these markets are connected in demand. Knowledge of how and which markets are connected in demand helps to explain why migration flows occur between them and how local economic shocks get propagated through the economy. In this example, the high home price correlation between Allentown and Philadelphia/New York is predictive of the high willingness of many residents to substitute between them.[3]

Studies of housing markets that lack a framework for demand substitution fail to explain certain empirical patterns in housing data. For example, research on housing supply elasticity provides

* Fax: +1 858 876 9102.
E-mail addresses: rmcduff@Princeton.edu, RDM@QuantEconomics.com

[1] See Rosen (1979), Roback (1982), Blomquist et al. (1988), Greenwood et al. (1991), Kahn (1995), Gabriel et al. (2003), Gabriel and Rosenthal (2004), Chen and Rosenthal (2008).

[2] The 2000 US Census indicates that from 1995–2000 Allentown attracted roughly 0.59% and 0.16% of Philadelphia and New York residents, respectively, compared to only 0.33% and 0.06% for Harrisburg and 0.18% and 0.04% for Scranton.

[3] Home price correlation would have also predicted that Allentown is a better substitute for Philadelphia than New York and for Scranton than Harrisburg. The 2000 Census also confirms these substitution patterns since Allentown attracted more residents from Philadelphia than New York (0.59% vs 0.16%) and from Scranton than Harrisburg (1.53% vs 0.71%).

0094-1190/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jue.2011.02.004

Author's personal copy

Table 1
Similarity example: Allentown-Bethlehem, Pennsylvania.

	Allentown	Scranton	Harrisburg	Philadelphia	New York	Mean	SD
Miles to Allentown	0	77	84	64	95	999	747
2000 Population (1000s)	638	625	629	6,188	21,200	819	1958
Housing supply elasticity	1.54	1.34	1.27	1.10	0.64	1.59	0.85
Manufacturing	21.2%	18.2%	13.3%	12.7%	10.7%	18.7%	3.6%
Business services	10.9%	9.9%	12.4%	14.2%	16.5%	11.4%	2.8%
HPC with Allentown	1.000	0.632	0.357	0.857	0.631	0.187	0.350

Notes: The table above displays city similarity and home price correlation (HPC) for Allentown-Bethlehem, PA and four nearby cities. Population and industrial shares come from the 2000 US Census, housing supply elasticity comes from Saiz (2010), and home price correlation comes from the pairwise correlation of annual log returns on the OFHEO metropolitan-level home price indices. Allentown-Bethlehem is much more similar in observables to Scranton and Harrisburg compared to Philadelphia and New York, but the home price correlation measure indicates that Allentown-Bethlehem's housing market has a stronger relationship with the latter, despite the other observable differences. Means and standard deviations are reported for all 297 cities in the sample.

explanations for the occurrence of price fluctuations and long-run differences in housing growth,[4] but do not explain why price fluctuations tend to correlated across some markets but not others. Research on macroeconomic housing cycles recognize changing demand over time,[5] but do not explain the sizeable variation in housing markets across cities.[6]

The role of demand substitution is similarly underrepresented in the intercity migration literature. This research has at its core the notion that migration occurs in response to spatial disequilibrium generated by changing population preferences for amenities over time or changing individual preferences for amenities over the course of the life cycle.[7] Yet these papers rarely examine the channels over which migration occurs to achieve equilibrium. That is, spatial disequilibrium in one market causes migration to or from several substitute markets, but which ones? This is the type of question on which demand substitution provides some understanding.

Demand substitution occurs when individuals on the margin between certain city pairs affect demand patterns in the aggregate. Two cities are substitutes in demand if price changes in one city affect demand for the other city. This often occurs when cities are similar in amenities or characteristics that people care about. For example, an individual may be tempted to move from Los Angeles to Phoenix but not to Seattle if warm weather is an amenity that he or she values. Using a discrete model of city choice, I show that varying individual preferences for city amenities causes certain pairs of cities to be better substitutes than others.

I derive two empirical tests that use migration data to validate candidate measures of city demand substitutes. Migration data reflects city substitution by definition since individuals who move from city j to city k observably substitute one city for the other. I show that gross migration flows in the steady state and net migration flows in response to labor demand shocks should be higher between city pairs that are better substitutes. For the latter, I use the methodology of Bartik (1991) to generate exogenous variation in labor demand across cities based on national industry growth rates and industry shares in each city.

In addition to considering measures of observable city similarity, I propose home price correlation as a priced-based measure of city substitution.[8] Home price correlation is defined as the pairwise correlation in price changes between two cities over time. Home price correlation may be informative for identifying and understanding demand substitution since demand shocks show up in prices when housing is inelastic in the short-run. To the extent that shocks to any one city spill over via demand substitution to other cities, substitutability between those cities will show up in home price correlation.

Regressions on gross and net migration flows from the US Census suggest several observable measures that are associated with substitute city pairs, including average temperature, political affiliation, home values, and per capita income. Relative to observable measures of city similarity, home price correlation is particularly informative for predicting migration flows. This finding is robust to a variety of empirical specifications.

In summary, this paper makes three primary contributions related to understanding demand substitution across cities. First, I define demand substitution in the context of spatial equilibrium across cities. Second, I evaluate the occurrence of city substitution by deriving and estimating empirical substitution predictions for US Census migration data. Third, I propose home price correlation as a measure of demand substitution and evaluate its predictive power relative to known measures of observable similarity.

The remainder of the paper proceeds as follows. Section 2 discusses the methodology. Section 3 describes the data. Section 4 presents the results. Section 5 concludes.

2. Methodology

2.1. Substitute cities

The classical definition of substitute goods is that the cross-price elasticity of demand of one good with respect to the price of the other is positive. In a city context, this means that the de-

[4] See Glaeser et al. (2008), Gyourko et al. (2006), and Saiz (2010) for discussion and references.

[5] See Brunnermeier and Julliard (2008), Himmelberg et al. (2005), Mayer and Sinai (2007), and Shiller (2007).

[6] For example, the most recent home price boom from 1997–2006 was concentrated in Florida, California, the Northeast, and the Mid-Atlantic. Some parts of the Northwest, Southwest, and the coastal South experienced similar booms, but price growth in these regions was far less uniform across cities. Finally, Texas, the Southwest, the Midwest, and the Rocky Mountain region had little, if any participation in the price boom. In fact, the mean pairwise home price correlation for cities studied in this paper is only 0.18.

[7] See Mueser and Graves (1995), Costa and Kahn (2000), Graves and Knapp (1988), Graves and Waldman (1991), Rappaport (2004), Rappaport (2007), Walker (2006), Bishop (2007), and Chen and Rosenthal (2008).

[8] Financial economists have a long tradition of studying asset price correlation, dating back to original CAPM models by Sharpe (1963) and Lintner (1965). In light of recent volatility in US home prices, much attention is being given to uncertainty and risk structure of home prices, including papers by Davidoff (2006), Englund et al. (2002), Flavin and Yamashita (2002), Leung (2007), and Sinai and Souleles (2005). Fewer studies, however, identify what factors drive home price correlation and what can be learned from. McDuff (2010) is one exception that examines home price correlation in local housing markets in terms of geography, price, and home type in an attempt to identify local markets which tend to co-move in price. To my knowledge, only one other paper studies home price correlation between city pairs. Sinai and Souleles (2009) observe that individuals frequently move between city pairs with highly correlated home values. However, they take no real position on causality, emphasizing not only that identification is difficult but that causality possibly runs in both directions. In addition to these points, I attempt to provide evidence that home price correlation reflects city demand substitution above and beyond correlated economic shocks via an empirical strategy with uncorrelated, exogenous demand shocks. One other notable differences is that Sinai and Souleles (2009) utilize covariance in a risk-hedging model, whereas correlation is more appropriate for the purposes here because it abstracts from the amplitudes of price changes.

Author's personal copy

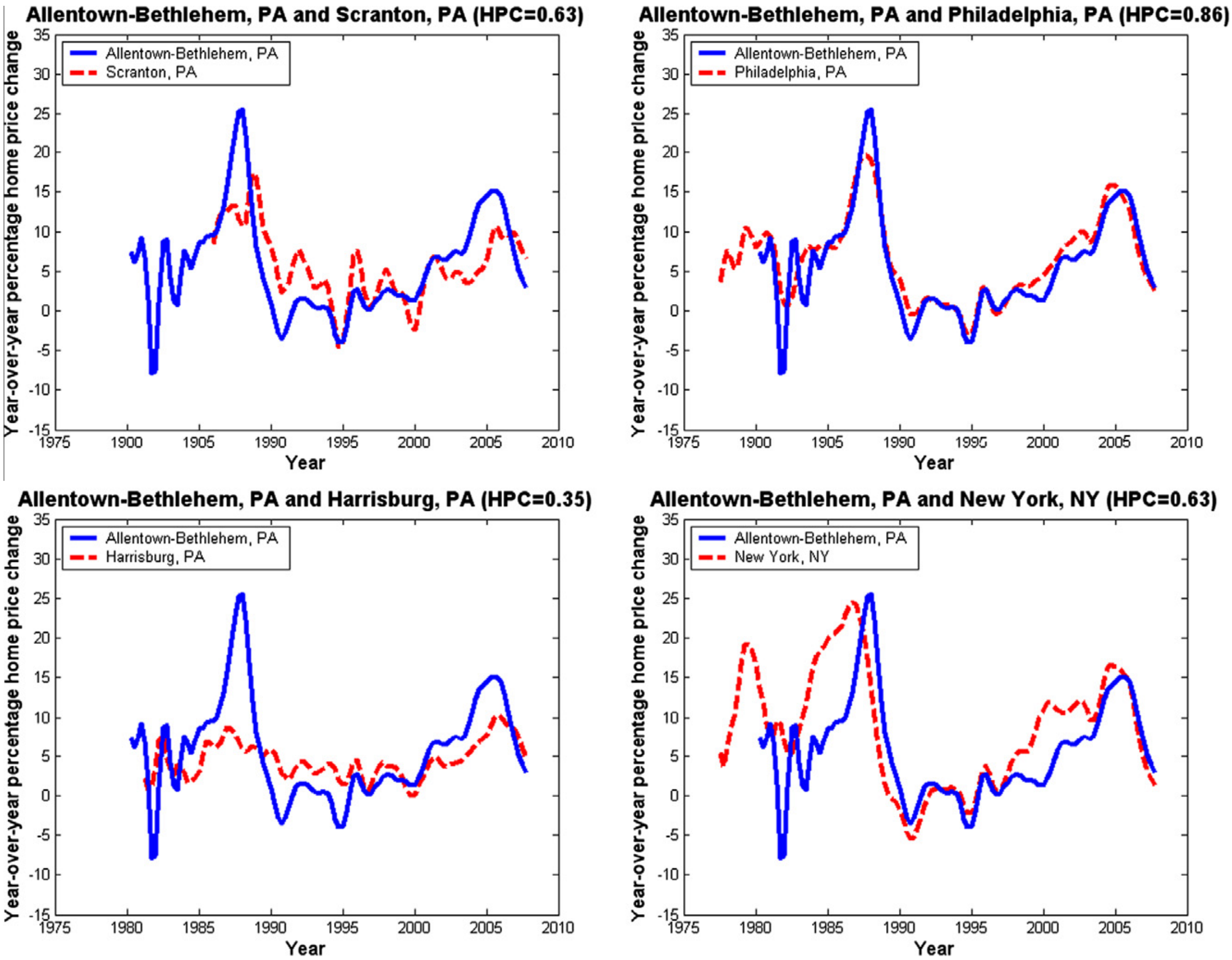

Fig. 1. The figure above plots year-over-year home price appreciation for four cities compared to Allentown-Bethlehem, PA from 1975–2008. Home price appreciation for each city is calculated using the OFHEO metropolitan-level home price index for each city. Despite the fact that Allentown-Bethlehem is more similar to Harrisburg and Scranton on observable measures of city similarity, its housing market more closely resembles Philadelphia's and New York's.

