A Dynamic Model of Location Choice and Hedonic Valuation

Kelly C. Bishop*
Department of Economics
Arizona State University

Abstract

Hedonic equilibrium models allow researchers to recover willingness to pay for spatially delineated amenities by using the notion that individuals "vote with their feet." However, the hedonic literature and, more recently, the estimable Tiebout sorting model literature, have largely ignored both the costs associated with migration (financial and psychological), as well as the forward-looking behavior that individuals exercise in making location decisions. Each of these omissions could lead to biased estimates of willingness to pay. Building upon dynamic migration models from the labor literature, I estimate a fully dynamic model of individual migration at the national level that explicitly controls for moving costs and forward-looking behavior. By employing a two-step estimation routine, I avoid the computational burden associated with the full recursive solution and can then include a richly-specified, realistic state space. With this model, I am able to perform non-market valuation exercises and learn about the spatial determinants of labor market outcomes in a dynamic setting. Including dynamics has a significant positive impact on the estimates of willingness to pay for air quality. In addition, I find that location-specific amenity values can explain important trends in observed migration patterns in the United States.

*Department of Economics, Arizona State University; kelly.bishop@asu.edu. I would like to thank Chris Timmins, Peter Arcidiacono, Pat Bayer, Marty Smith, and seminar participants at Duke, Claremont McKenna, Berkeley ARE, Washington University, New York Fed, Maryland AREC, Brown University, SUNY Albany, Iowa State, University of Michigan, St. Louis Fed, University of Pennsylvania, and Arizona State University for their helpful comments and suggestions.

1 Introduction

Since the seminal theoretical models of Hicks (1932) and Sjaastad (1962), economists have been interested in examining both the determinants and consequences of individuals' location decisions. Individuals choose their location for a variety of reasons, including employment opportunities and family-related motivations. Importantly, they also care about local public goods and amenities when making these decisions. This feature of their behavior provides a basis for the non-market valuation of local attributes using the ideas in Tiebout (1956).[1] Complicating such an exercise, however, is the fact that location choice is also an inherently dynamic decision process. Individuals face high costs associated with moving. We therefore expect that they would look to the future with regard to wage opportunities and time-varying location attributes when choosing where to live today. The hedonic and empirical Tiebout sorting literatures have, however, essentially ignored these complications. In the case of amenity valuation, ignoring dynamic considerations would bias estimates of willingness to pay downwards if individuals chose locations with high predicted, but low current levels of the amenity in question. Similarly, ignoring moving costs would also bias estimates downwards if individuals had ties, either financial or psychological, to areas with low amenity levels. My analysis addresses these shortcomings by developing and estimating a fully dynamic model of national migration and using it to recover individual willingness to pay for air quality. Importantly, incorporating dynamic behavior results in significantly larger estimates of the willingness to pay for air quality.

Because of the computational burden associated with the traditional, full-solution method of dynamic analysis, previous dynamic models of migration have been forced to either limit the individual to a simple decision of move-stay, which is inappropriate for valuing amenities that vary spatially, or to essentially ignore the role of local amenity values in migration altogether. In a model of marital status and family location decisions, Gemici (2011) specifies the geographic choice set as the nine Census divisions in the United States. At this level of

[1]Notable papers which incorporate these ideas into hedonic and sorting models include: Rosen (1979), Roback (1982), Hoehn, Berger, and Blomquist (1987), Blomquist, Berger, and Hoehn (1988), Epple and Sieg (1999), Sieg, Smith, Banzhaf, and Walsh (2004), Bajari and Kahn (2005), Bayer, McMillan, and Rueben (2004), and Kuminoff (2012).

geography, it is impractical to infer preferences for spatially varying amenities. Kennan and Walker (2011) estimate a model of expected income and individual location decisions, allowing the choice set to be defined at the finer level of U.S. states. However, the computational burden of the traditional estimator limits the number and the type of parameters that can be estimated, particularly as the number of elements in the choice set grows. While including mean wages (deflated to reflect differences in cost of living), population, and historical averages of temperature to describe location amenities, Kennan and Walker specify these attributes as being fixed through time.

I model the location decisions of individuals using panel data from the National Longitudinal Survey of Youth (NLSY79), explicitly controlling for costs of moving, which are allowed to vary with a variety of factors, including age. Individuals are forward-looking and choose the location that maximizes the discounted stream of benefits in expectation. Wages are allowed to evolve stochastically and the notion of job search is included,[2] as individuals learn their full wage in a given location only after moving there and paying the associated moving costs.

To properly define local attributes, one must move to an even finer level of geography than the U.S. state. I estimate my model at the level of metropolitan area,[3] allowing individuals to be forward-looking with respect to a rich set of both fixed and time-varying individual and location attributes. I circumvent the computational burden of the traditional estimation routine by employing the two-step estimator proposed by Arcidiacono and Miller (2011).[4] This method avoids the need for the full recursive solution by estimating the components of the value function in two stages. This simplifies the main estimating equation to one that is

[2]Kennan and Walker allow the location-specific match component of earnings to evolve stochastically. I additionally allow the location-specific mean wage to evolve stochastically, as well as the location-specific cost of living, proxied by median house prices.

