Housing Prices and Inter-urban Migration

Andrew J. Plantinga*[a]

Cécile Detang-Dessendre[b]

Gary L. Hunt[c]

and

Virginie Piguet[b]

*Corresponding author

[a]Department of Agricultural and Resource Economics, 232B Ballard Extension Hall, Oregon State University, Corvallis, OR, 97330, USA. Email: plantinga@oregonstate.edu. Phone: (011) 1-541-737-1423. Fax: (011) 1-541-737-1441.

[b]INRA, Agrosup Dijon, UMR 1041 CESAER, 26 bd docteur Petitjean, BP87999, Dijon, F-21000, France. Email: cecile.detang-dessendre@dijon.inra.fr; virginie.piguet@dijon.inra.fr.

[c]School of Economics, 205C Stevens Hall, University of Maine, Orono, ME, 04469, USA. Email: Gary.Hunt@umit.maine.edu

Housing Prices and Inter-urban Migration

Abstract. Economic theory predicts that individual migration decisions for working-age adults will depend on area differences in wages, housing costs, and amenities. While the importance of wages and amenities is well-established from previous empirical studies, evidence regarding housing costs is far less conclusive. We develop and test a new method for representing housing prices in migration analyses. We first provide conditions under which utility-maximizing housing costs can be specified as a function of individual characteristics, similar to a Mincerian wage equation. Using large samples of individuals from the 2000 PUMS, we estimate the relationship between housing costs and individual attributes for each of 291 metropolitan areas in the U.S. Our approach accounts for rental and ownership decisions, the costs of rental and owned properties, and the costs of holding housing capital. We test our housing cost measure using observations of point-to-point migration decisions for a large sample of college-educated males. Our migration model includes additional controls for the wage each individual expects to earn in each area as well as a large set of area amenities. Our key finding is that our proposed housing cost measure yields the expected results (higher housing prices reduce the probability that an area is selected). We re-estimate the model using three alternative metropolitan area measures of housing costs: median house price, average apartment rent, and average urban land rent. These measures consistently produce counterintuitive positive effects of housing costs on area choice.

Classification Codes. R23 - Regional Migration; Regional Labor Markets; Population; Neighborhood Characteristics. R31 - Housing Supply and Markets.

Housing Prices and Inter-urban Migration

1. Introduction

Understanding the causes and consequences of human migration has long been of interest to urban and regional economists. Roback (1982) explained the equilibrium distribution of human population by differences in the non-traded amenities at each location. These amenity differences produce wage and rent differentials that, in equilibrium, leave households and firms indifferent to changing locations. Mueser and Graves (1995) modify the Roback model by making instantaneous adjustment to equilibrium costly for households and firms. Migration emerges in their model as a short-run response to disequilibrium in labor and housing markets. Absent any shocks to exogenous factors such as preferences and technology, the sequence of short-run equilibria in these markets converges to the Roback equilibrium in the long run.[1]

Empirical studies build on these theoretical results by estimating the effects of wages, housing prices, and amenities on migration. Findings with respect to amenities are clear. Area measures of population and migration as well as household location decisions are significantly related to climate (Mueser and Graves, 1995; Clark and Murphy, 1996; Hunt and Mueller, 2004; Cheshire and Magrini, 2006; Rappaport, 2007; Poston et al., 2009; Eichman et al., 2010), air quality (Seig et al., 2004; Bayer et al., 2009), recreational opportunities (Duffy-Deno, 1998; Lewis et al., 2002), cultural amenities (Clark and Hunter, 1992), and crime rates (Gottlieb and Joseph, 2006). Housing prices and wages are endogenous to area-level migration (Mueser and Graves, 1995), and so these variables are typically excluded from reduced-form analyses with aggregate data, or modeled as jointly determined in a simultaneous equations framework (Jeanty et al., 2010). In migration analyses with individual data, it is reasonable to treat households as

[1] If the system is in long-run equilibrium, then spatial variations in wages and housing prices would have no effect on migration decisions. The finding of significant effects of wages and housing prices suggests that the system is out of equilibrium (Greenwood et al., 1991).

price-takers in labor and housing markets and, thus, the effects of wages and housing prices on migration decisions can be measured. In such studies, higher relative wages or income are consistently found to increase the probability that a household will select a given location, all else equal (Berger and Blomquist, 1992; Davies et al., 2001; So et al., 2001; Hunt and Mueller, 2004; Bayer et al., 2009; Bishop, 2008; Kennan and Walker, 2011; Dahl and Sorenson, 2010). In contrast, the results for housing prices are much less clear. Studies that include area-level measures (e.g., median housing price for a county or metropolitan area) find a mix of negative, positive, and insignificant effects on migration decisions (Berger and Blomquist, 1992; Hunt and Mueller, 2004; Gottlieb and Joseph, 2006; Bishop, 2008).[2] Other studies do not control for housing prices or do not explicitly measure their effects (Davies et al., 2001; Bayer et al., 2009; Detang-Dessendre et al., 2008; Kennan and Walker, 2011; Dahl and Sorenson, 2010).[3]

The purpose of this paper is to investigate the determinants of inter-urban migration decisions, with a particular emphasis on the role of housing costs. An important contribution of the study is the development of a new method for representing housing prices in individual-level migration analyses.[4] Our proposed approach is inspired by the Mincerian wage equation commonly used to model wages in studies of household migration (e.g., Hunt and Mueller, 2004; Bayer et al., 2009). In the case of wages, a reduced-form wage equation is estimated for each area using observations of wage rates and corresponding individual and job characteristics, such as age, education, and weekly hours worked. These equations are then used to predict the wage

[2] It is important to distinguish between inter-area migration and intra-area location changes. While housing prices clearly matter for moves in both cases, we are primarily interested in their effect on migration at the scale of metropolitan areas, counties, and states. Intra-area studies that examine effects of housing prices on household location decisions include Chan (2001), So et al. (2001), Engelhardt (2003), Seig et al. (2004), and Ferreira et al. (2010).

[3] Chen and Rosenthal (2008) construct area-level quality of life indices that reflect wages and housing prices. They investigate how changes in these indices for migrants are influenced by individual-level factors such as age and gender.

[4] See Jeanty et al. (2010) for a careful treatment of population change and housing values in an analysis with aggregate data.

an individual would earn in unselected areas conditional on these attributes. We identify the form of the utility function under which a similar approach can be used to predict individual-specific housing expenditures for each area using individual characteristics. Our theoretical results guide the development of an empirical measure of housing costs that accounts for the decision to own or rent and the cost of holding housing capital.

We test our housing cost measure using data from the 5% sample of the 2000 Public Use Microdata Survey (PUMS) that identify point-to-point migration decisions for a large sample of individuals residing in 291 U.S. metropolitan areas. We estimate nested logit models of metropolitan area choice, controlling for wages, a large range of amenities, and expected housing costs. Our proposed method for measuring housing costs yields the expected result that, all else equal, higher housing costs reduce the probability that a metropolitan area will be chosen. This finding is robust to alternative specifications and samples. We then re-estimate our model using three alternative metropolitan area measures of housing costs: median house price, average 2-bedroom apartment rent, and average per-acre urban land rent. We find that these aggregate measures consistently produce counter-intuitive positive effects of housing costs on metropolitan area choice. Potential migrants are likely to base decisions on the costs of housing that they themselves would select, rather than what the average metropolitan area resident would choose, implying measurement error in the metropolitan area housing cost measures. Correlation between this measurement error and unobservable area attributes is a likely source of bias in the coefficient estimates on housing costs. In contrast, our proposed measure is a projection of individual-level housing costs into individual attributes. It is constructed in a way that controls for any unobservable area-level influences on migration decisions transmitted through housing markets.

The next section presents a model of household migration that provides the theoretical underpinnings for our housing cost measure and empirical analysis. Section 3 describes the data we use and the specification of choice sets for households. In section 4, we discuss the estimation procedures for the wage equations, housing cost measures, and nested logit models of metropolitan area choice. Section 5 presents a simulation analysis in which we use our estimated models to predict how changes in age affect metropolitan area populations. The simulations account for how age increases affect expected wages, rental and ownership decisions, expected housing costs, and the propensity of individuals to leave their origin location. Discussion and conclusions are provided in a final section.