mand for city j is increasing in the price of living in city k. That is, city j is a substitute for city k if the cross-price elasticity of demand is positive:

$$\varepsilon_{jk}^{D} = \frac{p_k}{D_j}\frac{\partial D_j}{\partial p_k} > 0, \tag{1}$$

where D_j represents the aggregate demand for city j and p_k represents the price of housing in city k. City pairs with higher cross-price elasticities of demand are better substitutes.[9]

Aggregate demand for cities is formulated by the sum of individuals' demand for various cities.[10] Understanding variation in substitution patterns across cities requires individuals with varying preferences for amenities. In this regard, I define individual utility using a random coefficients logit model as in Nevo (2000).[11]

Traditional urban economic models originating from Rosen (1979) and Roback (1982) typically assume identical workers with indirect utility over wages, rents, and amenities, $\upsilon(w,r;a)$, and identical firms with profit functions over a similar domain, $\pi(w,r;a)$. Amenities are fixed across cities, and spatial equilibrium occurs when wages and rents adjust so as to make consumers and firms indifferent across locations.[12]

I make several simplifying assumptions to connect the random coefficient model with the spatial equilibrium literature. First, I assume linearly separable utility over city amenities with varying preferences across individuals. Second, as in Glaeser and Gyourko (2007), I assume a fixed quantity of housing consumption so that wages and rents are additive. Third, I assume a constant linear relationship between rents and capitalized prices, which in theory varies according to adjustments in interest rates, taxes, etc. Thus, I define utility for agent i living in city j over wages (w_j), housing prices (p_j), and amenities (a_j) by:

9 Unfortunately, this definition is imperfect with respect to its transitivity across cities of different sizes. Under this definition, a small city could be a close substitute for a large city if only a small fraction of the large city's residents consider it so, whereas the opposite is not true. Price changes in the small city are unlikely to ever have a measurable impact on demand for the large city. Nevertheless, I move forward this definition but am careful to control for relative city sizes in the empirical portion of the paper.

10 While I focus here on issues relating to consumption demand, I note that investment demand for housing also may influence the degree to which cities are substitutes. To the extent that investors identify substitutable housing assets across cities, demand substitution will be strengthened. This is particularly evidence in the context of regional correlations in housing investment demand. Home prices appear to be correlated over time as much for investment substitution as consumption substitution. Still, the majority of this distinction is beyond the scope of this paper, yet certain aspects of demand substitution models may be informative for housing investment as well.

11 Other papers which are similar in spirit are Berry et al. (1995), Nevo (2001), Bayer et al. (2004), Bayer et al. (2007). Only the latter two have specific focuses on housing markets.

12 See Rosen (1979), Roback (1982), Blomquist et al. (1988), Kahn (1995), Gabriel and Rosenthal (2004), Chen and Rosenthal (2008).

Author's personal copy

$$u_{ij}(w_j - p_j; a_j) = \beta_i a_j + \lambda(w_j - p_j) + v_{ij}. \tag{2}$$

where β_i is a vector of agent i's preferences for various amenities and $\lambda > 0$ represents a common agent preference for non-housing consumption. The error term represents agent i's idiosyncratic preference for city j. For convenience, I assume that the error term has an extreme value distribution: $F(v_{ij}) = \exp(-\exp(-v_{ij}))$.

The extreme value error distribution produces a closed-form expression for individual city demand. Given a fixed vector of wages and home prices, the probability that agent i chooses city j is given by:

$$P_{ij} = P(Y_i = j|\beta_i) = \frac{\exp\lfloor \beta_i a_j + \lambda(w_j - p_j)\rfloor}{\sum_{k=1}^{J} \exp[\beta_i a_k + \lambda(w_k - p_k)]}. \tag{3}$$

For a given agent, the probability of living in city j increases with its wage: $\partial P_{ij}/\partial w_j = \lambda P_{ij}(1 - P_{ij}) > 0$, decreases with its price of housing: $\partial P_{ij}/\partial p_j = -\lambda P_{ij}(1 - P_{ij}) < 0$, and increases with lower wages and/or higher prices in all other cities: $-\partial P_{ij}/\partial w_k = \partial P_{ij}/\partial p_k = \lambda P_{ij} P_{ik} > 0$. In other words, agent i's likelihood of selecting city j is more sensitive to city k's wages and prices if his baseline probabilities of living in city j and city k are higher.[13]

For any set of wages and home prices, aggregate demand for each city is generated by integrating demand over the heterogeneous preference parameter distribution denoted by: $g(\beta) : D_j = \int P_{ij} g(\beta) d\beta$, with P_{ij} given in Eq. (3). D_j represents the demand for city j as a fraction of the total demand, so that $\sum_{j \in J} D_j = 1$. Firms locate across cities according to identical production functions given by $\pi(w_j, p_j; a_j)$. Equilibrium occurs when all individuals choose cities that maximize their utilities, firm profits in each city is identical, and total demand sums to one.

Differentiating D_j with respect to city k home price produces the aggregate cross-price elasticity of demand between city j and city k:

$$\begin{aligned} \varepsilon_{jk}^{D} &= \frac{p_k}{D_j}\frac{\partial D_j}{\partial p_k} = \frac{p_k}{D_j}\int \frac{\partial P_{ij}}{\partial p_k} g(\beta) d\beta = \lambda p_k \frac{\int P_{ij} P_{ik} g(\beta) d\beta}{\int P_{ij} g(\beta) d\beta} \\ &= \lambda p_k E[P_{ik}|Y_i = j] > 0. \end{aligned} \tag{4}$$

Eq. (4) provides the main intuition for this paper's hypothesis. Price changes in city k have a larger impact on the demand for city j when more individuals living in city j had high *ex ante* probabilities of living in city k, indicated by the conditional expectation as well as the multiplication of probabilities under the integral. City k is a good substitute for city j if city j residents tend to have high values of $\beta_i a_k$, which occurs when city j and city k tend to have similar amenities.

Work in the industrial organizational literature typically proceeds by assuming some distribution for as a function of individual observables and estimating the elasticities using simulation techniques with cross-sectional price and quantity data.[14] An advantage of city choice versus other product choice is that sequential cities are observed in migration data. Individuals moving from city j to city k reveal that they have relative high values for P_{ij} and P_{ik}. As Eq. (4) indicates, migration between cities is *prima facie* that city j and city k are substitutes.

[13] By construction, the model requires that all city pairs are substitutes rather than complements. However, it is interesting to consider how two cities might be complements in practice, so that the demand for city j decreases with the price of city k. Consider a situation where individuals consume city j mainly because of its proximity to city k, which is enjoyable to visit. If prices rise so much in city k that visiting is much less enjoyable, demand for city j might actually drop due to the complementarity of consumption.

[14] See Nevo (2000) for references and discussion.

2.2. Empirical predictions

If preferences were directly observable, Eq. (4) could be used to analytically determine which city pairs are substitutes in aggregate demand. In practice, only individuals' top city choices are observable, so I generate empirical predictions for migration data, where at least two sequential city choices are observed for the same individual.

This setup generates two specific empirical predictions for migration between substitute city pairs. All else being equal, city pairs which are better substitutes should have (1) higher gross migration flows in steady-state and (2) higher net migration flows in response to labor demand shocks.

The first empirical prediction is that gross migration flows will be higher between city pairs which are better substitutes in demand. Suppose that individuals draw new v_{ij}'s each year and move if their updated utilities have changed from one city to another between time periods.[15] This will generate observed migration between some city pairs even when prices and demand parameters are fixed. With i.i.d. error terms in each period and no cost of moving, one can interpret the probabilities in Eq. (3) literally so that agent i will live in city j at time t with probability P_{ij}. The time independence of the error terms dictates that agent i's joint probability of living in city j in time $t-1$ and city k in time t is $P_{ij}P_{ik}$. Thus, the expected fraction of city j residents moving to city k between times $t-1$ and t is:

$$\begin{aligned} M_{jkt} &= E[Y_{it} = k|Y_{it-1} = j] = \frac{E[Y_{it-1} = j \cap Y_{it} = k]}{E[Y_{it-1} = j]} \\ &= \frac{\int P_{ij} P_{ik} g(\beta) d\beta}{\int P_{ij} g(\beta) d\beta} = \frac{\varepsilon_{jk}^{D}}{\lambda p_k}, \end{aligned} \tag{5}$$

where the last equality comes from the expression for the cross-price elasticity of demand in Eq. (4). All else being equal, city pairs which are better substitutes as measured by a higher cross-price elasticity of demand should exhibit higher levels of gross migration flows between them.[16] Thus, the first empirical validation is to evaluate candidate measures of city substitutability based on how well they predict gross migration between city pairs.

The second empirical prediction is that net migration flows in response to labor demand shocks will be higher between city pairs that are closer substitutes in demand.[17] Suppose that an exogenous labor demand shock idiosyncratically increases the demand for labor in city k; that is, firms' production in city k, $\pi(w_k, p_k; a_k)$, is suddenly higher at every wage and price. With perfectly competitive labor markets, wages rise in the short-run, causing workers to migrate towards city k. Wages and prices adjust according as this migration occurs according to the shapes of the utility and profit functions until a new equilibrium is reached.[18]

[15] All formulas in this paper assume no cost of moving. Alternatively, one could imagine individuals moving only if utility gains from switching cities exceed some fixed cost, c. This change would decrease the overall magnitude of migration between cities but would not have any impact on the main theoretical implications. Even if the cost of moving was some increasing function of distance, the main qualitative predictions remain unchanged. In any case, the main empirical specifications control for distance between cities.

[16] Note that if the dimensionality of β is sufficiently large and $g(\beta)$ has positive support over its entire domain, there will exist at least some agents on the margin between every city pair. If not, many city pairs will have no gross migration between them.

[17] Recent research notes the propensity of migration towards economic opportunity. See Kennan and Walker (2010a, and Notowidigdo (2010) for discussion and references.

[18] See Blomquist et al. (1988), Gyourko and Tracy (1991) and Blomquist (2006). Note that the extent to which labor demand shocks clear through wages or prices affects only the magnitude of migration flows, not the relative cities from which they draw. Thus, the empirical results relating to relative migration from various cities are unaffected when compared on a relative rather than absolute basis.