[3]To avoid issues with sparse data, I limit my analysis to the fifty most populous metropolitan statistical areas (MSAs) in the NLSY79. I additionally create nine Census division, non-MSA "catch-alls" to limit attrition from the panel. Thus, the choice set is comprised of fifty-nine locations. With more data describing migration behavior (see footnote 18), my proposed model could handle a much larger choice set.

[4]The use of Hotz and Miller-style two-step approaches has become increasing popular, particularly to estimate dynamic games in Industrial Organization, as they dramatically reduce the computational burden of these models.See, for example, Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky and Berry (2007), and Pesendorfer and Schmidt-Dengler (2008).

linear in the structural parameters and that can handle a large descriptive state space.

The two-step approach in a migration model uses the fact that individuals optimally choose where to live and work based on both current and expected future local attributes. Thus, the probability of choosing a particular location at a given time incorporates the full information set available that period. Therefore, it is possible to replace the difficult terms within the future value component of lifetime utility with future conditional choice probabilities. In the first step of the estimation routine, I flexibly estimate the future probability of choosing a particular location, conditional on the agent being at any one of the potential points in the state space. The remaining structural parameters are then estimated in the second stage, after plugging in the conditional choice probabilities and integrating out over the transitions of the stochastic variables.

This paper makes several contributions to the literature. While previous dynamic models of location choice were forced to essentially ignore the role of amenity values, the use of a two-step approach allows me to include a rich descriptive set of local attributes that are allowed to transition through time, in addition to location fixed-effects (which control non-parametrically for all time-invariant local attributes). This extends the existing labor literature of migration models and allows me to estimate willingness to pay for non-market amenities which vary spatially. The framework also extends the existing literature on Tiebout sorting models, which typically ignore the costs associated with migration (both financial and psychological)[5] as well as the forward-looking behavior that individuals exercise in making location decisions.

I find that the inclusion of moving costs and forward-looking behavior has a significant impact on estimates of willingness to pay. I estimate a willingness to pay for a one unit decrease in air pollution (measured as a one microgram per cubic meter of air decrease in particulate matter)

[5] An exception is Bayer, Keohane, and Timmins (2009), which allows for moving costs that depend only on an individual's birth location, but which ignores other types of dynamics. Deriving a willingness to pay for air quality using a static Tiebout sorting model, they find that ignoring moving costs substantially underestimates the value of clean air.

of \$198.22,[6] which implies an elasticity for the average individual of -0.17.[7] This willingness to pay is 2.2 times larger than that estimated by the dynamic model without ties to birth location and 2.6 times larger than that estimated by a simple static model.[8]

Finally, the estimated model is used to perform counterfactual simulations of individual migration decisions. In particular, I analyze the effects of geography and location-specific amenity values on observed patterns of migration. I simulate behavior under the assumption of uniform amenity values across locations and find that the inclusion of amenity values (as opposed to solely falling wages in declining industries) helps to explain much of the recent migration out of the North East and to the South and West of the United States.

This paper is organized as follows. I describe my model in Section 2. Sections 3 and 4 discuss the main data sources and the two-step estimation routine in detail. I present the results in Section 5. Finally, Section 6 concludes.

2 Model

I model the location decisions of individuals in a finite-horizon framework. In each period, individuals receive flow utility associated with their current location and incur a moving cost if they decide to relocate in the following period. Locations will be defined as a set of U.S. metropolitan areas.

The timing of the decision process is important and highlights the effect of expectations on the location decision. The decision period will be two years in length.[9] An agent begins period t in a location denoted j. The agent knows the full utility flow associated with their current location. If they choose to move (*i.e.* begin period $t+1$ in a new location k), they

[6]The willingness to pay is reported in inflation-adjusted 2000 dollars. For ease of comparison to previous estimates (see, for example, Chay and Greenstone (2005) and Bayer, Keohane, and Timmins (2009)), which are expressed in constant 1982-1984 dollars, divide all figures by 1.72.

[7]The elasticity is calculated at a mean income of $\$32,997.26$ (in 2000 dollars) and a mean particulate matter concentration of 27.79 $\mu g/m^3$.

[8]See Smith and Huang (1995) for an excellent review of the willingness to pay for air quality literature.

[9]In my primary data source, the NLSY79, individuals are interviewed biennially.

pay all associated moving costs (which are known in full) in the current period, t. However, the individual only has expectations over the value of living in k, which will not be realized until period $t + 1$, after all costs have been paid.

Agents receive flow utility from income, living in the location or region of their birth, the amenities associated with their current location, and an additively-separable choice-specific shock. Moving costs are specified as a function of the agent's age, the distance between the locations, and whether or not the agent has previously lived in the chosen location. Lifetime utility is given by current flow utility and the discounted stream of expected future per-period utilities. Individuals are forward looking with respect to income and time-varying location attributes. Uncertainty comes in the form of the transition of these variables, a location- and individual-specific match component of income, and an idiosyncratic shock to utility.