2. Theory

Individuals are assumed to choose locations conditional on expected wages and housing costs, the amenities of the area, and costs associated with moving. For individual *i*, the utility in the area *j* is specified:

(1) $$U_{ij} = U(X_{ij}, z_{ij}; A_j, C_i)$$

where X_{ij} is a vector of housing attributes and z_{ij} is a composite numeraire good. These are choice variables for individual *i*. The utility derived by the *i*th individual in the *j*th area also depends on the individual's characteristics (age, gender, etc.), denoted by the vector C_i, and the amenities in area *j*, denoted by the vector A_j. If area *j* differs from the starting location, then A_j includes measures of the dis-amenities associated with moving (e.g., moving costs). Conditional on choosing area *j*, the individual maximizes utility subject to the budget constraint:

(2) $$P_j X_{ij} + z_{ij} = I_{ij}$$

where P_j is a vector of implicit prices for housing attributes and I_{ij} is the income that individual *i* expects to earn in area *j*. Individuals are assumed to be price takers in housing, labor, and goods markets. For simplicity, we assume that the cost of housing is a linear function of housing attributes and that the price of the composite good is constant across areas.

In the migration problem, an individual will choose the area that gives the highest utility. Thus, we must solve the utility maximization problem for each area to find the indirect utility function V_{ij}. We assume a quasi-linear utility function of the following form:

(3) $$U_{ij} = u(X_{ij}; A_j, C_i) + z_{ij}$$

where $u(\cdot)$ has the standard properties of a utility function. This specification assumes additive separability between the numeraire good and goods associated with the migration choice (housing attributes and area amenities). This specification permits individuals to choose different housing bundles in different areas and to make trade-offs among these attributes. However, it restricts individuals from trading off housing attributes and the numeraire good, which by construction gives constant marginal utility. The level of the numeraire good can vary by individual and area. We adopt this specification because it will allow us to specify housing expenditures as a reduced-form function of individual attributes, as we now show.

The solution to the utility maximization problem gives the demands $X^*(P_j; A_j, C_i)$ and $z^*(P_j, I_{ij}; A_j, C_i)$. With positive consumption of the numeraire good, which we assume here, an individual allocates a portion of their income to housing and any remaining income is spent on the numeraire good; i.e., $z^*(P_j, I_{ij}; A_j, C_i) = I_{ij} - P_j X^*(P_j; A_j, C_i)$. Demands for housing attributes do not depend directly on income. This does not imply that individuals spend the same amount of income on housing in every location. The prices of housing attributes and amenities vary by location and this affects the composition and cost of the chosen housing bundle. Individuals also must satisfy their budget constraints in every location. Because an individual's income differs by location, a housing bundle that would be purchased in one location may not be affordable in other locations.

Housing expenditures can be written as an area-specific function of an individual's characteristics:

$$
\begin{aligned}
H_{ij}^{*} &= P_j X^{*}(P_j; A_j, C_i) \\
&= H_j(C_i)
\end{aligned} \tag{4}
$$

where the form of the function $H_j(\cdot)$ varies by location *j* due to area differences in the implicit prices of housing attributes and amenities. Following the hedonic wage literature, income can also be written as a function of an individual's characteristics; that is, $I_{ij}^{*} = I_j(C_i)$. The form of the income function $I_j(\cdot)$ varies by area because of differences in industrial composition, transportation costs, amenities, and other factors. The similarity between the housing and income functions is due to the fact that both housing and jobs are differentiated goods. Buyers of housing obtain a bundle of housing attributes in the same way that workers buy (at negative price) bundles of job characteristics, including working conditions, opportunities for advancement, and so on.[5] Thus, in our model, migrants make choices among locations conditional on optimal expenditures on housing and job attributes in each area.

Substitution of $X^{*}(P_j; A_j, C_i)$ and $z^{*}(P_j, I_{ij}; A_j, C_i)$ into the utility function (3) gives the indirect utility function:

---

[5] Unlike the housing market in which sellers of housing are presumed to be indifferent to the attributes of buyers, employers clearly care about the characteristics of workers. Thus, wages are also determined by the interaction of supply and demand for worker characteristics, implying that a correctly-specified hedonic wage equation must include job characteristics (Rosen, 1979), which we subsume in C_i.

(5)
$$\begin{aligned} V_{ij} &= U(X^*(P_j; A_j, C_i)) + I_{ij}^* - H_{ij}^* \\ &= V(H_j(C_i), I_j(C_i), P_j, A_j, C_i) \end{aligned}$$

Equation (5) is the theoretical basis for the empirical model of migration. An increase in housing expenditures in a given area, H_{ij}^*, reduces utility because it reduces consumption of the numeraire good. Conversely, an increase in area income, I_{ij}^*, increases consumption of the numeraire good, thereby raising utility. Individuals who choose area *j* also obtain utility from that area's amenities A_j. The prices of housing attributes, P_j, enter into (5) because they affect the composition of the housing bundle. In the empirical application below, we do not observe these price vectors in every area, and so they are excluded from the model. As such, housing expenditures H_{ij}^* become a control for the "price" of the composite housing good chosen by individual *i* in area *j*.

The specification of housing expenditures in (4) has a clear advantage over metropolitan area variables used in previous studies, such as median house price. Such measures implicitly assume that all individuals within the metropolitan area consume the same housing bundle, while (4) allows the composition of housing attributes to vary among individuals. Accordingly, the migration decision is dependent on the cost of the housing bundle that an individual would optimally select, rather than the cost of the average housing bundle in an area. In general, we would expect a given individual to choose different housing bundles in different areas. But, this presents a puzzle: why would an individual ever move to an area with high housing costs? Given downward-sloping demands for housing attributes, a small amount of housing would be optimally chosen in an area with high implicit prices. Yet, it is clear from (4) that housing expenditures in a high-price and a low-price area could be the same (high prices and low quantities in the first area, low prices and high quantities in the second). Given that an individual gets higher utility from consuming more housing, why would the high-price area ever be chosen? The puzzle is solved by recognizing that the higher prices must be due to better amenities. These amenities raise the utility obtained in the high-price area—recall that A_j is an argument in (5)—and compensate the individual for accepting less housing at higher prices. The implication for our empirical analysis is that the expected negative relationship between utility and housing expenditures will be identified only if we properly control for area amenities (Hunt, 1993).

3. Data and Choice Set Specification

Our main data source is the 5% sample of the 2000 Public Use Microdata Survey (PUMS), which includes approximately 14 million U.S. individuals (Ruggles et al., 2010). The survey provides a large number of demographic and socioeconomic variables, including measures of age, income, employment, and educational attainment (Table 1). Residence in 2000 is reported at the level of the Public Use Microdata Area (PUMA). PUMAs are geographic areas, designated by the Bureau of the Census, that contain at least 100,000 people. Respondents to the PUMS are also asked about their residence in 1995. If an individual's residence changed, the former residence is reported at the level of the Migration PUMA (MIGPUMA). MIGPUMAs are agglomerations of one or more PUMAs. In the 2000 PUMS, there are 2,101 PUMAs and 1,050 MIGPUMAs. The PUMS allows us to model point-to-point migration decisions between 1995 and 2000. While the PUMS provides observations of migration decisions at the level of PUMAs and MIGPUMAs, we model migration between metropolitan areas (MAs) for reasons discussed below.

Our empirical analysis involves the estimation of wage, housing cost, and migration models. For each component, we use a different sample extracted from the PUMS. Data on all

MA residents in the U.S. (approximately 5.4 million observations) are used to estimate wage equations for each MA. MA-specific housing cost equations are estimated with data on all household heads residing in MAs (approximately 3.8 million observations) and, finally, our basic migration model is estimated with a smaller sample of working-age, college-educated male MA residents (24,604 observations). The migration sample represents a population of approximately 13.5 million individuals.