Author's personal copy

Differentiating Eq. (5) with respect to w_k indicates that this wage shock will produce migration from city j to city k that increases in the cross-price elasticity of demand:

$$\frac{\partial M_{jkt}}{\partial w_{kt}} = \frac{\left(\varepsilon^D_{jk}\right)^2}{\lambda p_k^2} + \frac{\varepsilon^D_{jk}}{p_k}\left(1 - 2\frac{\int P_{ij}P_{ik}^2 g(\beta)d\beta}{\int P_{ij}P_{ik}g(\beta)d\beta}\right) > 0. \tag{6}$$

The derivation can be found in Appendix A, where it is shown that $\partial M_{jkt}/\partial w_{kt}$ is always positive. Assuming that $\int P_{ij}P_{ik}^2/\int P_{ij}P_{ik}$ is sufficiently small (a sufficient but not necessary condition for the term in parentheses to be positive is that $P_{ik} < 1/2$ for all individual), differentiating $\partial M_{jkt}/\partial w_{kt}$ with respect to ε^D_{jk} shows that the migration response to exogenous increases in the wage rate should be higher for city pairs with higher cross-price elasticities of demand:

$$\frac{\partial^2 M_{jkt}}{\partial \varepsilon^D_{jk}\partial w_{kt}} \sim \frac{2\varepsilon^D_{jk} + \lambda p_k}{\lambda p_k^2} > 0. \tag{7}$$

In other words, the net migration response to a labor demand shock in city k will be larger for cities that are closer substitutes. Thus, the second empirical validation is to evaluate candidate measures of city substitutability based on whether they predict the differential migration across cities in terms of how many residents move in response to exogenous shocks.

2.3. Empirical estimation

I propose a number of candidate proxies for city substitutability and evaluate them using migration data. As discussed, migration data are useful because individuals moving from one city to another observably substitute between them. I motivate this empirical estimation strategy using predictions derived from the city choice model in Section 3.2. Candidate measures that are most representative of demand substitution patterns should be the best predictors of migration flows.

The first empirical validation technique comes from the prediction that gross migration flows will be higher between cities which are substitutes. Let ϕ_{ij} represent a candidate measure of substitutability, either observable similarity in some amenity or home price correlation.[19] The main specification for evaluating ϕ_{ij} regresses migration from city i to city j between times $t-1$ and t as a fraction of city i's population, $M_{ijt} = m_{ijt}/pop_{it-1}$, on the propose measure and several controls:

$$M_{ijt} = \delta_{it} + \beta_1 \times \phi_{ij} + \beta_2 \times \chi_{ij} + \beta_3 \times pop_{jt} + v_{ijt}, \tag{8}$$

χ_{ij} represents controls for the distance between city i and city j and whether or not the two cities are in the same state, and pop_{jt} controls for the population of city j to avoid gross migration resulting from differences in city sizes.[20] The regressions also control for city i characteristics with origin city times time fixed effects. Migration from city j to city i is implicitly included since each left-hand side city shows up on the right-hand side of other city observations. Thus, the outward migration setup is sufficient to capture migration flows between every city pair combination.

The second empirical validation technique comes from the prediction that net migration flows in response to labor demand shocks should be higher between cities which are substitutes. I generate exogenous variation in labor demand by assuming that industry demand growth in each city occurs at national industry growth rates. Cities with larger fractions of rapidly growing industries receive higher shocks. This instrument, proposed by Bartik (1991) and popularized by Blanchard and Katz (1992), has been demonstrated to have a causal effect on migration in a number of studies of mobility and labor demand (see Bound and Holzer, 2000; Wozniak, 2006). The methodology used to construct the shocks is reviewed in Section 3.5. The main specification regresses migration from city i to city j on city j's labor demand shock interacted with candidate measures of substitutability:

$$\begin{aligned} M_{ijt} = \delta_{it} + \beta_1 \times LS_{jt} + \beta_2 \times LS_{jt} \times \phi_{ij} + \beta_3 \times LS_{jt} \times \chi_{ij} + \beta_4 \\ \times LS_{jt} \times pop_{jt} + \beta_5 \times \phi_{ij} + \beta_6 \times \chi_{ij} + \beta_7 \times pop_{jt} + v_{ijt}, \end{aligned} \tag{9}$$

where LS_{jt} denotes the labor demand shock for city j. Larger shocks should attract workers to city j since it has a greater short-run demand for labor. The interaction terms test whether net migration in response to labor demand shocks are larger for cities that are more similar.[21]

An alternative approach to identifying substitute cities would be to regress migration flows directly on prices or price changes in other cities. Unfortunately, this strategy is problematic since quantities and prices are endogenous and the timing is unclear. For example, if home prices are rising in city k, it seems equally plausible to observe net migration towards the city (causing the price increase) or away from the city (because of the price increase). In theory, one could execute such a strategy if exogenous variation in prices was large enough, yet the instrument to generate exogenous variation is large enough only to validate substitution rather than identify them directly.

One advantage of empirical specifications with migration from city j to city k is that migration flows are defined as a consistent proportion of the origin city's population. Alternative specifications were considered that regress gross or net migration flows as a proportion of the sum of the origin and destination populations, but achieving consistency in this regard was more difficult. Still, those specifications yielded similar qualitative results. See Appendix C.

The empirical specifications include city i fixed effects and allow identification off of similarity differences and labor demand shocks in city j.[22] The specifications omit destination fixed effects since controlling for them would remove the labor shock variation of interest. In addition, city j differences and labor demand shocks appear on the right-hand side of every city i, so that migration towards city j is considered relative to every city migration combination. Still, results are robust to controlling for destination region and state fixed effects. See Appendix C.

The main results in this paper are robust to a number of specification choices, including outward migration as the dependant variable, metropolitan versus state versus region observations, distance interactions, and destination controls. See Appendix C for a complete discussion.

[19] If home price correlation is informative above and beyond observables, it should have a positive coefficient even when the other observable measures are included as regressors. Importantly, the coefficient on home price correlation is not meant to have a causal interpretation. However, the regression format is still useful because it readily provides explained versus unexplained variance and allows multiple proxies to be compared at once. Typical endogeneity problems that arise when regressing quantities on prices are not relevant here because estimating the quantity response to price is not the goal.

[20] Ideally, I would also control for housing supply elasticity in city j, but I only have data available for 92 US cities. As a robustness check, I repeat these regressions with supply elasticity controls for this subsample. I also test specifications which control for more flexible functional forms for distance and population, including cubic polynomials. See Appendix C. In both cases, I find no qualitative difference in the main results, so I omit them and report only the specification described in Eq. (8).

[21] I include levels of city similarity so that the interaction term estimates only net migration above and beyond what would be expected via gross migration. I make sure to interact the labor demand shock with distance, same state, and population since the shocks should have a larger effect in cities that are closer, that are in the same state, and that have larger labor pools. In general, the coefficients on the interaction term should be larger for measures that better identify substitute cities.

[22] Because I interact the measures of similarity with the shocks, regressions that use net labor shocks (LS_{jt}–LS_{it}) are not identical to regressions that use city j shocks (LS_{jt}) even with city i fixed effects. Using net shocks allows the variation in city i's shock to impact β_2, β_3, and β_4. Although both regressions are similar in spirit and produce similar results, I use LS_{jt} to prevent city i variation from driving the coefficients.

Author's personal copy

3. Data

3.1. Overview

The data originate from a variety of sources, including the Office of Federal Housing Enterprise Oversight (OFHEO) metropolitan-level home price indices, the Integrated Public Use Microdata Series (IPUMS) US Census 5% sample, and the IPUMS March Consumer Population Survey. Merging the data sources together is possible by standardizing the Metropolitan Statistical Area (MSA) and Primary Metropolitan Statistical Area (PMSA) codes and checking the matching manually.

One unit of observation is a metropolitan area. The IPUMS Census database reports residential locations in terms of Public Use Microdata Sample Areas (PUMAs) for 1990 and 2000 and county groups for 1980. PUMAs and county groups map individuals into MSAs and PMSAs according to the 2000 Census definitions.[23] PMSAs are used where Census geographical definitions allow it and MSAs elsewhere; both are hereafter referred to as MSAs. Out of 327 potential MSAs, the sample includes 247 MSAs for 1975–1980, 289 MSAs for 1985–1990, and 297 MSAs for 1995–2000, producing 231,906 total city pairs over the three decades.[24]

3.2. Observable similarity

Ten measures of observable similarity are identified and tested: average temperature (in degrees Fahrenheit), whether cities are within 100 miles of a coastline, political affiliation based on the Republican voting share in the 2004 presidential election, age composition, racial composition, education composition, industrial composition, mean house value, per capita income, and log population density.[25] For age, race, education, education, and industrial compositions, similarity is measured based on the sum of absolute share differences between city pairs.[26]

Table 2 presents the summary statistics for these ten similarity measures plus controls for distance and whether two cities are in the same state. The variables are categorized as time-varying or time-invariant depending on whether they are calculated over time or based on a single year. Taken together, these measures provide substantial variation in observable similarity across city pairs. For the empirical estimation, all similarity measures are normalized with means equal to 0 and standard deviations equal to 1.

3.3. Home price correlation

As discussed, I propose home price correlation as a price-based measure of demand substitutability.[27] Theory predicts that home price correlation can represent substitution patterns if shocks to any one city affect demand in other cities. Home price correlation will be especially pronounced when short-run housing supply is inelastic.[28] In theory, home prices might be correlated without being substitutes in demand if they are subject to correlated economic shocks. In Appendix B, I provide a two-city simultaneous equations model to demonstrate correlation resulting from substitution versus correlated economic shocks. The model shows that, all else being equal, home price correlation is higher for cities with higher cross-price elasticities of demand. Ultimately, whether home price correlation represents demand substitution is an empirical question to be tested.

I calculate pairwise correlation in home price changes between cities using quarterly home price indices published by the Office of Federal Housing Enterprise Oversight (OFHEO) for 381 MSAs from 1975–2008. The home price indices are constructed using a weighted repeat-sales methodology for single-family homes, so price increases reflect average home price appreciation over pairs of sales or refinancing of the same homes. Home price indices are also published at the state-level under the same general guidelines.[29]

Home price correlation between two cities is calculated as follows:

$$\rho_{ij} = corr(\Delta \log p_{it}, \Delta \log p_{jt}) = corr\left(\log \frac{p_{it}}{p_{it-1}}, \log \frac{p_{jt}}{p_{jt-1}}\right), \tag{10}$$

where p_{it} and p_{jt} refer to home prices in cities i and j in time t. Price correlation picks up simultaneous trend deviations rather than differences in long-run growth rates. Correlation is used in lieu of

---

[23] One challenge is that MSA/PMSA definitions are not constant but in fact change as cities grow and shrink over time. The OFHEO home price indices use constant MSA definitions according to the most current definitions, making them subject to historical revision as MSA definitions change. I use the 2000 Census definitions in order to maintain consistency for all Census years but the results are robust to alternative metropolitan definitions that change over time. In any case, the state specifications discussed in Appendix C do not have this problem and thus represent an additional robustness check on this dimension.

[24] By "potential MSAs", I refer to the number of MSAs that could be identified if Census data reported individuals' counties of residence. Because the public Census data report residences only as finely as the PUMAs or county groups, the number of identifiable MSAs is reduced. The total number of city pairs is $231{,}906 = 247 \times 246 + 289 \times 288 + 297 \times 296$.