2.1 Income

In each location in each period, an individual receives the location- and time-specific mean income, a return to their individual characteristics (including age), and an idiosyncratic error on income. Following Kennan and Walker (2011), I assume that this error term can be divided into three distinct components: a fixed location-specific match component, an individual fixed effect, and a transitory component. Thus, the income of individual i, in location j, at time t is specified as:[10]

$$inc_{i,j,t} \;=\; \omega_i'\gamma \;+\; f(age_{i,t}) \;+\; \mu_{j,t} \;+\; (\theta_{i,j} + \eta_i + e_t)$$

where ω_i is the vector of fixed characteristics of individual i, $age_{i,t}$ is individual i's age in period t, and $\mu_{j,t}$ is the location-specific mean wage at time t. The error term is comprised of the match component $\theta_{i,j}$, the individual fixed effect η_i, and the transitory earnings component e_t. All three are assumed to be i.i.d. (across individuals, locations, and time) normally distributed with mean 0 and respective variances σ^2_θ, σ^2_η, and σ^2_e.

[10]In this specification, the return to individual attributes is fixed across locations. In practice, I run the model on a homogenous sample, so these effects are captured non-parametrically by the location-specific mean wage. Individual characteristics (besides age, but including education) are held fixed through time.

Individuals know their individual fixed effect and the current values of their location-specific match component and time-specific transient component. However, an individual does not know the value of their match component in other locations or the value of the transient component in future periods (but does know the distributions from which they are drawn). I additionally allow individuals to know the value of their match component in their prior location.[11]

Income plays an important role in the decision to migrate. However, only the components of income that vary with location, $\mu_{i,j,t}$ and $\theta_{i,j}$, drive migration decisions. Thus, it is possible to ignore the other components in the estimation, as they will drop out of utility. While it is straightforward to include the location- and time-specific mean income, I do not observe the individual's match component. Thus, in estimation I employ the signal extraction of Kennan and Walker.

The signal extraction uses observed wage histories to extract estimates of match components for each individual in each of their observed locations. The central idea of the extraction is that the population-wide distributions can be updated for each individual based on their observed wage histories. For example, if an individual moves and earns an above-average income in both observed locations, it is possible to increase the estimate of the individual's fixed effect (η_i). Likewise, if an individual earns an unexpectedly high income following a move, it is possible to update the estimates of the respective match components ($\theta_{i,j}$).

Isolating the unobserved component of income for each individual in each period and denoting it $\Omega_{i,j(t)}$:

$$\Omega_{i,j(t)} = \theta_{i,j(t)} + \eta_i + e_t$$

it is possible to estimate the population-wide variances of the error (σ_θ^2, σ_η^2, and σ_e^2) by taking sample averages of the elements of the variance-covariance matrix, $\Omega_i \Omega_i'$, where Ω_i is the stacked vector of $\Omega_{i,j(t)}$.[12]

[11]This introduces "memory" in the form of Kennan and Walker's limited history approach.

[12]The sample average of the diagonal elements will yield an estimate of $\sigma_\theta^2 + \sigma_\eta^2 + \sigma_e^2$. The sample average of the off-diagonal elements that correspond to the same location will yield an estimate of $\sigma_\theta^2 + \sigma_\eta^2$. Finally, the sample average of the off-diagonal elements that correspond to different locations will yield the estimate of σ_η^2, which can be used to separately solve for σ_θ^2, σ_η^2, and σ_e^2.

Using these population-wide estimates, Kennan and Walker show that it is possible to update the conditional distribution of the fixed effect for each individual, based on observed income history Ω_i:

$$(\eta_i|\Omega_i) \sim N(\hat{\eta}_i, \hat{\sigma}^2_{i,\eta})$$

where:

$$\hat{\eta}_i = \hat{\sigma}^2_{i,\eta}\left[\sum_{\lambda=1}^{\Lambda_i}\frac{\Omega_{i,\lambda}}{\zeta_{i,\lambda}}\right], \quad \hat{\sigma}^2_{i,\eta} = \left[\frac{1}{\sigma^2_\eta} + \sum_{\lambda=1}^{\Lambda_i}\frac{\Omega_{i,\lambda}}{\zeta_{i,\lambda}}\right]^{-1}, \quad \zeta_{i,j} = \sigma^2_\theta + \frac{\sigma^2_e}{S_{i,j}}$$

and where Λ_i represents the number of locations visited by individual i and $S_{i,j}$ represents the number of periods that individual i spends in a given location j.

Using the expected value of individual i's fixed effect by drawing from the above distribution,[13] it is possible to generate a conditional distribution of observed location- and individual-specific match components:

$$(\theta_{i,j}|\Omega_i, E[\eta_i]) \sim N(\hat{\theta}_{i,j}, \hat{\sigma}^2_{i,\theta})$$

where:

$$\hat{\theta}_{i,j} = \sigma^2_{i,\theta}\left[\frac{\Omega_{i,j} - E[\eta_i]}{\pi_{i,j}}\right], \quad \hat{\sigma}^2_{i,\theta} = \left[\frac{1}{\sigma^2_\theta} + \frac{1}{\pi_{i,j}}\right]^{-1}, \quad \pi_{i,j} = \frac{\sigma^2_e}{S_{i,j}}$$

I discretize this distribution to the following three points of support, where the points are fixed over locations, individuals, and time: $\mu_j + \Phi^{-1}(\frac{1}{6})$, μ_j, $\mu_j - \Phi^{-1}(\frac{1}{6})$, where $\Phi(\cdot)$ is the standard normal distribution function.[14]

Thus, the signal extraction yields a probability that each individual received a "low," "medium," or "high" match component in each of their observed locations. When considering a new location, an individual expects to receive one of these draws, each with equal probability in expectation, based on the chosen points of discretization.