We focus the migration analysis on a sub-population of MA residents for several reasons. Given our emphasis above on wages as a key determinant of migration decisions, we consider only non-institutionalized adults of working age (25-64 years). Results from Hunt and Mueller (2002) indicate that labor market relationships differ for males and females and so we limit our attention to male migrants and leave female migrants for future research. Further, preliminary estimation revealed that pooling samples of individuals with different levels of educational attainment was not justified and so we estimated separate models by educational level. We present results, below, on individuals with a 4-year college degree or higher, the most mobile segment of the working-age population. As a check on the generality of our findings, we discuss estimation results done with samples of male migrants with an educational attainment of some college.

Given the emphasis on working-age adults, it is necessary that our set of origins and destinations conform to distinct labor markets. MAs are likely to satisfy this criterion because they are delineated so that the communities within them exhibit a high degree of social and economic integration. Although we would wish to include non-MA residents, it is difficult to identify distinct labor markets in non-MA areas. There is no counterpart that we know of to the MA defined for non-MA areas. Even if this were available, it would need to match reasonably well the geographical scale of MIGPUMAs. For non-MA areas, MIGPUMAs are frequently too large to reasonably correspond to labor market areas.[6] Our sample is thus restricted to MA residents, defined as individuals who lived in an MA in 1995 and 2000.[7]

Area attributes were developed for the final set of 291 MAs. The main data source is the State and Metropolitan Area Data Book 1997-98 (U.S. Bureau of the Census, 1998), which provides observations of demographic, social, and economic variables for all MAs and years ranging from 1990 to 1997. We use lagged area measures (as close to 1995 as possible) to explain migration decisions occurring between 1995 and 2000. Additional measures are constructed using county data from the 1990 U.S. Census and McGranahan (1999). Because MAs are agglomerations of counties, we can compute MA averages using county-level observations. Table 2 provides a list of and sources for the area measures that were developed.

4. Methods

Wage Equation Estimation

The PUMS provides information on the wages earned by individuals in 2000. Of course, we observe wage only for the location where an individual lived and worked. Estimates of wages in unselected MAs are needed for estimation of the migration model. To this end, we estimate a log-linear wage equation for each MA using data on all non-institutionalized working-age individuals in the PUMS who resided there in 2000. We use observations of all individuals

---

[6] For example, a single MIGPUMA in eastern Oregon encompasses an area greater than the combined area of Connecticut, Rhode Island, Massachusetts, New Hampshire, and Vermont.

[7] In 2000, there were 324 MAs in the U.S. This figure includes 251 metropolitan statistical areas (MSAs), 12 New England Consolidated Metropolitan Areas (NECMAs), and 61 Primary Metropolitan Statistical Areas (PMSAs). We matched each of the 324 MAs to one or more MIGPUMAs. If a portion of a MIGPUMA lay outside of an MA boundary, we retained the MIGPUMA only if at least 75% of its population lived within the MA. In the case of 24 MAs, no matches to MIGPUMAs could be made using this criterion. These MAs had relatively small populations (on average, approximately 126,000 persons) and were dropped from the analysis. We also excluded eight MAs in Alaska, Hawaii, and Puerto Rico, and one MA (Auburn-Opelika) with missing data, leaving us with 291 MAs, comprised of 576 MIGPUMAs.

in an MA, not just working-age, college-educated males, to increase variation in the data.[8] On average, we have approximately 13,500 observations per MA. The dependent variable is the natural log of average weekly salary wages in 2000 and independent variables are a vector of individual and job attributes (C_i in section 2) that includes gender, age, race, marital and family status, language, educational attainment, usual hours worked, type of position, and sector of employment (Table 1).[9] We dropped individuals in the top and bottom 1% of the wage distribution to reduce the influence of very high and very low wages on our estimates. We find that dropping these extreme values reduces skewness and excess kurtosis in the distribution of predicted errors, thus yielding a better approximation of a normal distribution.

The MA-specific wage equations are used to estimate the (unlogged) wage each individual would earn in each MA, conditional on the individual's gender, age, race, etc. These estimates are denoted $Wage_{ij}$. Ideally, we would have estimated the wage equations with 1995 data, so that wage predictions are lagged with respect to the migration decision. However, PUMS data are available in either 1990 or 2000. If the parameters of the wage equation are not changing appreciably over time, then either data set can be used. If they are changing, then the 1990 data have the disadvantage that the wage prediction and the migration decision are separated by 5 to 10 years. The 2000 data are preferable in this respect (the separation is between 0 and 5 years), but have the shortcoming that some of the information may have been unobservable to potential migrants in 1995. A further consideration is that definitions of PUMAs and MIGPUMAs differ somewhat between the 1990 and 2000 PUMS. This complicates

---

[8] All estimations in this study are weighted using the person weight variable in the PUMS, denoted *Weight*.

[9] There is a large literature examining the potential endogenous relationship between wages and educational attainment. We do not make adjustments for this since, according to results of a meta-analysis by Ashenfelter et al. (1999), alternative estimators (e.g,, instrumental variables) tend to produce similar estimates.

the use of wage data from the 1990 PUMS and the migration data from the 2000 PUMS, and partially explains our decision to estimate the wage equations with 2000 data.

Each wage equation includes a constant term that is an important part of the identification strategy employed in the migration analysis. Because we estimate separate wage equations for each MA, the constant term in each equation captures the influence of any common area effects on wages, including those arising from amenity differences among MAs (Blomquist et al., 1988). These constant terms are included when we form the wage estimates ($Wage_{ij}$) for each individual and area. As discussed below, these constant terms act as fixed effects in the migration model, controlling for any unobservable area-level influences on migration decisions transmitted through labor markets.

Housing cost estimation

A similar approach is used to predict the cost of housing for each individual in each MA. In developing the housing cost measure, the first issue to contend with is the choice an individual makes between renting and owning. We allow the likelihood that an individual owns or rents to differ by area because, for example, the same individual may rent an apartment if they live in New York City, but buy a house if they live in Miami, Florida. We assume that the probability of ownership depends on individual attributes. This formulation implicitly accounts for the role of income in influencing home ownership since, as in section 2, income is a function of individual attributes. The PUMS includes a variable indicating whether a household head lives in a rented (*Renter*) or owned home (*Owner*). Homes acquired with a mortgage or other lending arrangements are classified as owned. Using the full sample of household heads for each area (on average, about 9,500 observations per MA), we estimate probit models for the binary ownership decision. This yields area-specific functions for the probability of ownership that depends on individual attributes. These functions are used to estimate the ownership probability, denoted π_{ij}, for each individual *i* and area *j*.

Separate housing price equations are then estimated for rented and owned homes.[10] The PUMS indicates the monthly rent paid by household heads or, for owners, the value of their housing unit. We multiply the monthly rent by 12 to obtain the annual rent. The logs of the annual rent (*Annualrent*) and owned home value (*Ownervalue*) variables are regressed on the corresponding set of individual attributes using the full sample of household heads in each area. As with the wage and ownership analysis, this yields functions that are used to estimate annual values of rented and owned homes for each individual and each area. These estimates are denoted $Annualrent_{ij}$ and $Ownervalue_{ij}$. The housing price models include separate constant terms for each MA that capture the influence of any common area effects on housing costs.

The final step is to derive an annual housing cost measure (H_{ij}^* in section 2). To do so, we must express the value of owned housing as an annualized equivalent, which requires an estimate of the cost of holding a unit of housing capital. Following the Jorgensonian cost of capital formulation, we specify this as the financial cost of holding housing capital less the rate of housing price appreciation.[11] The first term is approximated with the January 1, 2000 rate of return on 3-month Treasury bills (5.33%) and the second is estimated as the average annual percentage change in the metropolitan area median house price between 1990 and 2000 (*House value change*). The difference between these two terms gives a metropolitan area-specific capital cost r_j. Our annual housing cost measure is, thus,

[10] We do not distinguish between apartments, single detached houses, etc. Thus, all types of housing may be included in rented and owned homes.
[11] We assume that marginal tax rates are constant across individuals and areas and that there are no investment tax credits or depreciation allowances.

(6) $$Housingcost_{ij} = \pi_{ij} r_j Ownervalue_{ij} + \left(1 - \pi_{ij}\right) Annualrent_{ij}$$

Consistent with the theoretical development in section 2, $Housingcost_{ij}$ is estimated for each area and individual using individual attributes and functions specific to each MA. As with wages, the housing cost estimates include constant terms that measure any unobservable area-level effects transmitted through housing markets.