[25] Literature that estimates quality of life differences between cities tends to control for a more diverse set of climate variables, including precipitation, cooling degree days, heating degree days, relative humidity, sunshine, and wind speed. See Blomquist et al. (1988) and Gyourko and Tracy (1991). Since my aim is merely to evaluate whether observable similarity is predictive of substitution patterns, for simplicity I use average temperature throughout the year. However, alternative specifications have been explored. See Appendix C. Other measures of city similarity are generally consistent with the types of controls in the quality of life and migration literature, which control for differences in distance (Davies et al., 2001), coastline (Blomquist et al., 1988; Gyourko and Tracy, 1991), age (Gabriel et al., 2003; Kennan and Walker, 2010a), education (Costa and Kahn, 2000; Chen and Rosenthal, 2008), house value/housing expenditures (Gyourko and Tracy, 1991; Kennan and Walker, 2010a), per capita income (Davies et al., 2001), and population density (Whisler et al., 2008).

[26] There are seven categories for age (0–17, 18–24, 25–34, 35–44, 45–54, 55–64, and >65), four categories for race (White, Black, Hispanic, and Other), four categories for education (Less than high school, High school, Some college, and College), and 10 categories based on broad industry codes: Agriculture/Mining (1–59), Construction (60–99), Manufacturing (100–399), Transportation (400-499), Trade (500-699), Business Services (700–760), Personal Services/Entertainment (761–811), Healthcare/Law/Education (812–861), Other Services (862–899), and Government/Military (900–960).

[27] There are several reasons to prefer home price correlation to other priced-based measures such as wage or rent correlations. First, home prices are notoriously volatile while wages and rents are notoriously sticky. Thus, correlation in prices is likely to provide more information than correlations in the others. See Case and Shiller (1987, 1989) for a descriptive discussion of home price indices over time. Second, heterogeneity of home price changes within cities is small compared to heterogeneity in wage changes, which tend to be dominated more by occupation than by city (see Case and Cook (1989) and McDuff (2010) for discussion and references of how home prices move within metropolitan areas). Lastly, OFHEO home prices are available for 381 metropolitan areas, compared to just a handful where rent series are available, making the increased coverage appealing from a practical viewpoint.

[28] Glaeser et al. (2008) show that price volatility is larger and more frequent in cities for which housing supply is more inelastic but do not discuss the implications for demand substitution across city pairs.

[29] The OFHEO home price indices have two primary disadvantages relative to other home price indices, most notably the Case–Shiller/S&P (CS) home price indices which are published by the Chicago Mercantile Exchange. First, OFHEO uses refinance appraisals in addition to sales which may increase price persistence, especially in a market downturn. The CS indices which exclusively use sale prices have estimated more rapid price declines in the last few years. Second, the OFHEO housing sample is truncated at $417,000 according to Fannie Mae and Freddie Mac guidelines whereas the CS indices do not have this problem since they use transactions from county and state records databases. There are a number of less important differences, but the main outcome is that OFHEO home price indices tend to show more price persistence in the short-term than the CS indices. In the end, I use the OFHEO indices because they are available for 381 cities versus only 20 published CS home price indices.

Author's personal copy

Table 2
Summary statistics.

Variable	Year	Definition	N	Mean	SD	Min	Max
Single city							
Population (1,000,000s)	1980	pop_{it-1}	247	0.62	1.03	0.10	8.22
	1990		289	0.63	1.06	0.10	8.85
	2000		297	0.71	1.18	0.10	9.52
Outward migration	1980	$M_{it} = 100 \times m_{it}/\text{pop}_{it-1}$	247	25.67	7.26	13.50	67.87
	1990		289	27.32	6.92	14.66	56.73
	2000		297	26.56	6.74	15.39	55.33
Labor demand shock	1980	LS_{it}	126	0.120	0.023	−0.008	0.155
	1990		242	0.071	0.016	0.012	0.123
	2000		287	0.062	0.014	−0.008	0.092
Housing supply elasticity	2000	ε_i^S	92	1.58	0.85	0.57	5.16
City pairs							
Outward migration to city *j*	1980	$M_{ijt} = 100 \times m_{ijt}/\text{pop}_{it-1}$	60,762	0.10	0.40	0.00	24.69
	1990		83,232	0.09	0.39	0.00	24.33
	2000		87,912	0.09	0.38	0.00	16.55
Time-invariant city similarity							
Distance		−ln(distance in miles)	87,912	−6.71	0.78	−8.54	−2.52
Same state		1 = same state	87,912	0.03	0.18	0.00	1.00
Average temperature (°F)		$-\lvert\text{temp}_i - \text{temp}_j\rvert$	87,912	−8.97	6.62	−38.50	0.00
Coastline		1 = both coastal or not coastal	87,912	0.38	0.49	0.00	1.00
Political affiliation		$-\lvert(\%\ \text{Rep.})_i - (\%\ \text{Rep.})_j\rvert$	87,912	−11.82	9.06	−66.00	0.00
Time-varying city similarity							
Age		$-(\lvert s_{i1} - s_{j1}\rvert + \ldots + \lvert s_{ik} - s_{jk}\rvert)$	231,906	−1.63	1.00	−8.98	−0.09
Race		$-(\lvert s_{i1} - s_{j1}\rvert + \ldots + \lvert s_{ik} - s_{jk}\rvert)$	231,906	−1.28	1.00	−5.93	−0.01
Education		$-(\lvert s_{i1} - s_{j1}\rvert + \cdots + \lvert s_{ik} - s_{jk}\rvert)$	231,906	−1.88	1.00	−6.87	−0.03
Industry		$-(\lvert s_{i1} - s_{j1}\rvert + \cdots + \lvert s_{ik} - s_{jk}\rvert)$	231,906	−2.43	1.00	−8.58	−0.22
Home value		$-(\lvert hv_i - hv_j\rvert/\$100{,}000)$	231,906	−0.77	0.66	−5.21	0.00
Per capita income		$-(\lvert\text{inc}_i - \text{inc}_j\rvert/\$10{,}000)$	231,906	−2.04	2.15	−22.04	0.00
Population density		$-(\lvert\ln(\text{ppsq}_i) - \ln(\text{ppsq}_j)\rvert)$	231,906	−1.02	0.83	−7.43	0.00
Between city price correlations							
OFHEO home price corr. 1Q		$\text{corr}(\Delta\ln\_P_{it}, \Delta\ln\_P_{jt})$	87,912	0.11	0.18	−0.72	0.91
OFHEO home price corr. 4Q		$\text{corr}(\Delta\ln\_P_{it}, \Delta\ln\_P_{jt})$	87,912	0.18	0.29	−0.73	0.98
OFHEO home price corr. 8Q		$\text{corr}(\Delta\ln\_P_{it}, \Delta\ln\_P_{jt})$	87,912	0.18	0.37	−0.87	0.99
OFHEO home price corr. 20Q		$\text{corr}(\Delta\ln\_P_{it}, \Delta\ln\_P_{jt})$	87,912	0.18	0.50	−0.96	1.00

LS = labor demand shock, M = outward migration (%), m = outward migration (# individuals), pop = population, ε^S = housing supply elasticity
temp = degrees Fahrenheit, % Rep = % Republican in 2004 presidential election, s_{ik} = share of city *i* population in category *k*, hv = home value
inc = per capita income, ppsq = population per square mile, Δln_P = log change home prices, Δln_R = log change in rents

Notes: The above data come from the IPUMS March Current Population Survey and the IPUMS Census databases. Observations occur over fixed, 5-year time intervals: 1975–1980, 1985–1990, and 1995–2000 in accordance with Census migration data. Migration flows are calculated using the 5% Census sample which contains individuals' current residences and residences 5 years prior. Labor demand shocks are constructed according to 2-digit industry shares from the previous Census and national labor growth for each industry according to the March CPS survey. The top panel summarizes the 297 MSAs/PMSAs, some of which are not identified in earlier Census years. The subsequent panels describe variables representing city pairs, so that $N_{1980} = 60{,}762 = 247 \times 246$, $N_{1990} = 83{,}232 = 289 \times 288$, and $N_{2000} = 87{,}912 = 297 \times 296$. Home price correlations are calculated using the OFHEO home price indices, the Case–Shiller/S&P home price indices, and the BLS Consumer Price Index rent series, the latter being available for only 18 cities ($N = 306 = 18 \times 17$). Housing supply elasticity is provided for 92 cities by Saiz (2010).

covariance to remove differences in the amplitude of price changes and instead focus on timing. This is important for city pairs which are demand substitutes but have differing housing supply elasticities.[30]

Table 2 presents summary statistics for home price correlation between city pairs. Correlations are calculated ranging from 1 to 20 quarters, although the main specifications utilize 4-quarter correlation to capture short-term correlations while avoiding issues of seasonality inherent in the single-quarter correlations.

3.4. Migration

I use migration data from the IPUMS US Census 5% sample for 1980, 1990, and 2000 which report individuals' current residences and residences 5-years prior to the survey. In order to make consistent comparisons across cities and time, I define outward migration from city *i* to city *j* between $t-1$ and t as the number of migrating individuals, m_{ijt}, divided by city *i*'s population in the base year (1975, 1985, or 1995): $M_{ijt} = m_{ijt}/pop_{it-1}$. Thus, a single observation consists of migration from city *i* to city *j* as a fraction of city *i*'s population over one of the following three time periods: 1975–1980, 1985–1990, or 1995–2000.

Table 2 presents summary statistics for migration between city pairs. On average, outward migration from a given city over a 5-year migration interval is around 25% of its population and the migration to any specific city is 0.1%. Outward migration is sharply skewed since few destination cities receive the majority of outward migration from any one city. Nearly three-quarters of city pairs have zero migration flows between them in the 5% Census sample. Migration flows are calculated based on a restricted sample of individuals 30 and older to remove migration flows related to youth moves and college attendance.

3.5. Labor demand shocks

Exogenous variation in labor demand is obtained using the methodology of Bartik (1991) by assuming that industry demand

[30] Gyourko et al. (2006) present empirical evidence that long-run price growth is higher in cities for which housing supply is more inelastic since new demand raises prices instead of quantities when construction is limited. With this in mind, using correlation instead of covariance is important for capturing demand substitution that occurs between cities of different price levels, long-run growth rates, and the variance of price shocks since correlation by definition abstracts from these long-run parameters.

Author's personal copy

Table 3
Similarity correlations.

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Migration	1.00													
2	Distance	0.30	1.00												
3	Same state	0.42	0.42	1.00											
4	HPC	0.18	0.19	0.24	1.00										
5	Temperature	0.08	0.36	0.17	0.08	1.00									
6	Coastline	−0.03	−0.22	−0.04	0.04	−0.13	1.00								
7	Political	0.01	0.11	0.01	0.08	0.00	−0.04	1.00							
8	Age	0.03	0.13	0.01	−0.04	0.17	−0.05	0.06	1.00						
9	Race	0.05	0.36	0.06	−0.02	0.30	−0.09	0.04	0.11	1.00					
10	Education	0.06	0.20	0.05	0.08	0.06	−0.04	0.17	0.32	0.31	1.00				
11	Industry	0.03	0.15	0.06	0.06	0.14	−0.06	0.12	0.22	0.12	0.28	1.00			
12	Home value	0.01	0.25	0.03	0.01	0.01	−0.12	0.22	0.23	0.25	0.34	0.21	1.00		
13	Income	0.03	0.02	0.00	0.04	0.02	−0.04	0.08	0.34	0.10	0.33	0.28	0.26	1.00	
14	Density	−0.02	0.19	0.02	0.00	0.03	−0.02	0.17	−0.02	0.19	0.06	0.14	0.17	0.00	1.00

Notes: The above table displays pairwise correlations for a number of the similarity measures from Table 2 for the year 2000. HPC stands for home price correlation. The migration variable equals the sum of the migration flows between city pairs divided by the sum of the populations. The sample size of $N = 43{,}956 = 297 \times 296/2$ is generated by 297 identifiable metropolitan areas in the 2000 US Census.

within each city grows at national industry growth rates. Cities with large fractions of nationally growing industries receive greater shocks. First, national industry growth rates are calculated as follows:

$$g_{kt} = \frac{L_{kt} - L_{kt-1}}{L_{kt-1}} - 1, \tag{11}$$

where L_{kt} is the national labor in industry k at time t. Next, city-level labor demand shocks are imputed based on industry shares in each city:

$$LS_{jt} = \sum_{k=1}^{K} \left(\frac{L_{jkt-1}}{L_{jt-1}}\right)(1 + g_{kt}) - 1, \tag{12}$$

where L_{jkt-1}/L_{jkt-1} is the fraction of industry k in city j at time $t-1$ and g_{kt} is the national growth rate of that industry between $t-1$ and t.