2.2 Moving Costs

In each period, an individual currently located in j chooses a location k from the set of all locations, J. Based on the timing of the model, individuals receive a utility flow associated

[13]In practice, I simulate using 1,000 draws for each individual.

[14]Discretizing in this fashion is shown to be optimal in Kennan (2004).

with their current location j and pay moving costs if their choice of location k is different from j.

The moving cost ($M_{i,j,k}$), depends on the individual's current location (j), choice location (k), prior location (l), and age. In addition to having full information about their current location j, individuals have more information regarding previous locations than locations that they have never visited. In particular, I allow individuals to know the value of the location- and individual-specific match component of income in their prior location. This form of "memory" follows the limited history approach of Kennan and Walker (2011) in allowing individuals to have more information regarding job prospects in places where they have lived previously.

I specify the moving costs associated with choosing location k while currently living in j as:

$$M_{i,j,k} = \psi_0 + \psi_1 distance_{j,k} + \psi_2 distance_{j,k}^2 - \psi_3 I_{k \in j.reg} - \psi_4 I_{k=l} + \psi_5 age_i$$

where ψ_0 is a fixed cost of moving, $distance_{j,k}$ is the distance in miles between current location j and chosen location k, $I_{k \in j.reg}$ is an indicator equal to one if k is in the same Census division as j, $I_{k=l}$ is an indicator equal to one if k is equal to individual i's prior location l, and age_i is individual i's current age.

I include both $distance_{j,k}$ and $distance_{j,k}^2$ to allow long-distance moves to be more expensive than more local moves, but with a decreasing effect. Moves that are within the same Census division, or return moves to a prior location, are thought to be less costly. Finally, age is included as it is thought that older individuals (prior to retirement) experience higher costs of moving than younger individuals.[15] Ties to birth location (an implicit form of moving cost) enter into utility directly, as individuals are expected to continually receive utility while residing in the location of their birth.

[15] In practice, I use individuals who are aged 21 to 45.

2.3 Utility Specification

Net of moving costs, individuals receive utility from income ($inc_{i,j,t}$), living in their birth location (b_i) or birth region, time-varying location attributes ($z_{j,t}$), fixed observable location attributes (χ_j), fixed unobservable location attributes (ξ_j), and a time-varying idiosyncratic unobservable location attribute ($\epsilon_{i,k,t}$). Thus, flow utility is given by:

$$\begin{aligned} u_{i,j,k,t} + \epsilon_{i,k,t} &= \alpha_{inc} inc_{i,j,t} + z'_{j,t}\alpha_z + \chi'_j \alpha_\chi + \xi_j \\ &\quad + \alpha_{bmsa} I_{j=b_i} + \alpha_{breg} I_{region(j)=region(b_i)} - I_{k \neq j} M_{i,j,k,t} + \epsilon_{i,k,t} \end{aligned}$$

where moving costs enter with an indicator for whether the individual chooses a location other than their current location ($I_{k\neq j}$). The idiosyncratic component of utility ($\epsilon_{i,k,t}$) and the unobservable location attribute (ξ_j) both enter linearly separable.

As the effects of the observable and unobservable fixed location attributes are not separately identified, I collapse all fixed location attributes into a mean flow utility, or overall quality of life term, δ_j, rewriting utility as:

$$\begin{aligned} u_{i,j,k,t} + \epsilon_{i,k,t} &= \alpha_{inc} inc_{i,j,t} + z'_{j,t}\alpha_z + \delta_j \\ &\quad + \alpha_{bmsa} I_{j=b_i} + \alpha_{breg} I_{region(j)=region(b_i)} - I_{k \neq j} M_{i,j,k,t} + \epsilon_{i,k,t} \end{aligned}$$

where $\delta_j = \chi'_j \alpha_\chi + \xi_j$.

2.4 State Space

The vector of state variables for individual i at time t is denoted $x_{i,t}$. It describes the state of the world at time t and is comprised of all individual characteristics and location attributes that affect utility. The state vector for a given period t includes the individual's current, previous, and birth locations, their age, their current and previous location-specific match components of income, and the current values of location attributes (observed and unobserved) for all locations. The decision variable is equal to the individual's choice at time t and is denoted $d_{i,t}$. Thus, flow utility can be explicitly written as a function of $x_{i,t}$ and $d_{i,t}$:

$$u_t(x_{i,t}, d_{i,t}) \; + \; \epsilon(d_{i,t})$$

I assume that the transition of the state is Markovian, so that $x_{i,t+1}$ depends on $x_{i,t}$ and $d_{i,t}$ only; no additional information is gained by knowing $x_{i,t-1}$. The transition probability of the state vector $x_{i,t}$ is denoted $q(x_{i,t+1}|x_t, d_{i,t})$.