Migration decisions

Following Hunt and Mueller (2004), we estimate a nested logit model of migration over the period 1995 to 2000. Individuals decide whether to remain in the same location (the MA where they lived in 1995) or move to a new MA. Conditional on moving, the individual selects an MA and to maximize utility. According to (5), the maximum utility from location *j* depends on expected wage ($Wage_{ij}$), expected housing cost ($Housingcost_{ij}$), individual characteristics (C_i), and area attributes (A_j). We specify the utility that individual *i* obtains from the *j*th MA as:

(7) $$V_{ij} = \begin{cases} \alpha_0^{'} C_i + \beta^{'} A_j + \gamma Wage_{ij} + \delta Housingcost_{ij} + \varepsilon_{ij} & j = 0 \\ \alpha_1^{'} C_i + \beta^{'} A_j + \gamma Wage_{ij} + \delta Housingcost_{ij} + \varepsilon_{ij} & j = 1, 2, \ldots \end{cases}$$

where *j*=0 indicates the individual's origin (the 1995 MA), $\alpha_0, \alpha_1, \beta, \gamma, \delta$ are conformable parameter vectors, and ε_{ij} is a random disturbance with a type I extreme value distribution.[12]

Several remarks on the specification in (7) are in order. First, only variables that differ among MAs can explain area choice. As such, we use the individual attributes in C_i only to explain the decision to stay or move. The parameters on these variables (α_0, α_1) must differ to capture utility differences associated with staying or moving. Second, the area attributes in A_j

[12] We do not account for the possibility that the disturbance terms are spatially correlated. Estimation of spatial discrete-choice models is challenging, especially with large datasets like the one we analyze. An alternative would be to include random effects for MAs. In this case, however, simulation bias becomes a concern because we are modeling choices (i.e., the selection of a particular MA) that are associated with very low probabilities.

differ across MAs but not across individuals. If one thought that the marginal utility of a given attribute is different among individuals, one approach would be to interact the area attributes with individual characteristics. For example, one might interact an MA-level measure of cultural amenities with an individual-level measure of educational attainment. Alternatively, one can accomplish the same result by estimating models for selected cohorts of individuals, the approach that we pursue below. Finally, the wage and housing cost variables differ by both MAs and individuals. As noted above, we include area-specific constants when forming these estimates, which is an important part of our strategy for identifying the parameters of the migration model. In (7), these constants act as MA-level fixed effects, controlling for any unobserved area attributes whose effects are transmitted through labor and housing markets.[13]

Following Train (2003), the probability that individual *i* chooses to stay (*m*=0) or move (*m*=1) is given by:

(8)
$$P_{im} = \frac{e^{\alpha_m C_i + \lambda_m I_{im}}}{\sum_{m=0}^{1} e^{\alpha_m C_i + \lambda_m I_{im}}}$$

where λ_m measures the degree of substitutability among alternatives under choice *m*,

(9)
$$I_{i0} = (\beta' A_0 + \gamma Wage_{i0} + \delta Housingcost_{i0}) / \lambda_0$$
$$I_{i1} = \ln \sum_{j=1}^{J} e^{(\beta' A_j + \gamma Wage_{ij} + \delta Housingcost_{ij}) / \lambda_1}$$

[13] The inclusion of the constant terms also ensures that $\beta' A_j$ measures the total contribute of the area attributes to utility. Suppose that a local amenity provides the utility-equivalent of $50 and causes a downward adjustment in wages equal to $20. If we neglect the compensating differential (e.g., by omitting the intercept term when we calculate expected wage), then utility implicitly rises by the utility-equivalent of $20. The term for this amenity in (7) would then add only the utility-equivalent of $30. In contrast, if we control for the compensating differential by including the intercept term, the amenity term in (7) adds the utility-equivalent of $50, the total contribution of the amenity to utility.

and J is the total number of MAs. The terms $\lambda_0 I_{i0}$ and $\lambda_1 I_{i1}$ are equal to the expected utility that the individual obtains from the choices in the stay and move nests and, thus, provide the link between the stay/move and area choice decisions. Conditional on moving, the probability that the ith individual selects the jth MA equals:

$$P_{ij|m=1} = \frac{e^{(\beta' A_j + \gamma Wage_{ij} + \delta Housingcost_{ij})/\lambda_1}}{\sum_{j=1}^{J} e^{(\beta' A_j + \gamma Wage_{ij} + \delta Housingcost_{ij})/\lambda_1}} \tag{10}$$

whereas the probability that the individual selects the origin, conditional on staying, is one (i.e., $P_{i0|m=0} = 1$). This model is a case of the partially degenerated nested logit model analyzed in detail by Hunt (2000), which requires the normalizations $\lambda_0 = 1$ and $\alpha_0 = 0$ for estimation.[14]

Because of the large number of individuals and alternatives, two steps are taken to reduce the size of the estimation problem, thereby speeding up the estimation of the migration model. The PUMS provides approximately 1.96 million observations of working-age males. First, we randomly select 15,000 individuals who remained in the same location and 45,000 individuals who changed locations. Movers are over-sampled in order to retain information about the large number of potential point-to-point location changes. Sampling weights are adjusted so that a representative sample is maintained. From the 60,000 individuals, we construct a sample of 24,604 highly-educated individuals (4,276 stayers, 20,238 movers) ensuring that this sub-sample matches the demographic characteristics (in terms of race, age, and marital status) of the national sample of working-age, college-educated males. Second, we limit the number of alternatives in the choice set. We include the origin and the selected MA (if different from the origin) and then randomly sample from the unselected MAs to bring the total size of the choice set to 100. This

[14] This specification imposes the Independence from Irrelevant Alternatives (IIA) property on destinations (i.e., choices within the move nest) but not between the origin and any given destination. See Train (2003) for more discussion of the IIA property.

procedure has been shown to give consistent estimates of the parameters for the model with the full choice set (Ben-Akiva and Lerman, 1985).

The individual characteristics used to explain the stay or move decision include *Age*, *Married*, *White*, and *Black* (the omitted categories are separated, divorced, single, and other race) (Table 1). The constant term in the upper nest captures any net fixed costs associated with moving. In addition to the *Wage* and *Housingcost* variables, area choice is assumed to depend on migration costs and MA-level amenities (definitions and data sources are given in Table 2). Migration costs are measured with the variables *Distance*, equal to the radial distance between the centroids of the most populous county in the origin and destination MAs, and *Population Spread*, equal to the absolute value of the logged differences in population between the origin and destination MA. If we suppose that an individual's origin MA reveals their preference with respect to MA size, then *Population Spread* controls for the cost in utility terms associated with migration to larger or smaller MAs. MA-level amenities include four climate variables (*January temperature*, *July temperature*, *July humidity*, and *Precipitation*), two topography variables (*Mountain, Plainshills*), variables for proximity to major water bodies (*Gulf Coast, Great Lakes*, *North Atlantic*, *Pacific*, and *South Atlantic*), and variables measuring air quality (*Ozone*) [15], crime (*Violent crime*), and economic opportunity (*Employment growth*). We hypothesize that, all else equal, individuals prefer warmer winters, cooler and drier summers, and less annual precipitation. We expected positive signs on the topography variables, indicating that migrants prefer MAs with hills and mountains, and on the coastal variables.[16] Poor air quality and crime

[15] Bayer et al. (2009) estimate a migration model with MA-level fixed effects and then regress the estimated effects on MA-level covariates, treating air quality (measured as particulate matter concentrations) as an endogenous regressor. We attempt to mitigate potential endogeneity problems associated with the ozone measure by controlling for a broad set of MA-level climatic and topographical characteristics as well as including regional indicators for proximity to major water bodies.

[16] The omitted topography variable is the proportion of the MA land area classified as plains or tablelands. The omitted coastal variable is an indicator variable for MAs that share no border with a major water body.

are expected to lower utility, whereas lagged MA employment growth is expected to increase the attractiveness of an area.