Labor demand shocks are estimated for each city in the sample over three time periods relevant for migration flows: 1975–1980, 1985–1990, and 1995–2000. Industries are based on over 200 industry categories from the first two digits of Census industry codes. The previous Census is used to estimate city-level industry shares and the IPUMS CPS is used to calculate national industry growth rates.

Table 2 presents summary statistics for the labor demand shocks.[31] The standard deviations of the labor shocks are low relative to the means, suggesting substantial variation which is not cross-sectional. Still, there exists sufficient variation to produce meaningful results using a large sample of city pairs. The labor demand shocks are treated as exogenous, which will be true as long as industry shares adjust slowly relative to national industry growth rates and single cities do not meaningfully affect the results.[32]

4. Results

4.1. Similarity correlations

Table 3 presents the raw city-pair correlations between migration and the full set of similarity measures. In this table, migration is measured as the sum total migration flows between two cities divided by the sum of the populations. As expected, most of the similarity measures are positively correlated with one another, and positively correlated with migration. Correlations with migration are highest for same state (0.42) and distance (0.30), followed by home price correlation (0.18), temperature (0.08), education (0.06), and race (0.05). The low magnitude of these correlations is unsurprising given the wide range of factors that influence migration choices.

Home price correlation is positively correlated with the majority of similarity measures. Correlations are highest with temperature (0.08), political affiliation (0.08), and education (0.08). Of all of the similarity measures, home price correlation has the highest raw correlation with migration (0.18), nearly twice as large as temperature (0.08), the next highest. The fact that home price correlation is positively correlated with distance (0.19) and same state (0.24) as well as the majority of similarity measures is consistent with the hypothesis that home price correlation is a useful measure of substitutability.[33]

4.2. Two empirical validation examples

Two empirical validation techniques are employed to test candidate measures of substitutability. The first evaluates measures as predictors of cross-sectional gross migration flows. The second evaluates measures as predictors of net migration in response to

[31] Because the granularity of Census geography increases over time, using the previous Census to calculate industry shares reduces the sample size since an MSA must be identified in two consecutive decades. This problem is especially severe in 1980 since the 1970 county groups identify many fewer MSAs than the 1980 Census. Table 2 indicates that labor demand shocks are available for only 126 MSAs in 1980, 242 MSAs in 1990, and 287 MSAs in 2000. I am careful to make sure that this incomplete coverage does not influence the main results of this paper by running all regressions over a variety of "consistent samples" that use only cities which are identified in every time period as well as "maximized samples" which impute industry shares using the current Census. All results in Section 5 are robust to these subsample specifications. In addition, state specifications do not have this problem and thus provide an additional robustness check. Because results are not seem sensitive to sample selection, I present gross migration with the full migration sample and net migration only where labor demand shocks are available even though the samples size differs slightly for the two tests.

[32] Although the main results are presented with national industry growth rates calculated using the full CPS sample, I have tested labor demand shocks that leave out the city/state in question. The results are not sensitive to the methodology choice at the city or the state level, so I present the simpler formulation that uses the raw national labor for each industry.

[33] A previous version of this paper included a regression of home price correlation on distance, same state, and the other measures of observable similarity. Together, these variables accounted for roughly 10% of the variance in home price correlation, indicating that there is quite a lot of information in home price correlation not contained in the other measures.

Author's personal copy

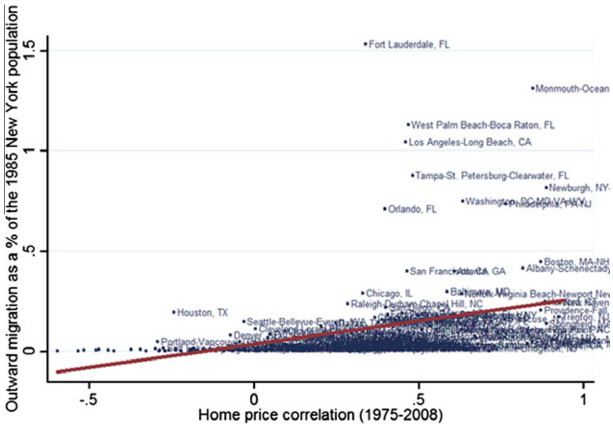

Fig. 2. The figure above displays outward migration from New York, NY to other metropolitan areas across the United States from 1985–1990. The *x*-axis displays the year-over-year home price correlation between New York and the metropolitan areas, and the *y*-axis displays outward migration to each city as a percent of New York's 1985 population. The positive slope of the fitted line indicates that, on average, individuals tend to move to cities whose home price indices have a high correlation with New York's.

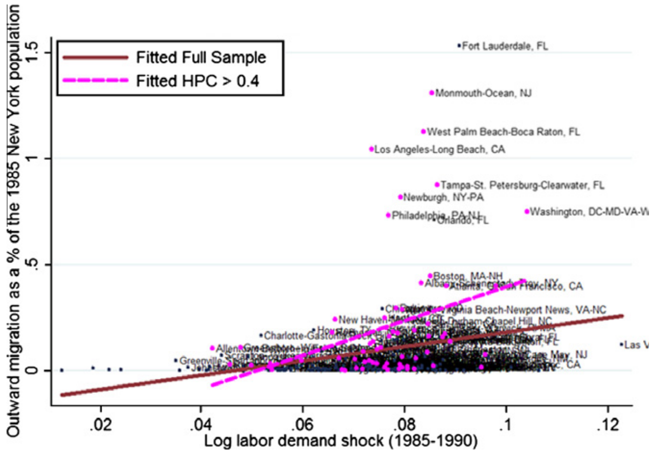

Fig. 3. The figure above displays outward migration from New York, NY to other metropolitan areas across the United States from 1985–1990 in response to labor demand shocks. Labor demand shocks for each city are calculated by assuming that industry demands in each city grow at national industry growth rates, so that cities with higher fractions of nationally growing industries experience larger shocks. The *x*-axis displays the labor demand shocks, and the *y*-axis displays net outward migration to each city as a fraction of New York's 1985 population. The slope of the fitted line for city pairs for which the home price correlation is greater than 0.4 is steeper than the fitted line for the full sample, indicating that migration is more sensitive to labor demand shocks among cities whose home price indices have a higher correlation with New York's.

exogenous labor demand shocks. I present two examples in the context of evaluating home price correlation as a substitutability measure.

Fig. 2 displays a scatter plot of outward migration flows from New York between 1985 and 1990 and home price correlations between New York and other US cities. The *y*-axis plots outward migration as a fraction of New York's 1985 population, and the *x*-axis plots year-over-year home price correlations with New York's home price index. The positive relationship between these two variables is immediately obvious from the figure. Cities with higher home price correlations with New York tend to attract a higher percentage of New York's residents.[34] As discussed, home price correlation is highly correlated with distance and other similarity measures, so multivariate regressions analysis is required to determine whether this relationship holds when controlling for other factors.

Fig. 3 displays a scatter plot of outward migration flows from New York between 1985 and 1990 and labor demand shocks for US cities over that period. The *y*-axis plots outward migration as a fraction of New York's 1985 population and the *x*-axis plots labor demand shocks for the cities. The fitted full sample relationship indicates that cities with higher labor demand shocks tend to attract higher outward migration flows from New York. When the sample is restricted to cities with home price correlation above 0.4, the relationship is even stronger. The full sample has a response of about 0.03 percentage points of migration per percentage point shock, whereas the high-HPC sample has a response of 0.07. As discussed, analyzing these types of relationships using continuous specifications and controlling for other factors requires multivariate regression analysis.

4.3. Empirical validation 1: Gross migration flows

The first empirical validation is whether candidate measures of similarity predict gross migration flows between city pairs.

Table 4 summarizes the main results from the regression specification in Eq. (8), which regression outward migration from city *i* to city *j* on a variety of similarity measures. All regressions include city *i* times time fixed effects, so that the coefficients are identified

Table 4
Gross migration.

Dependent variable: city-to-city migration = $M_{ijt} = 100^{*}\ (m_{ijt}/\mathrm{pop}_{it-1})$	(1)	(2)
Distance	0.210*	0.206*
	(0.018)	(0.022)
Same state	0.336*	0.333*
	(0.030)	(0.030)
Home price correlation		0.045*
		(0.008)
Temperature		−0.051*
		(0.023)
Coastline		0.004
		(0.008)
Political affiliation		0.032*
		(0.007)
Age		−0.011
		(0.012)
Race		0.017
		(0.016)
Education		0.008
		(0.010)
Industry		−0.012
		(0.010)
Home value		0.028
		(0.023)
Income		0.032*
		(0.009)
Density		0.008
		(0.014)
Population in city *j*	0.228*	0.235
	(0.040)	(0.037)
Fixed effects: year* city *i*	Yes	Yes
N	231,906	231,906
R^2	0.2628	0.2692

Notes: The table above regresses outward migration from city *i* to city *j* as a fraction of city *i*'s population on a variety of city similarity measures. All right-hand side variables measure differences between city pairs on the indicated measure. All regressions include year times city *i* fixed effects and controls for distance, same state, and population. All left-hand side and right-hand side variables have been normalized to have mean 0 and standard deviation 1. Standard errors are reported in parentheses and are clustered for city *j*'s state.
* Significant at the 10% level.

[34] Nassau–Suffolk NY has a home price correlation of 0.872 and a migration rate of 4.8% and is too high on the *y*-axis to be included. However, the positive relationship holds with or without including this city.

Author's personal copy

off of the similarity in potential cities to which individuals might move. As discussed, all variables have been normalized to mean 0 and standard deviation 1, so that coefficients can be interpreted as a ratio of standard deviations (*i.e.*, a 1 standard deviation change in *X* is associated with how many standard deviations in *Y*). All standard errors are clustered for city *j*'s state, so that error terms are correlated within but not across states.

Column (1) displays the basic specification which includes only distance, same state, and the population of city *j* on the right-hand side. Unsurprisingly, all three coefficients are large in magnitude and statistically significant. The distance coefficient (0.210) indicates that a 1 standard deviation more similar in distance is associated with a 1/5 standard deviation increase in migration flows. The same state coefficient (0.336) indicates that being in the same state is associated with a 1/3 increase in migration flows. The R^2 of 0.2628 is large and indicates that a substantial portion of gross migration can be explained by local controls alone.