Birth location is fixed and, along with time-invariant location attributes, does not transition. Age, current location, and prior location transition deterministically; in period $t+1$ an agent is two years older, their current location is given by $d_{i,t}$, and, if $d_{i,t}$ necessitates a move, the period t current location will become the period $t+1$ prior location. An individual's location-specific match component of income evolves stochastically in that the value stays fixed if the individual does not move, the value reverts to a known value if the individual returns to their prior location, and, if $d_{i,t}$ involves a move to a new location, the individual gets a draw from the known distribution of match components. Time-varying location attributes evolve stochastically; individuals have beliefs about the distributions from which future amenity values will be drawn. I assume rational expectations, so these distributions are the observed distributions.

The potential size of the state space, or the total number of values state vectors can take on, is the limiting factor in the traditional, full-solution method of estimation. As the number of state variables increases, the size of the state space grows exponentially. The full-solution method, described by Rust (1987), quickly becomes infeasible as the value function needs to be evaluated at every possible combination of the state variables.[16] The two-step estimation method circumvents this problem.

[16]In practice, I take 100 draws for each location in each time period for the five continuous variables – mean wage, crime, housing price, pollution, and population. This implies an approximate state space size of 1.12E+184.

2.5 Value Functions

In a world with significant costs of migration, rational agents will not be myopic in their location decisions. Individuals will consider the future stream of utility associated with choosing a new location today, as opposed to the behavior described by a static model where individuals care only about current flow utility. It is possible to write lifetime utility, $U(\mathbf{x}, \epsilon, \mathbf{d})$, as the suitably discounted sum of per-period utilities. Thus, individuals choose a decision rule based on:

$$\max_{\mathbf{d}^*=(d_{i,1};\ldots;d_{i,T})} E_{\mathbf{d}}\Big[U(\mathbf{x}, \epsilon, \mathbf{d})\Big] \quad = \quad \max_{\mathbf{d}^*=(d_{i,1};\ldots;d_{i,T})} E_{\mathbf{d}}\Bigg[\sum_{t=1}^{T} \beta^{t-1} \cdot \Big(u_t(x_{i,t}, d_{i,t}) + \epsilon(d_{i,t})\Big)\Bigg]$$

where β is the discount factor.[17]

In the lifetime optimization problem current decisions affect both current-period utility and future-periods' utility through the current decision's effect on future states. By assuming (i) that ϵ is i.i.d. over time and (ii) that the evolution of the state is Markovian, I am limiting the patterns of dependence in the dynamic process. These, along with the additive separability of flow utility, are the basic assumptions of Rust (1987). This allows me to write the following Bellman equation:

$$V_t(x_{i,t}, \epsilon(d_{i,t})) \quad = \quad \max_{d_{i,t} \in J}[v_t(x_{i,t}, d_{i,t}) + \epsilon(d_{i,t})]$$

where:

$$v_t(x_{i,t}, d_{i,t}) \quad = \quad u_t(x_{i,t}, d_{i,t}) \; + \; \beta \int \sum_{x_{i,t+1}} V_{t+1}(x_{i,t+1}, \epsilon(d_{i,t+1}))\; q(x_{i,t+1}|x_{i,t}, d_{i,t})\; dF(\epsilon(d_{i,t+1}))$$

An individual receives flow utility this period, $u_t(x_{i,t}, d_{i,t})$, based on their current state and current location decision, as well as the discounted stream of future payoffs associated with

[17] In practice, I set β to 0.9025, or $(0.95)^2$, to account for the biennial nature of the data.

that decision, $\beta \int \sum_{x_{i,t+1}} V_{t+1}(x_{i,t+1}, \epsilon(d_{i,t+1}))\, q(x_{i,t+1}|x_{i,t}, d_{i,t})\, dF(\epsilon(d_{i,t+1}))$. The precise value of this future stream is unknown at time t, as $x_{i,t+1}$ and $\epsilon(d_{i,t+1})$ are only known in expectation.

Assuming that the idiosyncratic error term, $\epsilon(d_{i,t})$, is distributed i.i.d. Type 1 Extreme Value, $\beta \int \sum_{x_{i,t+1}} V_{t+1}(x_{i,t+1}, \epsilon(d_{i,t+1}))\, q(x_{i,t+1}|x_{i,t}, d_{i,t})\, dF(\epsilon(d_{i,t+1}))$ can be replaced with the familiar Logit inclusive value:

$$v_t(x_{i,t}, d_{i,t}) \;=\; u_t(x_{i,t}, d_{i,t}) \;+\; \beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Big(v_{t+1}(x_{i,t+1}, d_{i,t+1} = j)\Big)\Bigg]\; q(x_{i,t+1}|x_{i,t}, d_{i,t})$$

This is the main estimating equation of the model. The computational burden of the traditional solution method comes from the recursive nature of this equation; there is a v_t on the left-hand side and v_{t+1} on the right-hand side. Specifically, there are J number of v_{t+1} terms on the right-hand side, which describe the value of choosing any of the J locations in period $t+1$.