We estimate four versions of the migration model that differ according to the housing price variable included. Version I includes our proposed measure, *Housingcost*. For version II, we use the median value of owner-occupied housing in 1990 (*Median house value*). This variable is collected as part of the Census of Population and Housing and is based on respondents' estimates of how much their properties would sell for were they to be offered for sale. Similar measures have been used in a number of previous migration studies (Clark and Hunter 1992, Bishop 2008, Scott 2010). Version III includes the 40th percentile monthly rental rate for a 2-bedroom apartment in 1995 (*Apartment rent*). These data are collected by the Department of Housing and Urban Development and used to determine federal housing assistance payments. Unlike the median house value, this variable measures the rent for a standardized housing unit. As well, it is based on a combination of self-reported census and telephone survey data. Finally, version IV uses a measure of urban land rent developed by Lubowski (2002) (*Land rent*). This variable was constructed by subtracting the value of structures from county measures of single-family housing prices. This residual is a proxy for the average price of land in each county used to build residential housing. *Land rent* is expressed as the present discounted value of an infinite stream of annual land rents per acre. All versions of the migration model use MA-level measures of the housing price. Because the median house value, apartment rent, and land rent variables are all measured at the county level, we average them to form MA-level variables.

Summary statistics are presented in Table 3 for the variables used in the migration analysis. The average age of working-age, college-educated male MA residents in our sample is

42 years. Eighty-four percent of this sample is white and 71 percent is married. Turning next to the area variables, we see that approximately 34 percent of the land area in the MAs is classified as mountain or hills and 23 percent of the MAs are located next to a major water body. On average, there are approximately 600 violent crimes per 100,000 persons and the ozone standard is exceeded about 2 days per year. The average monthly apartment rent is $516 and the median house value is approximately $80 thousand, on average, which is similar to the average per acre present value of land rents for urbanized land. Averaged over MAs and individuals, the mean value of our housing cost variable is $6.48 thousand per year. The average weekly wage is $979, or about $50,000 on an annual basis.

5. Results

Wages and Housing Costs

Although we estimate separate wage equations for all 291 MAs, we present estimates for a national-level model in Table 4 to indicate the general nature of the results. The signs of the coefficients are consistent with expectations (Lemieux, 2006; Chiswick and Miller, 2010). Wages are higher for white, college-educated married men who are fluent in English. Wages also increase with age, but at a decreasing rate. Wages fall for blacks (relative to the other race category) and separated or divorced people, but number of children does not have a significant effect. Wage rise for those who work more hours per week, but at a diminishing rate. As expected, executives receive higher wages. Relative to working in the manufacturing sector, wages fall for workers in agriculture, commerce, services, education, and administration and rise for workers in mining and energy, transportation, information/communication, and finance/insurance.

Also for illustration purposes, we produce national-level results for the probability of ownership and for the rental and owned value equations (Table 5). The results show that the likelihood of ownership increases for males, whites, and married households, and with age, but at a decreasing rate. The higher is the educational attainment, the higher is the probability of owning. Finally, executives are more likely to own than non-executives, which likely reflects the influence of income on ownership. These results are consistent with those found by Hendershott et al. (2009) using Australian data and by Painter et al. (2001) and Jepsen and Jepsen (2009) using U.S. data.

For the rental and homeowner value models, an illustrative set of national-level results (Table 5) suggest that housing is a normal good. That is, factors that increase wages (Table 4) tend to also increase housing expenditures, and vice-versa. When the household head is male, the rental and the owned value are higher than when the household head is female. Expenditures on housing increase with age, but at a diminishing rate, and they increase with educational attainment, number of children, and executive status. Expenditures are highest for married household heads, followed by single and separated or divorced heads. At the national level, the rental and owned values for blacks are smallest, followed by whites and other race. The finding that whites spend less on housing than households heads of other races contrasts with the result that whites have higher wages (Table 4), but is consistent with the study by Ihlanfeldt and Martinez-Vazquez (1986).

The key implication of our theory is that housing costs can be predicted using only individual characteristics (equation 4). In particular, we need not include measures of housing attributes in the housing cost equations because attributes are themselves a function of individual characteristics. As a robustness check, we re-estimate the annual rent (*Annualrent*) and owned home value (*Ownervalue*) equations for each MA, including variables for the year the housing unit was built, the number of rooms and bedrooms, the type of heating, and whether the unit is a detached house or other type of house.[17] We then predict housing costs in the MA selected in 2000 for each individual in our sample of working-age college-educated males. The inclusion of housing characteristics yields higher predicted housing costs, on average, but the difference is small (4 percent). In addition, the new housing cost predictions are highly correlated (90 percent) with the original estimates.

Migration choice

The results for the four migration models are presented in Table 6. For Model I, which includes our individual- and area-specific measure of housing costs (*Housingcost*), all of the coefficient estimates are significantly different from zero at the 1% level. The estimated value of the dissimilarity parameter (λ_1 in equations 8-10) lies in the unit interval, indicating that our model is consistent with utility maximization for all possible values of the explanatory variables (Train 2003). As well, because the estimate of λ_1 is different from one, the null hypothesis of IIA is rejected for our nesting structure. The results indicate that the likelihood of moving declines for older and married individuals and is lower for blacks and whites relative to other races. As expected, higher wages increase the likelihood that an MA is chosen, all else equal. MAs that are a greater distance from the origin MA, our proxy for higher moving costs, are less likely to be chosen, whereas a greater difference in population between the origin and destination MA increases the likelihood of the latter MA being chosen. We hypothesized, in contrast, that individuals would prefer MAs of similar size.

[17] We also examined two alternative specifications that omit controls for the number of bedrooms and controls for the number of rooms. This produced negligible differences in the results and so we focus our discussion on results produced with the original specification.

Most of the coefficients on the area variables in Model I have plausible signs. MAs on the Pacific and South Atlantic coasts are more likely to be chosen than inland MAs, whereas MAs adjacent to the Gulf Coast, the Great Lakes, and the North Atlantic coast are less likely to be selected. Higher lagged employment growth and fewer high ozone days increase the likelihood that an MA will be chosen. MAs with higher January temperatures and lower July temperatures and humidity and less annual precipitation are more desirable to migrants. Contrary to expectations, the coefficient on *Violent crime* is positive and the negative coefficients on the *Mountain* and *Plainhills* variables suggest that varied topography is less desirable to migrants. It is possible that these variables are correlated with other MA attributes, such as the effectiveness of policing in the case of the crime variable.

The coefficient of primary interest is the one on the housing cost variable. The coefficient on *Housingcost* is negative and significantly different from zero at the 1% level, indicating that migrants are more likely to select areas with lower housing costs, all else equal. This is in contrast to the results for Models II-IV. The three alternative MA-level housing cost measures, *Median house value*, *Apartment rent*, and *Land rent*, have positive and significant coefficients. Notably, the coefficients on the other variables are similar in sign and magnitude across the four versions of the model. As a robustness check, we repeat the analysis using a sample of working-age male MA residents with lower educational attainment (1-3 years of college). The results (not reported) are similar to those in Table 6. The estimated coefficient on *Housingcost* is negative and significantly different from zero, whereas two of three alternative housing cost measures (*Median house value*, *Apartment rent*) have positive coefficients. With this sample, the coefficient on *Land rent* has the expected negative sign.

6. Simulated Changes in the Population Distribution

The estimated migration model (Model 1) can be used to simulate changes in the distribution of the U.S. population given changes in attributes of individuals. In this section, we evaluate how changes in age induce migration among working-age, college-educated males. As in many parts of the developed world, the median age of the U.S. population increased throughout the 20th century, with large gains in the share of the population over 45 years (Hobbs and Stoops, 2002). Within our model, changes in age affect location decisions through multiple channels: expected wage in each MA, the likelihood of renting and owning and expected rental and owned home values in each MA, and the probability that an individual will choose to move. The simulation results show spatial changes in the population that are not revealed in the econometric estimates.