Column (2) displays the regression specifications with candidate measures of similarity on the right-hand side. As before, distance and same state have statistically significant coefficients of 0.206 and 0.333, respectively. Only a handful of similarity measures are statistically significant after controlling for local controls of distance and same state: home price correlation (0.045), temperature (−0.051), political affiliation (0.032), and income (0.032). Several other measures (race, education, home value, and density) have positive and statistically significant coefficients when distance and same state are not included (results not shown here).

The negative coefficient on temperature suggests a weakness in the theory. However, two explanations provide insight into this result. First, when interacted with distance (results not shown), the coefficient on temperature (−0.035) and the coefficient on the interaction term (0.077) suggest that temperature has the expected positive coefficient at close distances (over which most migration occurs) but has a negative coefficient for longer distances. The negative coefficient is driven by long-distance, different-temperature city pairs where much migration occurs (*e.g.* New York and Florida) and long-distance, same-temperature city pairs where little migration occurs (*e.g.* California and Texas).[35] The result provides support for migration models that study migration patterns resulting from changes in life circumstance, such as retirement (see Graves and Knapp, 1988;Graves and Waldman, 1991; Chen and Rosenthal, 2008).

Second, the negative coefficient on temperature actually increases the utility of home price correlation as a market-based measure of substitutability. Unlike temperature, for which there are intuitive stories for either a positive or negative coefficient, home price correlation provides a specific coefficient prediction based on an agnostic approach to the importance of specific city amenities. As such, it is empirically one of the most explanatory and most robust measures of city similarity. For example, the New York–Florida migration channel has a high home price correlation even though the temperature difference is large. This suggests that other, non-temperature factors—perhaps difficult-to-measure factors such as culture, lifestyle, etc.—may make these cities substitutes in demand.

Table 5 displays the goodness of fit for each measure of city similarity, which is another useful statistic for evaluating the predictive power of the various substitutability proxies. The table provides R^2 measures for regressions on individual similarity measures as well as those measures interacted with distance. The latter is useful since the similarity measures may be more informative when cities are closer together. Compared to the baseline R^2 of 0.2628 for distance, same state, and city *j* population, the measures that add the most predictive power are: temperature (0.2647), home price correlation (0.2646), and political affiliation (0.2641). When interacted with distance, the measures that add the most predictive power are: home price correlation (0.2801), temperature (0.2729), and race (0.2666).

Table 5
Gross migration goodness of fit.

Dependent variable: city-to-city migration = $M_{ijt} = 100 \times (m_{ijt}/pop_{it-1})$	R^2 Linear	R^2 ×Distance
Distance + Same state + City *j* population	0.2628	n/a
Home price correlation	0.2646	0.2801
Temperature	0.2647	0.2729
Coastline	0.2629	0.2656
Political affiliation	0.2641	0.2656
Age	0.2628	0.2656
Race	0.2629	0.2666
Education	0.2632	0.2657
Industry	0.2628	0.2652
Home value	0.2640	0.2659
Income	0.2638	0.2656
Density	0.2630	0.2632
All observables	0.2677	0.2829
All observables + Home price correlation	0.2692	0.2970

Notes: The table above displays the goodness of fit for regressions of outward migration from city *i* to city *j* as a fraction of city *i*'s population on a variety of city similarity measures. The baseline regression includes only log distance, same state, city *j* population as well as city *i* fixed effects. Other variables are added as indicated. The interactions with distance include a variable that interacts the indicated measure with log distance in miles.

Table 6 displays gross migration results with varying home price correlation durations. Theory suggests that shorter durations should be more predictive, since long run adjustments in wages and housing supply tend to reduce home price correlation over time. Consistent with this hypothesis, coefficients on home price correlation are: 1-quarter, 0.045; 4-quarter, 0.045; 8-quarter, 0.039; and 20-quarter, 0.033. When all are included together, the 1-quarter and 4-quarter correlation measures are the only two that retain statistical significance.

4.4. Empirical validation 2: Net migration flows in response to labor demand shocks

The second empirical validation is whether candidate measures of similarity predict net migration flows in response to labor demand shocks.

Table 7 summarizes the main results from the regression specification in Eq. (9), which regression outward migration from city *i* to city *j* on labor demand shocks interacted with a variety of city similarity measures. All regressions include city *i* times time fixed effects, so that the coefficients are identified off of the similarity in potential cities to which individuals might move.[36] As discussed, all similarity measures have been normalized to mean 0 and standard deviation 1. All standard errors are clustered for city *j*'s state, so that error terms are correlated within but not across states.

Column (1) displays the basic specification which includes the net labor demand shock interacted with distance, same state, and population of city *j* on the right-hand side. The coefficient of 1.655 on the labor shock indicates that a 1% point increase in labor

[35] One referee suggested running separate panels for elderly and non-elderly samples. Results are qualitatively similar for both groups, although the negative coefficient on temperature is attenuated for the non-elderly (−0.037) and accentuated for the elderly (−0.095). See Table 8 for more discussion of robustness to sample selection.

[36] Results are robust to including destination region and state fixed effects. See Table 8.

Author's personal copy

Table 6
Gross migration with varying home price correlation durations.

Dependent variable: city-to-city migration = $M_{ijt} = 100^*$ (m_{ijt}/pop_{it-1})	(1)	(2)	(3)	(4)	(5)
Home price correlation-1Q	0.045*				0.027*
	(0.007)				(0.006)
Home price correlation-8Q		0.045*			0.050*
		(0.008)			(0.025)
Home price correlation-4Q			0.039*		−0.040
			(0.029)		(0.041)
Home price correlation-20Q				0.033*	0.017
				(0.008)	(0.017)
Observable similarity	Yes	Yes	Yes	Yes	Yes
Distance	Yes	Yes	Yes	Yes	Yes
Same state	Yes	Yes	Yes	Yes	Yes
Population in city j	Yes	Yes	Yes	Yes	Yes
Fixed effects: year* city i	Yes	Yes	Yes	Yes	Yes
N	231,906	231,906	231,906	231,906	231,906
R^2	0.2694	0.2692	0.2688	0.2686	0.2697

Notes: The table above regresses outward migration from city i to city j as a fraction of city i's population on home price index correlations of varying durations: 1 quarter, 4 quarters, 8 quarters, and 20 quarters. All left-hand side and right-hand side variables have been normalized to have mean 0 and standard deviation 1. All regressions include year times city i fixed effects and controls for distance, same state, and population. Standard errors are reported in parentheses and are clustered for city j's state.
* Significant at the 10% level

Table 7
Net migration in response to labor demand shocks.

Dependent variable: city-to-city migration = $M_{ijt} = 100^*$ (m_{ijt}/pop_{it-1})	(1)	(2)
City j labor shock, LS_j	1.655*	1.769×
	(0.385)	(0.373)
$LS_j *$ distance	0.444*	0.436*
	(0.173)	(0.168)
$LS_j *$ same state	1.414*	1.375*
	(0.340)	(0.355)
$LS_j *$ home price correlation		0.427*
		(0.087)
$LS_j *$ temperature		−0.424*
		(0.179)
$LS_j *$ coastline		0.007
		(0.068)
$LS_j *$ political affiliation		0.170*
		(0.093)
$LS_j *$ age		−0.171
		(0.110)
$LS_j *$ racial		0.088
		(0.165)
$LS_j *$ education		−0.065
		(0.100)
$LS_j *$ industry		0.076
		(0.116)
$LS_j *$ home value		0.263*
		(0.101)
$LS_j *$ income		−0.030
		(0.129)
$LS_j *$ density		0.094
		(0.142)
Observable similarity levels	Yes	Yes
Fixed effects: year* city i	Yes	Yes
N	156,154	156,154
R^2	0.3036	0.3050

Notes: The table above regresses outward migration from city i to city j as a fraction of city i's population on labor demand shocks interacted with a variety of city similarity measures. Labor demand shocks for each city are calculated by assuming that industry demand in each city grow at the national industry growth rates, so that cities with higher fractions of nationally growing industries experience larger shocks. All interacted variables measure differences between city pairs on the indicated measure. All regressions control for difference levels (not interacted), city i times time fixed effects, and interacted controls for distance, same state, and population. All right-hand side variables (except the labor demand shocks) have been normalized to have mean 0 and standard deviation 1. Standard errors are reported in parentheses and are clustered for city j's state.
* Significant at the 10% level.

demand in city j corresponds to a 0.017 percentage point increase in migration flows from city i to city j for a baseline city pair of average distance not in the same state. The coefficients on distance and same state suggest that the baseline coefficient for two cities in the same state with no distance between them is 5.431. As expected, migration between city pairs is more sensitive to labor demand shocks when cities are close together and in the same state.

Column (2) displays the regression specification with candidate measures of similarity interacted with labor demand shocks. As before, city measures are normalized so that the coefficient represents a 1 standard deviation increase in similarity. Only a handful of substitutability measures are statistically significant in this specification (although many more are so when included individually): home price correlation (0.427), temperature (−0.424), home value (0.263) and political affiliation (0.170). These results indicate that migration flows between city pairs are more sensitive to labor demand shocks when they are more similar on these dimensions (see Section 4.3 for a more in-depth discussion of the temperature variable).

Due to the substantial data requirements and decisions required to implement the empirical specifications in this paper, the main results of this paper—particularly the coefficient signs in both empirical validation tests—have been confirmed on a number of dimensions. See Table 8.

5. Conclusion

This paper makes several strides towards understanding demand substitution across US cities. First, I show how aggregate demand substitution arises from individuals with varying amenity preferences being on the margin between pairs of similar cities. Next, I propose several tests using migration data to evaluate candidate measures of substitutability and implement those tests empirically. Finally, I propose home price correlation as an alternative, price-based measure of substitutability.

Results suggest a number of substitutability measures that are predictive of migration flows, but few add substantial predictive power above and beyond distance and state controls. The most predictive observable similarity measures are temperature, political affiliation, home value, and per capita income. Home price correlation as a price-based measure, by and large, outperforms the identified measures of observable similarity in predicting migration flows.

Author's personal copy

Table 8
Robustness checks.