An insight of this paper is that the econometric tools developed in Arcidiacono and Miller (2011) can be naturally applied to dynamic sorting models. A first stage of estimation involves estimating conditional choice probabilities as well as transition probabilities. Using these first stage estimates, the conditional value function can be reduced to an estimating equation that is linear in the structural parameters and these parameters can be estimated using a straight-forward Logit procedure. The Appendix provides further details.

3 Data

3.1 Choice Set

As the spatial determinants of national migration decisions are the focus of my research, the geographic definition of the choice set is critical; if the locations are chosen on too fine a level, there will be few observed moves to and from each location; however, if the locations are chosen on too broad a level (such as states or regions), it becomes impractical to define spatially varying amenities or to infer preferences for them. In a national model of migration, metropolitan areas seem to be a natural unit of geography. However, with standard panel datasets, sparse data becomes an issue when including all 362 recognized metropolitan areas.[18]

Although my methodology could handle a much larger choice set, I use the 50 largest metropolitan statistical areas (MSAs) in the contiguous United States, as represented in my primary data source, the National Longitudinal Survey of Youth 1979 (NLSY79).[19] To limit attrition from the panel, I additionally include 9 "catch-all" locations, defined by non-MSA Census divisions. Thus, the final choice set is comprised of 59 locations. The 50 MSAs represent approximately 39% of the U.S. population in 2000 and approximately 60% of my NLSY79 sample. The geographic distribution of these locations can be seen in Figure 1.

3.2 Individual Data

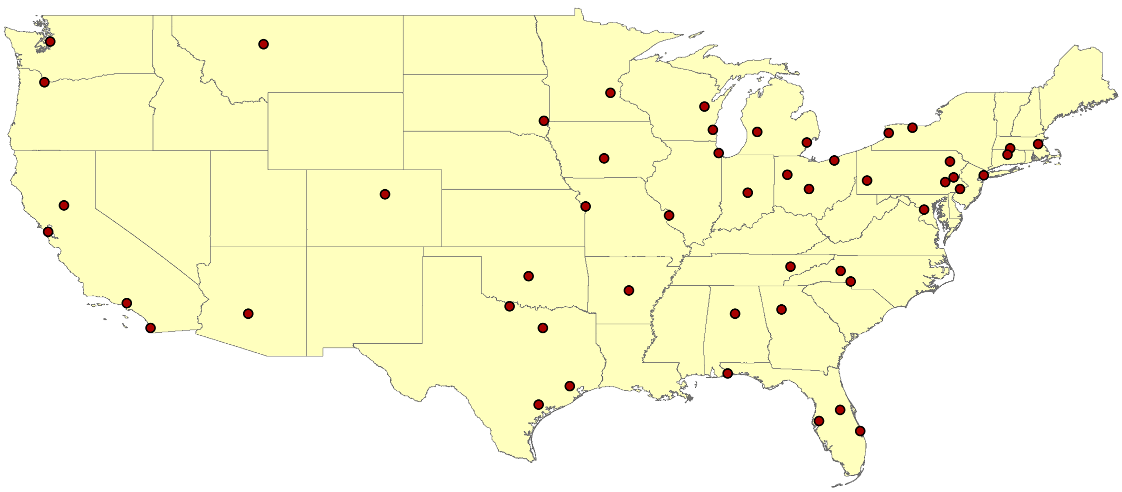
The primary dataset I employ is the National Longitudinal Survey of Youth, 1979 panel. This panel dataset follows over 12,000 individuals aged 14 to 22 in the first round of interviews in 1979. Beginning in 1994, the Bureau of Labor Statistics switched from an annual to a biennial interview basis. I use biennial interviews from 1986 to 2004, with individuals aged 21 to 29 in 1986 and 39 to 45 in 2004 (the panel is unbalanced). The restricted-access geocode supplement of the NLSY79 provides detailed information (county-level) on the individual's geographic location, including birth location, in addition to the detailed demographic data given by the public-access files. I aggregate this county-level data to the level of the MSA, as defined by the U.S. Census Bureau in 1990.

Following Kennan and Walker, I cut the sample to include a homogenous sample of high school

[18]In future work, the Longitudinal Employer - Household Dynamic (LEHD) panel dataset, which contains basic demographic, income, and detailed geographic data for millions of individuals based on Unemployment Insurance records, could be used to estimate a model with all 362 MSAs.

[19]While the largest MSAs in the NLSY include the most recognized cities in the U.S. (*e.g.*, New York, Chicago, Los Angeles), there are a number of smaller cities that contain a large number of NLSY respondents relative to their size – *e.g.*, Hickory-Morgantown-Lenoir, NC, Lima, OH, Great Falls, MT.