To conduct the simulations, we predict $P_{ij} = P_{im} \cdot P_{ij|m}$, the probability that the *i*th individual selects the *j*th MA, under baseline and alternative scenarios. These predictions make use of the estimated model parameters for the wage, housing cost, and migration models in addition to data on individual and MA attributes. Under the baseline scenario, we predict the initial distribution of working-age, college-educated males given their observed attributes. Specifically, for each location *j*, we estimate the number of individuals expected to choose that area as:

(11) $$N_j^b = \sum_i P_{ij}^b \cdot Weight_i$$

where $Weight_i$ is the person weight for individual *i*.[18] Because P_{ij}^b represents individual *i*'s initial location as the degenerate area choice associated with the stay decision, N_j^b explicitly accounts for the endogenous choice of the origin.

For the alternative scenarios, we compute $N_j^a = \sum_i P_{ij}^a \cdot Weight_i$ in the same way except that P_{ij}^a incorporates the following change: the age of all individuals 25-30 years is increased by 30 years to 60-65 years. This simulation allows us to predict how locations will change as individuals advance from the early to late stage in their career, holding other factors constant.[19] When then compute the percentage change in the population of working-age, college-educated males residing in each metropolitan area. The simulated change in age changes each individual's probability of selecting each MA. Thus, our results indicate the associated net effect of in- and out-migration on each MA's population of working-age, college-educated males relative to the baseline scenario with no age change.

The results are displayed in Figure 1, which depicts each of the 291 MAs in our study as a distinct polygon. The map reveals the strong draw of major MAs. Population increases are predicted for many MAs within the Boston-Washington corridor as well as Los Angeles, Phoenix, Houston, Minneapolis, Chicago, and Detroit. These results indicate that individuals who begin their careers in smaller MAs will tend to have moved to large MAs by the end of their careers.[20] As indicated in Tables 4-6, wages and housing costs tend to increase with age and age diminishes the likelihood that an individual moves. The simulation results suggest that the

[18] Equation (11) is computed using the entire sample of working-age, college-educated males, not just the sub-sample used to estimate the migration model.
[19] Of course, attributes such as marital status and number of children are likely to change with age.
[20] These results are consistent with trends for the U.S. population as a whole (Hobbs and Stoops, 2002). In 1950, 29 percent of the U.S. population lived in a MA with population over 1 million. By 2000, this figure had risen to 57 percent.

higher wages offered in larger MAs (and other MAs that gain population) offset the housing cost increases as well as utility losses from leaving the origin location.

7. Conclusions

Standard theory predicts that individual migration decisions for working-age adults will depend on area differences in wages, housing costs, and amenities. While the importance of wages and amenities is well-established from previous empirical studies, evidence regarding housing costs is far less conclusive. Many migration analyses do not include housing cost measures or find that aggregate measures such as median housing prices yield mixed results. In this study, we present and test an alternative approach to modeling housing costs. We show that if the utility function is quasi-linear in housing attributes and a numeraire good, then utility-maximizing housing costs can be represented as a function of individual attributes (age, race, etc.), as in a Mincerian wage equation. Housing costs and income can, thus, be interpreted as optimal expenditures on the differentiated goods housing and employment. Using large samples of individuals, we estimate the relationship between housing costs and individual attributes for each of 291 MAs in the U.S. Our approach accounts for rental and ownership decisions, the costs of rental and owned properties, and the costs of holding housing capital. As a robustness check, we show that adding variables for housing characteristics to the housing cost equations has little effect on predictions of housing expenditures.

We test our housing cost measure in a national-scale migration analysis focused on working-age, college-educated males. We find the expected result that higher housing costs reduce the likelihood that an MA is selected, holding wages and amenities constant. In contrast, three alternative MA-level measures yield counter-intuitive positive effects of housing costs on

MA choice. Similar results are found with an alternative sample of working-age males. A key feature of our housing cost measure is that it is specific to areas and individuals, unlike the aggregate variables that differ only by MA. Potential migrants are likely to base decisions on the costs of housing that they themselves would select, rather than what the average or median MA resident would choose, implying measurement error in the aggregate housing cost measures. Correlation between this measurement error and unobservable area attributes is a likely source of bias in the coefficient estimates on housing costs. In our analysis, the housing and wage variables control for any unobservable area-level effects, including amenities, transmitted through housing and labor markets.

Our migration analysis has a number of important additional features. First, area-specific wage equations specified in terms of individual attributes are used to predict labor market outcomes for individuals in unselected areas. Expected wages have the expected positive effect on the likelihood that an area is selected, all else equal. Second, we represent the origin as an endogenous choice and model the propensity to remain in the origin in terms of individual attributes. We find that within our sample of working-age, college-educated males, the likelihood of moving is lower for older, married whites. Third, we control for a large number of area amenities, including topography, climate, proximity to major water bodies, air quality, area employment growth, and violent crime. In most cases, the amenity variables have the expected effects on the likelihood that an area is selected. For example, consistent with earlier studies, migrants prefer MAs with higher winter temperatures, lower summer temperatures and humidity, and less annual precipitation. We use the estimated migration model to simulate changes in the MA populations given changes in the ages of working-age, college-educated males. The results suggest large metropolitan areas become more attractive to this group as they get older. Inter-urban migration has important implications for public policy as population restribution affects local consumption of natural resources such as water and needs for public services and infrastructure. An interesting extension of this study would be to predict how U.S. migration patterns would be affected by future climate change.

References

Ashenfelter, O., Harmon, C., Oosterbeek, H., 1999. A review of estimates of the school/earnings relationship, with tests for publication bias. Labour Economics 6, 453-470.

Bayer, P., Keohane, N., Timmins, C., 2009. Migration and hedonic valuation: The case of air quality. Journal of Environmental Economics and Management 58, 1-14.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press, Cambridge, MA.

Berger, M.C., Blomquist, G.C., 1992. Mobility and destination in migration decisions: The role of earnings, quality of life, and housing prices. Journal of Housing Economics 2, 37-59.

Bishop, K.C., 2008. A dynamic model of location choice and hedonic valuation. Unpublished paper, Olin Business School, Washington University, St. Louis, MO.

Blomquist, G.C., Berger, M.C., Hoehn, J.P., 1988. New estimates of the quality of life in urban areas. American Economic Review 78, 89-107.

Chan, S., 2001. Spatial lock-in: Do falling house prices constrain residential mobility? Journal of Urban Economics 49, 567-586.

Chen, Y., Rosenthal, S.S., 2008. Local amenities and life cycle migration: Do people move for jobs or fun? Journal of Urban Economics 65, 519-37.

Cheshire, P.C., Magrini, S., 2006. Population growth in European cities: Weather matters – but only nationally. Regional Studies 40, 23-37.

Chiswick, B.R, Miller, P.W., 2010. Occupational language requirements and the value of English in the U.S. labor market. Journal of Population Economics 23, 353-372.

Clark, D.E., Hunter, W.J., 1992. The impact of economic opportunity, amenities and fiscal factors on age-specific migration rates. Journal of Regional Science 32, 349-365.

Clark, D.E., Murphy, C.A., 1996. Countywide employment and population growth: An analysis of the 1980s. Journal of Regional Science 36, 235-256.

Dahl, M.S., Sorenson, O., 2010. The migration of technical workers. Journal of Urban Economics 67, 33-45.

Davies, P.S., Greenwood, M.J., Li, H., 2001. A conditional logit approach to U.S. state-to-state migration. Journal of Regional Science 41, 337-360.

Detang-Dessendre, C., Goffette-Nagot, F., Piguet, V., 2008. Life cycle and migration to urban and rural areas: Estimation of a mixed logit model on French data. Journal of Regional Science 48, 789-824.

Duffy-Deno, K.T., 1998. The effect of federal wilderness on county growth in the intermountain western United States. Journal of Regional Science 38, 109-136.

Engelhardt, G.V., 2003. Nominal loss aversion, housing equity constraints, and household mobility: Evidence from the United States. Journal of Urban Economics 53, 171-195.

Eichman, H., Hunt, G.L., Kerkvliet, J., Plantinga, A.J., 2010. Local employment growth, migration, and public land policy: Evidence from the Northwest Forest Plan. Journal of Agricultural and Resource Economics 35, 316-333.

Ferreira, F., Gyourko, J., Tracy, J., 2010. Housing busts and household mobility. Journal of Urban Economics 68, 34-45.