Empirical specification	As noted in Section 3.3, specifications with gross migration and net migration between city pairs have been tested as alternatives to the outward migration specifications in Eqs. (8) and (9). Ultimately, these specifications were removed due to difficulties associated with having consistent measures of migration flows between city pairs of varying sizes. Results are by and large consistent with reported results, particularly regarding the empirical relationship between migration and home price correlation
State and region observations	Empirical specifications in this paper are reported at the metropolitan level. Because of the empirical challenges related to MSA definitions, all specifications have been tested at the state-level and region-level using the OFHEO state and region home price indices. Results are both qualitatively and quantitatively similar, particularly as they relate to relative values of home price correlation versus other similarity measures
Distance interactions	As discussed, interacting home price correlation and similarity measures with distance has been tested and evaluated. While interesting and generally supportive, results have been omitted unless otherwise noted
Age samples	Various age samples have been tested, including the full census sample (all ages), ages 25 and up, ages 30 and up, ages 30–60, ages 30–65, ages 60 and up, and ages 65 and up. As noted, the coefficient magnitude of the temperature similarity measure increases for the elderly samples in both empirical validations, though the sign generally remains negative and statistically significant for all samples. Other results are largely unchanged
Migration matrix zeros	Because of the large number of zeros in the city-to-city migration matrix, results have been tested using an exclusively non-zero migration sample. Results are largely unchanged
Destination fixed effects	Destination region and state fixed effects have been added to account for regional differences in migration patterns. The coefficient on home price correlation is somewhat attenuated but still statistically significant for region and state fixed effects in gross migration specifications and essentially unchanged in net migration in response to labor demand shock specifications. Other results are largely unchanged
Similarity measure specification	A variety of methods for quantifying similarity measures have been tested, where possible, including sum of absolute differences, sum of squared differences, temperature differences defined by various monthly combinations (e.g., January differences, June differences, sum of absolute differences, etc.), distance, log distance, squared distance, cubic distance, population density, log population density, and others. Generally speaking, the directions of statistically significant signs are unchanged by variable specification
Locally weighted regression techniques	Locally weighted regressions of net migration in response to labor demand shocks have been investigated to determine sensitivity, if any, to the imposition of linear specifications. Generally speaking, distance and home price correlation in particular have more qualities of non-linearity, meaning that linear specifications will tend to understate the effect of being similar in distance or home price correlation
Alternative correlation measures	In addition to the OFHEO home price indices, correlation measures have been tested using Case–Shiller indices and rent series from the Bureau of Labor Statistics. Generally speaking, the magnitudes of the coefficients are smaller but still statistically significant. These results could only be run for a smaller sample for which these measures are available, but they are generally supportive of alternative correlations carrying similar information
Combinations of similarity measures and local controls	Various combinations of similarity measures have been tested, including single measures without local controls, single measures with local controls, combination measures without local controls, combination measures with local controls. The magnitude of coefficients on the similarity measures generally increases and frequently become statistically significant (if not already so) when distance and same state controls are not included. As expected, correlations with distance and same state tend to make the coefficient magnitudes attenuate when distance and same state are controlled for

Notes: The table above reviews some robustness checks to the main empirical results presented in this paper.

Relative to the urban quality of life literature, this paper expands the flexibility of individual city demand so that varying preferences for amenities form aggregate city substitution patterns. Understanding these substitution patterns has relevant implications for a variety of practical matters. For example, supply or demand shocks in one city may affect supply and demand in other cities if shocks are large enough and cities are sufficiently substitutable. Another example relates to housing price dynamics, where prices propagate over time and space according to city substitution patterns.

Similar intuition can be applied to substitution within metropolitan areas as well. Determining which neighborhoods are substitutes in demand may be informative for predicting how urban development in one area affects locally substitutable areas. Along with those applications discussed above, this seems like an area of useful future research.

Acknowledgments

I thank Markus Brunnermeier, Henry Farber, Maria Guadalupe, Lars Lefgren, Chris Mayer, Jesse Rothstein, Hyun Shin, Todd Sinai, Abigail Wozniak, two anonymous referees, and seminar participants at the Industrial Relations Section and the Bendheim Center for Finance for helpful conversations and suggestions. I also thank the National Science Foundation and the Industrial Relations Section for generous research funding.

Appendix A. Migration derivative with respect to wages

Assuming i.i.d. error terms in the city choice model presented in Section 3.2, the probability of an individual i living in city j at time t is just P_{ij} from Eq. (3). Thus, the expected fraction of city j residents moving to city k between time $t-1$ and t is:

$$M_{jkt} = E[Y_{it} = k | Y_{it-1} = j] = \frac{E[Y_{it-1} = j \cap Y_{it} = k]}{E[Y_{it-1} = j]} = \frac{\int P_{ij} P_{ik} g(\beta) d\beta}{\int P_{ij} g(\beta) d\beta} = \frac{\varepsilon^D_{jk}}{\lambda p_k}, \tag{A.1}$$

with the last equality coming from the expression for ε^D_{jk} in Eq. (4). Thus, gross migration from city j to city k should be higher between city pairs that have a higher cross-price elasticity of demand.

As discussed in Section 2.2, suppose that an exogenous labor demand shock idiosyncratically increases the demand for labor in city k; that is, firms' production in city k, $\pi(w_k, p_k; a_k)$, is suddenly higher at every wage and price. With perfectly competitive labor markets, wages rise in the short-run, causing workers to migrate towards city k. Wages and prices adjust according as this migration occurs according to the shapes of the utility and profit functions until a new equilibrium is reached.

Thus, the net migration in response to a labor demand shock that raises short-run wages in city k is given by differentiating (A.1) with respect to w_k:

Author's personal copy

$$\frac{\partial M_{jkt}}{\partial w_k} = \frac{\left(\int P_{ij}\right)\left(\int \frac{\partial (P_{ij}P_{ik})}{\partial w_k}\right) - \left(\int P_{ij}P_{ik}\right)\left(\int \frac{\partial P_{ij}}{\partial w_k}\right)}{\left(\int P_{ij}\right)^2}, \tag{A.2}$$

where $g(\beta)d\beta$ is removed from the integrals to simplify notation. The partial derivatives are calculated by differentiating Eq. (3) with respect to city k wages: $\partial P_{ik}/\partial w_k = \lambda P_{ik}(1-P_{ik}) > 0$ and $\partial P_{ij}/\partial w_k = -\lambda P_{ij}P_{ik} < 0$. Plugging these into Eq. (A.2) and simplifying yields:

$$\begin{aligned}\frac{\partial M_{jkt}}{\partial w_k} &= \frac{(\int P_{ij})(\int P_{ij}(\lambda P_{ik}(1-P_{ik}))) + P_{ik}(-\lambda P_{ij}P_{ik}) - (\int P_{ij}P_{ik})(\int P_{ij}P_{ik})}{(\int P_{ij})^2} \\ &= \lambda \frac{\int P_{ij}P_{ik}(1-2P_{ik})}{\int P_{ij}} + \lambda\left(\frac{\int P_{ij}P_{ik}}{\int P_{ij}}\right)^2 \\ &= \lambda\left(M_{jkt}^2 + M_{jkt}\left(1 - 2\frac{\int P_{ij}P_{ik}^2}{\int P_{ij}P_{ik}}\right)\right) \\ &= \left(\frac{\left(\varepsilon_{jk}^D\right)^2}{\lambda p_k^2} + \frac{\varepsilon_{jk}^D}{p_k}\left(1 - 2\frac{\int P_{ij}P_{ik}^2}{\int P_{ij}P_{ik}}\right)\right) > 0\end{aligned} \tag{A.3}$$

which is greater than zero from the middle line of (A.3). Assuming that $\int P_{ij}P_{ik}^2/\int P_{ij}P_{ik}$ is sufficiently small (a sufficient but not necessary condition for the term in parentheses to be positive is that $P_{ik} < 1/2$ for all individuals), differentiating $\partial M_{jkt}/\partial w_k$ with respect to ε_{jk}^D shows that the migration response to wages should be higher for city pairs with higher cross-price elasticities of demand:

$$\frac{\partial^2 M_{jkt}}{\partial \varepsilon_{jk}^D \partial w_{kt}} \sim \frac{2\varepsilon_{jk}^D + \lambda p_k}{\lambda p_k^2} > 0. \tag{A.4}$$

Appendix B. Home price correlation

In this section, I present a two-city simultaneous supply and demand equations model to show why home price correlation might be a useful price-based measure to identify cities which are substitutes in demand. The model distinguishes between two main factors that cause home price changes to be correlated across cities: correlated economic shocks and city demand substitution. City demand substitution can lead to dependence in home prices in a pair of cities even when their fundamental shocks are uncorrelated. In particular, I show that two cities will have more highly correlated price changes when the cross-price elasticity of demand for one with respect to the other is larger.

I begin with a set of cities in equilibrium defined by individuals' utility functions, $u_{ij}(w_j, p_j; a_j)$, and firms' production functions, $\pi_j(w_j, p_j; a_j)$. As before, I assume that home prices represent capitalized future rents. Equilibrium is defined by the set of wages, prices, and quantities that clear the labor and production markets.

Consider two cities, city A and city B, initially at equilibrium demand given by $\overline{D}_A = \int P_{iA}g(\beta)d\beta$ and $\overline{D}_B = \int P_{iB}g(\beta)d\beta$. As showed in Section 3, the partial derivatives of demand with respect to own home prices are negative and with respect to other cities' home prices are positive: $\partial D_A/\partial p_A = \alpha_A < 0$, $\partial D_A/\partial p_B = \delta_A > 0$, $\partial D_B/\partial p_A = \alpha_B > 0$, $\partial D_B/\partial p_B = \delta_B < 0$. For simplicity, I assume that these slopes are locally linear around equilibrium. I assume that housing supply, initially at equilibrium $\overline{S}_A$ and $\overline{S}_B$, is also locally linear based on a given housing supply elasticity: $\partial S_A/\partial p_A = \gamma_A > 0$, $\partial S_B/\partial p_B = \gamma_B > 0$. The result is a simultaneous equations system with housing demand and supply for two cities:

$$\begin{aligned} D_A &= \overline{D}_A + \alpha_A p_A + \delta_A p_B + v_A \\ D_B &= \overline{D}_B + \alpha_B p_A + \delta_B p_B + v_B \\ S_A &= \overline{S}_A + \gamma_A p_A + \eta_A \\ S_B &= \overline{S}_B + \gamma_B p_B + \eta_B \end{aligned}. \tag{B.1}$$

where v_A and v_B represent shocks to demand and η_A and η_B represent shocks to supply. Demand shocks may occur for a variety of reasons, including shocks to firms' production functions $\pi_j(w_j, p_j; a_j)$, investment shocks (*e.g.*, an asset price bubble), or other source. Supply shocks may occur from business cycles or other economic forces related to home building.

In theory, a shock to demand would affect market-clearing wages as well as prices, as in traditional Rosen–Roback and urban quality of life models. For simplicity, I model only home prices here. Yet one can imagine allowing wage flexibility into the model as well, where the amount of price adjustment on the demand side will depend on the shapes of the individuals' utility functions as well as the firms' production functions. In any case, the same qualitative results for this equilibrium will still hold, even if some amount of wage adjustment is occurring in the background.

Setting supply equal to demand in each city produces the equilibrium prices and quantities:

$$p_A^* = \frac{(\gamma_B - \alpha_B)(v_A - \eta_A) + \delta_A(v_B - \eta_B)}{(\gamma_A - \alpha_A)(\gamma_B - \alpha_B) - \delta_A\delta_B}, \tag{B.2}$$

and

$$\begin{aligned} q_A^* &= D_A^* = S_A^* \\ &= \frac{\gamma_A(\gamma_B - \alpha_B)v_A - (\alpha_A(\gamma_B - \alpha_B) + \delta_A\delta_B)\eta_A + \delta_B\gamma_A(v_B - \eta_B)}{(\gamma_A - \alpha_A)(\gamma_B - \alpha_B) - \delta_A\delta_B}, \end{aligned} \tag{B.3}$$

where p_B^* and q_B^* can be written analogously. If the cross-price elasticities are zero ($\delta_A = \delta_B = 0$), then Eqs. (B.2) and (B.3) simplify to the standard simultaneous equations equilibrium with one good: $p_j^* = (v_j - \eta_j)/(\gamma_j - \alpha_j)$ and $q_j^* = (\gamma_j v_j - \alpha_j \eta_j)/(\gamma_j - \alpha_j)$.