Figure 1: Metropolitan Areas of the Choice Set



educated white males, with no military or college experience.[20] In addition, I drop individuals who were born outside of the contiguous United States. Finally, I drop observations following a missed interview by the individual. This leaves me with a final sample of 1,123 individuals with 7,176 observations and 478 observed moves between locations in the choice set.[21]

[20]One may worry that if college attendance rates were changing over the period, differing cohort effects would limit the homogeneity of the panel. However, national college attendance rates for 25 to 29 year old white males remained fairly constant over the age-relevant period of 1982 to 1993. Please see the U.S. Health and Human Services Department report findings at http://aspe.hhs.gov/hsp/97trends/ea1-6.htm for more detail.

[21]Although this is a relatively small number of moves compared to static analyses, which use cross-sectional data, it is more than twice the number of moves observed by Kennan and Walker (213). Future access to richer panel datasets will increase this number significantly.

In particular, I use information on individuals' birth locations and migration decisions, age, AFQT score,[22] and income (defined as the sum of wage and business income).

Summary statistics are given in Table 1 (the sample is comprised of 7,176 person-year observations with 1,123 distinct individuals).

Table 1: Summary Statistics for the NLSY79 Sample

Variable	Obs.	Mean	Std. Dev.
Age	7176	31.88	5.63
Income (2000 dollars)	7176	32,997.26	22,572.47
Live in MSA of Birth	7176	0.68	0.46
Number of Moves	1123	0.49	0.93
Number of Interviews	1123	7.93	2.18
AFQT	1123	54.08	25.91

Finally, in the estimation of location-specific incomes, I supplement the NLSY79 sample with a sample of Current Population Survey (CPS) respondents. This sample is also trimmed to white, high school educated males with no college or military experience. In addition, I trim to U.S.-born individuals of the relevant age cohort.[23]

3.3 Location Attributes

As any time-invariant characteristics would be absorbed by the location-specific fixed effect, δ_j, I collect a panel of annual time-varying data comprised of crime, median housing price, air pollution, and population variables for each MSA over the relevant period. I construct a matrix of distances using the "Great Circle" algorithm[24] with geographic coordinate data taken from the U.S. Census Bureau.

[22] The Armed Forces Qualifying Test is an general aptitude test given to all respondents of the NLSY79.

[23] The five MSAs of Fort Pierce-Port St. Lucie, FL, Great Falls, MT, Lima, OH, Victoria, TX, and Wichita Falls, TX do not have full CPS coverage over the relevant sample period. In these cases, I include individuals residing within the larger Public Use Microdata Area (PUMA).

[24] This is also known as the Haversine formula (Sinnott, 1984).

3.3.1 Crime

The crime data I employ are taken from the Federal Bureau of Investigation's annual report entitled "Crime in the United States." These annual reports include the total reported violent and property crime incidents,[25] as reported by over 17,000 law enforcement agencies across the United States. These data are given at the city-, MSA-, state-, region-, and national-level on an annal basis beginning in 1930.

Following Savageau and D'Agostino (2000), I compute a measure of crime that collapses the violent crimes and property crimes into a single index for each MSA in each year. As it is assumed that individuals respond to violent crimes more than property crimes, the index weights violent crimes ten times as heavily. I use a slightly modified version of the index by calculating it in per-capita terms.[26] Specifically, I use:

$$crime_{j,t} = \frac{violent_{j,t} + \frac{property_{j,t}}{10}}{population_{j,t}}$$

3.3.2 Median House Price

I include a measure of median house price to control for differences in cost of living both across MSAs and across time periods. Using data from the Office of Federal Housing Enterprise Oversight's (OFHEO) Housing Price Index (HPI) and estimates from the National Association of Realtors, I construct a panel of median house prices. The HPI, which sets 1995 as a base year for each MSA, reports a weighted, repeat-sales index for single-family homes within each MSA going back as far as 1975. I translate these appreciation figures into annual median house prices using a 2004 cross-section of median single-family house prices provided by the National Association of Realtors. Finally, all prices are converted to 2000 dollars using the Consumer Price Index.

[25] Property crimes include burglary, larceny-theft, and motor vehicle theft. Violent crimes include murder, manslaughter, forcible rape, robbery, and aggrevated assault. As not all reporting agencies collect data on rape, I omit it from my aggregate measure of violent crime.

[26] I use the FBI's measure of population here to ensure a proper per-capita measure.

3.3.3 Air Pollution

The measure of air pollution I use is the ambient concentration of particulate matter. The Environmental Protection Agency provides annual emissions figures in their National Emissions Inventory from nearly $6,000$ different sources. The data I use are county-level estimates of particulate matter concentration (PM10) generated from source data on total particulates and sulfur dioxide, a PM10 precursor, using the source to county receptor matrix of the Climatological Regional Dispersion Model.[27] I compute annual MSA-level pollution estimates by aggregating the county-level data to the 1990 geographical definition of MSA boundaries.

3.3.4 Population

Finally, I collect annual county-level population estimates from the U.S. Census Bureau. I compute annual MSA-level population estimates by aggregating the county-level data to the 1990 geographical definition of MSA boundaries. Thus, this figure captures changes in population density without the confounding effects of changes to the MSA boundaries.[28]

Summary statistics of the location attributes are presented in Table 2.