Gottlieb, P.D., Joseph, G., 2006. College-to-work migration of technology graduates and holders of doctorates within the United States. Journal of Regional Science 46, 627-659.

Greenwood, M.T., Hunt, G.L., Rickman, D.S., Treyz, G.I., 1991. Regional equilibrium, and the estimation of compensating differentials. American Economic Review 81, 1382-1390.

Hendershott P.H., Ong, R., Wood, G.A., Flatau, P., 2009. Marital history and home wwnership: Evidence from Australia. Journal of Housing Economics 18, 13-24.

Hobbs, F.,Stoops, N., 2002. U.S. Census Bureau 2000 Special Reports, Series CENSR-4, Demographic Trends in the 20th Century. U.S. Government Printing Office, Washington, DC.

Hunt, G.L., 1993. Equilibrium and eisequilibrium in migration modeling. Regional Studies 27, 341-349.

Hunt, G.L., 2000. Alternative nested logit model structures and the special case of the partial degeneracy. Journal of Regional Science 40, 89-113.

Hunt, G.L., Mueller, R.E., 2002. A methodology for estimating returns to skills for Canadian provinces and U.S. states. Journal of Regional Science 42, 127-143.

Hunt, G.L., Mueller, R.E., 2004. North American migration: Returns to skill, border effects, and mobility costs. Review of Economics and Statistics 86, 988-1007.

Ihlanfeldt, K.R., Martinez-Vazquez, J., 1986. Alternative value estimates of owner-occupied housing: Evidence on sample selection bias and systematic errors. Journal of Urban Economics 20, 356-369.

Jeanty, P.W., Partridge, M., Irwin, E., 2010. Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Regional Science and Urban Economics 40, 343-352.

Jepsen, C., Jepsen, L.K., 2009. Does home ownership vary by sexual orientation? Regional Science and Urban Economics 39, 307-315.

Kennan, J., Walker, J.R., 2011. The effect of expected income on individual migration decisions. Econometrica 79, 211-251.

Lemieux, T., 2006. The ‘Mincer equation’ thirty years after schooling, experience, and earnings, in: Grossbard, S. (Ed.), Jacob Mincer: A Pioneer of Modern Labor Economics. Springer, New York, pp. 127-145.

Lewis, D.J., Hunt, G.L., Plantinga, A.J., 2002. Public conservation land and employment growth in the northern forest region. Land Economics 78, 245-59.

Lubowski, R.N., 2002. Determinants of Land-Use Transitions in the United States: Econometric Analysis of Changes among the Major Land-Use Categories. Ph.D. Dissertation, Harvard University, Cambridge, MA.

McGranahan, D.A., 1999. Natural Amenities Drive Population Change. Food and Rural Economics Division, Economic Research Service, U.S. Department of Agriculture Report 781.

Mueser, P.R., Graves, P.E., 1995. Examining the role of economic opportunity and amenities in explaining population redistribution. Journal of Urban Economics 37, 176-200.

Painter, G., Gabriel, S., Myers, D., 2001. Race, immigrant status, and housing tenure choice. Journal of Urban Economics 49, 150-167.

Poston, D.L., Zhang, L., Gotcher, D.J., Gu, Y., 2009. The effect of climate on migration: United States, 1995-2000. Social Science Research 38, 743-753.

Rappaport, J., 2007. Moving to nice weather. Regional Science and Urban Economics 37, 375-398.

Roback, J., 1982. Wage, rents, and the quality of life. Journal of Political Economy 90, 1257-1278.

Rosen, S., 1979. Wage-based indices of urban quality of life, in: Mieszkowski, P., Straszheim, M. (Eds.), Current Issues in Urban Economics. The Johns Hopkins University Press, Baltimore, MD.

Ruggles, S., Alexander, J.T., Genadek, K., Goeken, R., Schroeder, M.B., Sobek, M., 2010. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. University of Minnesota, Minneapolis.

Scott, A.J., 2010. Jobs or amenities? Destination choices of migrant engineers in the USA. Papers in Regional Science 89, 43-63.

Seig, H., Smith, V.K., Banzhaf, H.S., Walsh, R., 2004. Estimating the general equilibrium benefits of large changes in spatially delineated public goods. International Economic Review 45, 1047-1077.

So, K.M., Orazem, P.F., Otto, D.M., 2001. The effects of housing prices, wages and commuting time on joint residential and job location choices. American Journal of Agricultural Economics 83, 1036-1048.

Train, K.E., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge.

U.S. Bureau of the Census, 1998. State and Metropolitan Area Data Book 1997-98 (5th edition.). U.S. Government Printing Office, Washington, DC.

Table 1. Individual-level Variables

Variable	Description
Male	Indicator variable for male
Female	Indicator variable for female
Age	Age in years
White	Indicator variable for white race
Black	Indicator variable for black race
Other	Indicator variable for other race
Married	Indicator variable for married
Separated/divorced	Indicator variable for separated or divorced
Single	Indicator variable for single
Children	Number of children
Household	Indicator variable for head of household
English	Indicator variable for English fluency
No English	Indicator variable for lack of English fluency
Less than high school	Educational attainment is less than high school
High school	Educational attainment is high school
Some college	Educational attainment is 1-3 years of college
College or more	Educational attainment is 4 years college or more
Wage	Log of annual salary wages divided by number of weeks worked
Usual work hours	Typical number of hours worked per week
Executive	Indicator variable for executive position
Not executive	Indicator variable for non-executive position
Owner	Indicator variable for home ownership
Renter	Indicator variable for home rental
Home Value	Value of an owned home
Rent	Annual rent
Manufacturing	Indicator variable for employment in manufacturing sector
Agriculture	Indicator variable for employment in agriculture sector
Mining and energy	Indicator variable for employment in mining and energy sector
Construction	Indicator variable for employment in construction sector
Commerce	Indicator variable for employment in commerce sector
Transportation	Indicator variable for employment in transportation sector
Information/communication	Indicator variable for employment in information or communication sector
Finance/insurance	Indicator variable for employment in finance or insurance sector
Services to enterprises	Indicator variable for employment in services to enterprises sector
Education	Indicator variable for employment in education sector
Services to individuals	Indicator variable for employment in services to individuals sector
Adminstration	Indicator variable for employment in administration sector
Weight	Person weight

Note: all variables are measured in 2000 and taken from the PUMS 5% sample.

Table 2. Metropolitan Area Measures

Variable	Description	Source
Distance	Radial distance between centroids of metropolitan areas	Authors' calculation
Population spread	Absolute value of log of ratio of populations between pairs of metropolitan areas	Authors' calculation
Gulf Coast	Indicator variable for border with Gulf Coast	Authors' calculation
Great Lakes	Indicator variable for border with Great Lakes	Authors' calculation
North Atlantic	Indicator variable for border with North Atlantic	Authors' calculation
Pacific	Indicator variable for border with Pacific Ocean	Authors' calculation
South Atlantic	Indicator variable for border with South Atlantic	Authors' calculation
Violent crime	Violent crimes per 100,000 population, 1995	State and Metropolitan Area Data Book 1997-1998
Employment growth	Average annual growth rate in total employment, 1990-1995	State and Metropolitan Area Data Book 1997-1998
January temperature	Mean temperature for January, 1941-1970	McGranahan (1999)
July temperature	Mean temperature for July, 1941-1970	McGranahan (1999)
July humidity	Mean relative humidity for July, 1941-1970	McGranahan (1999)
Plainshills	Proportion of land area classified as plains with hills or mountains	McGranahan (1999)
Mountain	Proportion of land area classified as hills and mountains	McGranahan (1999)
Precipitation	Annual precipitation in inches	Lawrence Berkeley Laboratory (http://eande.lbl.gov/IEP/high-radon/data/lbnl-met.html)
Ozone	Maximum number of days any monitor exceeds ozone standard, 1995	U.S. Environmental Protection Agency (http://www.epa.gov/aqspubl1/annual_summary.html)
Median house value	Median value of owner-occupied housing units, 1990	County and City Data Book 1994
House value change	Percent change in median value of owner-occupied housing units, 1990-2000	County and City Data Books, various dates
Land rent	Present discounted value of per-acre land rents for all urbanized land, 1995	Lubowski (2002)
Apartment rent	40th percentile monthly rental rate for 2 bedroom apartments, 1995	U.S. Department of Housing and Urban Development (http://www.huduser.org/datasets/fmr.html)