The equilibrium solution indicates how supply or demand shocks in any one city can affect equilibrium prices and quantities in both cities. In particular, the partial derivatives of Eqs. (B.2) and (B.3) with respect to v_B show how an exogenous demand shock in city B can affect the price and quantity of city A: $\partial p_j^*/\partial v_B = \delta_A/\xi$ and $\partial q_j^*/\partial v_B = \delta_A\gamma_A/\xi$, where $\xi = (\gamma_A - \alpha_A)(\gamma_B - \alpha_B) - \delta_A\delta_B$ is positive as long as the own price elasticities are larger in magnitude than the cross-price elasticities. If the cross-price elasticity of city A with respect to city B prices is positive, then a positive demand shock to city B will raise both prices and quantities in city A. City B's demand shock will have a bigger impact on city A if this parameter is large. If both cross-price elasticities are zero, then all cross-derivatives equal zero and the own-derivatives are equivalent to the comparative statics for a standard simultaneous equations model for a single good.

Because isolated shocks rarely occur in practice, I consider how equilibrium prices in the two cities are related over time in the presence of shocks to housing supply and demand. In order to simplify the analysis, I assume that both cities have identical parameters for elasticity of supply ($\gamma_A = \gamma_B = \gamma$), own-price elasticity of demand ($\alpha_A = \alpha_B = \alpha$), and cross price elasticity of demand ($\delta_A = \delta_B = \delta$). I assume that supply shocks and demand shocks may be correlated across cities, but I do not allow for the demand and supply shocks themselves to be correlated: $var(v_j) = \sigma_v^2$, $var(\eta_j) = \sigma_\eta^2$, $corr(v_A, v_B) = \rho^D$, $corr(\eta_A, \eta_B) = \rho^S$, and $corr(v_j, \eta_j) = 0$.

With these simplifications, home price correlation for a given city pair in terms of the structural parameters of the model are as follows:

$$\begin{aligned}\rho_{AB} &= corr(p_A^*, p_B^*) \\ &= \frac{2\delta(\gamma - \alpha)}{\left(\delta + (\gamma - \alpha)^2\right)} + \frac{\left(\rho^D\sigma_v^2 + \rho^S\sigma_\eta^2\right)}{\left(\sigma_v^2 + \sigma_\eta^2\right)}\frac{\left(\delta^2 + (\gamma - \alpha)^2\right)}{\left(\delta + (\gamma - \alpha)^2\right)},\end{aligned} \tag{B.4}$$

I separate Eq. (B.4) into the sum of two terms, the first representing the pure substitution effect and the second representing the two

Author's personal copy

cities having correlated supply and demand shocks. Clearly the total price correlation is the sum of both, but the comparative statics are easier to understand when each term is examined in isolation.

Case 1 – The cross-price elasticities are zero ($\delta = 0$):

$$\rho_{AB} = \frac{\left(\rho^D\sigma_\nu^2 + \rho^S\sigma_\eta^2\right)}{\left(\sigma_\nu^2 + \sigma_\eta^2\right)}. \tag{B.5}$$

Home price correlation for a given city pair is higher when either demand shocks or supply shocks have a higher correlation since $\partial\rho_{AB}/\partial\rho^D > 0$ and $\partial\rho_{AB}/\partial\rho^S > 0$. If supply shocks are uncorrelated ($\rho_S = 0$), then the price correlation is directly proportional to correlation in the demand shocks, and vice versa. Whether supply correlation or demand correlation is more important for determining overall home price correlation depends on the relative sizes of the shocks and ultimately is an empirical question.

Case 2 – The cross-price elasticities are positive ($\delta > 0$) and the shocks are uncorrelated ($\rho^D = \rho^S = 0$):

$$\rho_{AB} = \frac{2\delta(\gamma - \alpha)}{(\delta + (\gamma - \alpha)^2)} \quad \text{and} \quad \frac{\partial\rho_{AB}}{\partial\delta} = \frac{2\delta(\gamma - \alpha)^3}{(\delta + (\gamma - \alpha)^2)^2} > 0 \tag{B.6}$$

Eq. (B.6) shows how home price correlation can exist as a result of demand substitution, even when the fundamental economic shocks between cities are uncorrelated. Home price correlation is increasing in the substitutability parameter (δ) since $\alpha < 0$. The total derivative of Eq. (B.4) with respect to δ is also positive, but I do not show it here. However, the intuition should be clear: home price correlation is higher between city pairs that are better substitutes because price increases in one city increases demand for the other city.

In summary, this two city simultaneous supply and demand equations model demonstrates that home price correlation is the sum of two factors: correlated economic shocks and city demand substitution. The latter factor can induce home price correlation between two cities, even when their fundamental economic shocks are uncorrelated.

References

Bartik, T.A., 1991. Who benefits from state and local economic development policies? W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.
Bayer, P., McMillan R., Rueben, R., 2004. An equilibrium model of sorting in an urban housing market. Working Paper.
Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. Journal of Political Economy 115, 588–638.
Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.
Bishop, K.C., 2007. A dynamic model of location choice and hedonic valuation. Working Paper.
Blanchard, O.J., Katz, L.F., 1992. Regional evolutions. Brookings Papers on Economic Activity 1, 1–75.
Blomquist, G., 2006. Measuring quality of life. In: Arnott, R., McMillen, D. (Eds.), A Companion to Urban Economics. Wiley–Blackwell, Malden, pp. 480–501.
Blomquist, G., Berger, M., Hoehn, J., 1988. New estimates of the quality of life in urban areas. American Economic Review 78, 89–107.
Bound, J., Holzer, H.J., 2000. Demand shifts, population adjustments, and labor market outcomes during the 1980s. American Journal of Labor Economics 18, 20–53.
Brunnermeier, M., Julliard, C., 2008. Money illusion and housing frenzies. Review of Financial Studies 21, 135–180.
Case, K.E., Cook, L., 1989. The distributional effects of housing price booms: winners and losers in Boston, 1980–88. New England Economic Review, 3–12.
Case, K.E., Shiller, R.J., 1987. Prices of single-family homes since 1970: new indexes for four cities. New England Economic Review, 45–56.
Case, K.E., Shiller, R.J., 1989. The efficiency of the market for single-family homes. American Economic Review 79, 125–137.
Chen, Y., Rosenthal, S.S., 2008. Local amenities and life-cycle migration: do people move for jobs or fun? Journal of Urban Economics 64, 519–537.
Costa, D.L., Kahn, M.E., 2000. Power couples: changes in the locational choice of the college educated, 1940–1990. The Quarterly Journal of Economics, 1287–1315.
Davies, P.S., Greenwood, M.J., Li, H., 2001. A conditional logit approach to US state-to-state migration. Journals of Regional Science 41, 337–360.
Davidoff, T., 2006. Labor income, housing prices, and homeownership. Journal of Urban Economics 59, 209–235.
Englund, P., Hwang, M., Quigley, J.M., 2002. Hedging housing risk. Journal of Real Estate Finance and Economics 24, 167–200.
Flavin, M., Yamashita, T., 2002. Owner-occupied housing and the composition of the household portfolio. American Economic Review 92, 345–362.
Gabriel, S.A., Mattey, J.P., Wascher, W.L., 2003. Compensating differentials and evolution and the quality-of-life among U.S. states. Regional Science and Urban Economics 33, 619–649.
Gabriel, S., Rosenthal, S., 2004. Quality of the business environment versus quality of life: do firms and households like the same cities? Review of Economics and Statistics 86, 438–444.
Glaeser, E.L., Gyourko, J., 2007. Arbitrage in housing markets. Working Paper, 2007.
Glaeser, E.L., Gyourko, J., Saiz, A., 2008. Housing supply and housing bubbles. Journal of Urban Economics 64, 198–217.
Greenwood, M., Hunt, G., Rickman, D., Treyz, G., 1991. Migration, regional equilibrium, and the estimation of compensating differentials. American Economic Review 81, 1382–1390.
Graves, P., Knapp, T., 1988. Mobility behavior of the elderly. Journal of Urban Economics 24, 1–8.
Graves, P., Waldman, D., 1991. Multimarket amenity compensation and the behavior of the elderly. American Economic Review 81, 1374–1381.
Gyourko, J., Tracy, J., 1991. The structure of local public finance and the quality of life. Journal of Political Economy 99, 774–806.
Gyourko, J., Mayer, C., Sinai, T., 2006. Superstar cities. Working Paper, 2006.
Himmelberg, C., Mayer, C., Sinai, T., 2005. Assessing high house prices: bubbles, fundamentals and misperceptions. Journal of Economics Perspectives 19, 67–92.
Kahn, M., 1995. A revealed preference approach to ranking city quality of life. Journal of Urban Economics 38, 221–235.
Kennan, J., Walker, J.R., 2010a. The effect of expected income on individual migration decisions. Working Paper.
Kennan, J., Walker, J.R., 2010b. Wages, welfare benefits and immigration. Journal of Econometrics 156, 229–238.
Leung, C.K.Y., 2007. Equilibrium correlations of asset price and return. Journal of Real Estate Finance and Economics 34, 233–256.
Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics 47, 13–37.
Mayer, C., Sinai, T., 2007. Housing and behavioral finance. Working Paper.
McDuff, D., 2010. Local market shocks, home price risk, and index hedging. The Journal of Real Estate Finance and Economics. doi:10.1007/s11146-010-9255-2.
Mueser, P., Graves, P., 1995. Examining the role of economic opportunity and amenities in explaining population redistribution. Journal of Urban Economics 37, 176–200.
Nevo, A., 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. Journal of Economics and Management Strategy 9, 513–548.
Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. Econometrica 69, 307–342.
Notowidigdo, M.J., 2010. The incidence of local labor demand shocks. Working Paper.
Rappaport, J., 2004. Why are population flows so persistent? Journal of Urban Economics 56, 554–580.
Rappaport, J., 2007. Moving to nice weather. Regional Science and Urban Economics 37, 375–398.
Rosen, S., 1979. Wage-based indexes of urban quality of life. In: Mieszkowski, P., Straszheim, M. (Eds.), Current Issues in Urban Economics. Johns Hopkins University Press, Baltimore.
Roback, J., 1982. Wages, rents, and the quality of life. The Journal of Political Economy 90, 1257–1278.
Sharpe, W., 1963. A simplified model for portfolio analysis. Management Science 9, 277–293.
Shiller, R., 2007. Historic turnings points in real estate. Cowles Foundation Discussion Paper.
Saiz, A., 2010. The Geographic Determinants of Housing Supply 125, 1253–1296.
Sinai, T., Souleles, N.S., 2005. Owner-occupied housing as a hedge against rent risk. Quarterly Journal of Economics 120, 763–789.
Sinai, T., Souleles, N.S., 2009. Can owning a home hedge the risk of moving? Working Paper.
Walker, J.R., 2006. Economic perspectives on family and migration. Working Paper.
Whisler, R.L., Waldorf, B.S., Mulligan, G.F., Plane, D.A., 2008. Quality of life and the migration of the college-education: a life-course approach. Growth and Change 39, 58–94.
Wozniak, A., 2006. Why are college graduates more responsive to distant labor market opportunities? Working Paper.