Table 3. Summary Statistics for Variables in the Migration Models

Variables	Individual Mean	Individual Standard Deviation	Area Mean	Area Standard Deviation	Individual/area Mean	Individual/area Standard Deviation
Age	42.38	10.15				
Married	0.71	0.45				
Black	0.06	0.24				
White	0.84	0.37				
Wage					978.75	277.52
Distance					1727.51	1124.11
Population spread					1.78	1.13
Mountain			0.34	0.44		
Plainhills			0.08	0.24		
Gulf Coast			0.05	0.23		
Great Lakes			0.05	0.23		
North Atlantic			0.04	0.20		
Pacific			0.05	0.23		
South Atlantic			0.04	0.21		
Employment growth			0.02	0.01		
Violent crime			600.2	308.3		
January temperature			36.52	12.68		
July temperature			75.96	5.45		
July humidity			58.54	14.60		
Precipitation			36.47	13.53		
Ozone			1.75	6.71		
Housingcost ($1000/year)					6.85	4.43
Mean house value ($1000)			80.28	46.28		
Apartment rent ($/month)			515.85	128.49		
Land rent ($1000)			82.71	82.80		
Number of observations	24604		291		2460400	

Notes: Because we randomly sample 100 MAs for each individual, the number of observations for the individual/area variables is 2460400=24604×100.

Table 4. An illustrative national-level wage equation

Variable	Parameter	Standard error	*t*-statistic
Intercept	1.4846	0.001920	771.3
Male	0.0815	0.000249	327.8
Age	0.0167	0.000084	198.1
Age squared	-0.0002	0.000001	-163.6
White	0.0083	0.000353	23.5
Black	-0.0124	0.000456	-27.2
Married	0.0399	0.000313	127.5
Separated/divorced	-0.0008	0.000369	-2.2
Children	0.0001	0.000100	0.8
Household	0.0562	0.000240	233.8
English	0.0551	0.000432	127.4
Less than high school	-0.0600	0.000398	-150.9
Some college	0.0531	0.000265	200.1
College or more	0.1621	0.000311	521.3
Usual work hours	0.0234	0.000035	660.3
Usual work hours squared	-0.0002	0.000000	-442.3
Executive	0.1106	0.000264	419.3
Agriculture	-0.1508	0.001030	-145.7
Mining and energy	0.0544	0.000858	63.4
Construction	0.0005	0.000474	1.1
Commerce	-0.0562	0.000374	-150.3
Transportation	0.0135	0.000531	25.5
Information/communication	0.0137	0.000631	21.8
Finance/insurance	0.0029	0.000470	6.1
Services to enterprises	-0.0250	0.000433	-57.8
Education	-0.0760	0.000369	-206.1
Services to individuals	-0.1168	0.000411	-284.0
Administration	-0.0116	0.000484	-23.9

Dependent variable = natural log of average weekly salary wages in 2000
The omitted categories are female, other, single, no English, high school, not executive, manufacturing

No. observations = 5379510
Adj. R-squared = 0.40

Table 5. Illustrative national-level ownership, rental value, and owned value equations

	Parameter estimates		
Variable	Ownership	Rental value	Owned value
Intercept	-4.049	5.915	10.396
	0.003	0.008	0.009
Male	0.113	0.032	0.042
	0.0004	0.001	0.001
Age	0.103	0.013	0.036
	0.0001	0.0004	0.0004
Age squared	-0.0008	-0.0002	-0.0003
	0.000002	0.000004	0.000005
White	0.399	-0.010	-0.038
	0.0006	0.001	0.002
Black	-0.116	-0.167	-0.339
	0.0007	0.002	0.002
Married	0.848	0.135	0.222
	0.0005	0.001	0.002
Separted/divorced	0.114	0.001	-0.030
	0.0005	0.001	0.002
Children	0.084	0.021	0.034
	0.0002	0.0005	0.0005
English	0.538	-0.016	-0.068
	0.0007	0.002	0.002
Less than high school	-0.303	-0.141	-0.239
	0.0007	0.001	0.002
Some college	0.116	0.121	0.190
	0.0005	0.001	0.001
College or more	0.204	0.289	0.507
	0.0006	0.001	0.001
Executive	0.152	0.141	0.153
	0.0004	0.001	0.001
Dependent variable	*Owner*	*ln(Annualrent)*	*ln(Ownervalue)*
No. observations	2773017	907015	1866002
Adj. R-square	NA	0.18	0.22

Notes: The omitted categories are female, other, single, no English, high school, not executive. Standard errors are given below parameter estimates.

Table 6. Estimation Results for the Nested Logit Models of Migration Choice

	Model I		Model II		Model III		Model IV	
Variable	Parameter	z-statistic	Parameter	z-statistic	Parameter	z-statistic	Parameter	z-statistic
Move/stay decision								
Intercept	1.292	-858.7	1.186	296.5	1.304	344.42	1.257	266.9
Age	-0.071	-106.2	-0.067	-828.9	-0.071	-859.07	-0.072	-694.8
Married	-0.168	-129.6	-0.189	-121.0	-0.167	-105.68	-0.112	-57.1
Black	-0.299	-107.7	-0.345	-150.3	-0.299	-129.79	-0.294	-102.2
White	-0.394	340.9	-0.411	-114.1	-0.394	-107.56	-0.533	-117.1
MA choice								
Wage	0.0004	231.4	0.0003	157.4	0.0003	176.54	0.0004	167.4
Distance	-0.0002	-342.0	-0.0002	-210.8	-0.0002	-333.88	-0.0002	-279.7
Population spread	0.209	355.2	0.214	249.5	0.201	344.5	0.212	291.5
Mountain	-0.018	-42.1	-0.047	-93.3	-0.021	-50.41	-0.029	-51.9
Plainhills	-0.024	-33.4	-0.033	-44.1	-0.038	-54.92	-0.021	-23.8
Gulf Coast	-0.019	-24.3	-0.022	-26.6	-0.017	-21.98	-0.012	-11.8
Great Lakes	-0.136	-197.5	-0.146	-160.2	-0.122	-184.16	-0.130	-152.8
North Atlantic	-0.025	-46.6	-0.081	-114.6	-0.061	-108.8	-0.037	-55.4
Pacific	0.030	37.3	0.008	8.8	0.021	26.53	0.022	21.7
South Atlantic	0.015	21.6	0.010	13.2	0.002	2.81	0.029	31.4
Employment growth	3.124	214.7	4.921	189.3	3.508	240.01	3.447	186.6
Violent crime	0.00001	13.8	-0.00004	-68.7	0.00001	18.74	0.00001	12.2
January temperature	0.004	173.0	0.003	121.7	0.003	139.29	0.003	91.1
July temperature	-0.009	-167.7	-0.009	-130.5	-0.007	-137.85	-0.006	-87.2
July humidity	-0.001	-73.3	-0.001	-80.3	-0.001	-73.43	-0.001	-64.7
Precipitation	-0.001	-36.0	0.001	49.3	-0.0002	-15.32	0.00001	0.5
Ozone	-0.002	-131.0	-0.001	-80.0	-0.002	-121.99	-0.002	-108.3
Housingcost	-0.0007	-16.0						
Mean house value			0.0008	140.0				
Apartment rent					0.0002	137.6		
Land rent							0.0003	72.0
lambda_1	0.193	0.0005*	0.212	0.0009*	0.189	0.0005*	0.197	0.0007*
Number of individuals	13456202		13456202		13456202		8895995	

* Standard error

Note: The models are estimated with the sample of 24,604 working-age, college-educated male MA residents, who represent approximately 13.5 million individuals. Due to missing values of the land rent variable, this number is smaller for Model IV.

Figure 1. Simulated Change in the Distribution of Working-age College-Educated Males (measured as a percentage deviation from the baseline with the number of MAs in each category indicated in parentheses)

Beginning of career (age 25-30) changes to end of career (age 60-65)