4 Estimation

The computational burden of the model is greatly reduced by the two-step estimation strategy described in Arcidiacono and Miller (2011). The first step includes all of the estimation procedures that are performed outside of the dynamic routine. In particular, I estimate incomes, transition probabilities of the variables that evolve stochastically, and conditional choice probabilities. In the second step, I estimate the remaining structural parameters of the utility function, taking the first-stage estimates as given.

[27] The county-level dataset was created by Nat Keohane who generously provided it.

[28] The geographic boundaries of some U.S. counties changed over the period of my sample. However, these changes did not affect the definition of any of the fifty chosen MSAs.

Table 2: Summary Statistics for the Location Attributes

Variable	Obs.	Mean	Std. Dev.
Crime (per capita)	50	0.0116	0.0046
Median House Price (hundred thousand 2000 dollars)	50	130.55	64.58
Pollution (PM10 in $\mu g/m^3$)	50	27.79	7.60
Population (millions)	50	2.02	2.19
New England	50	0.07	0.25
Middle Atlantic	50	0.15	0.36
East North Central	50	0.17	0.38
West North Central	50	0.10	0.30
South Atlantic	50	0.15	0.36
East South Central	50	0.05	0.22
West South Central	50	0.12	0.33
Mountain	50	0.07	0.25
Pacific	50	0.12	0.33

4.1 First Step

4.1.1 Income

In the estimation of location-specific mean income levels, I supplement the NLSY79 sample with data taken from the CPS over the relevant periods, locations, and cohort. Using the CPS, I regress reported income on a set of age and MSA dummies, yielding estimates of $f(age_{i,t})$ and $\mu_{j,t}$:

$$inc^{CPS}_{i,j,t} \;=\; f(age_{i,t}) \;+\; \mu_{j,t} + \epsilon^{CPS}_{i,j,t}$$

I am then able to take these estimates to the NLSY79 sample, where I regress the deviations from predicted income on individual characteristics:[29]

[29] As the sample is cut to a homogenous sample, I follow Kennan and Walker in using AFQT score as the

$$inc_{i,j,t} \;-\; \hat{f}(age_{i,t}) \;-\; \hat{\mu}_{j,t} \;=\; \omega_i'\gamma \;+\; (\theta_{i,j} + \eta_i + e_t)$$

allowing me to isolate the unobserved, idiosyncratic portion of income:

$$inc_{i,j,t} \;-\; \hat{f}(age_{i,t}) \;-\; \hat{\mu}_{j,t} \;-\; \omega_i'\hat{\gamma} \quad = \quad \theta_{i,j} + \eta_i + e_t \quad = \quad \Omega_{i,j(t)}$$

I find population-wide estimates of standard deviations for $\theta_{i,j}$, η_i, and e_t to be (in 2000 dollars):[30]

$$\begin{aligned} \hat{\sigma}_\eta &= \$7,073.40 \\ \hat{\sigma}_\theta &= \$15,499.21 \\ \hat{\sigma}_e &= \$13,869.55 \end{aligned}$$

and (also in 2000 dollars) the estimated points of support of the distribution of match components in each location to be : ($\mu_{j,t}$ - \$8,775.40, $\mu_{j,t}$, $\mu_{j,t}$ + \$8,775.40), where $\mu_{j,t}$ is the mean income in location j in period t.

4.1.2 Transition Probabilities

Also in the first stage, I estimate the transitions for each of the five stochastic variables of the model (mean income, crime rate, housing price, pollution level, and annual population) by assuming AR-1 processes. To maximize use of the data, I pool observations across locations and regress the current value of these variables on a constant term, the lagged value of the variable, a set of Census division dummies, and a set of dummies describing the MSA's quintile of 1990 population. This yields a set of predicted values for each variable in each period. The residuals from these regressions are used to define the distributions from which each expected value is drawn.

sole individual characteristic.

[30]These estimates imply that approximately 41% of the total variance of earnings can be attributed to the transitory component of income. This figure is larger than that estimated in Kennan and Walker (33%). However, both are in line with the previous research of Gottschalk and Moffitt (1994).

4.1.3 Conditional Choice Probabilities

Finally, the conditional choice probabilities are recovered before proceeding to the dynamic estimation problem. In an ideal world, a first-stage estimate of these probabilities would be found using a fully non-parametric method, such as a "bin" estimator, where the probability of choosing a particular location could be estimated as the fraction of individuals who choose that location (conditional on being at a particular point in the state space). However, these probabilities need to be calculated at each and every possible state of the world, with this number approaching infinity in the current specification. Thus, I estimate a reduced-form approximation of the conditional choice probabilities, specifying a flexible functional form. Within the Logit framework, I include a sizeable number of higher-order and interaction terms, estimating over 40 parameters in addition to a set of location fixed-effects, which enter linearly.[31]

4.2 Second Step

The first step yields transitions of the state variables and the conditional probabilities of choosing any particular location, given any particular state. These estimates are used in the second-step, where the remaining structural parameters of the specified utility function are estimated.

Using the insights of Arcidiacono and Miller (described in Section 2.5 and the Appendix), the second step is estimated as a linear-in-parameters Logit. As the location fixed-effects (the δ's) enter linearly, it is possible to employ a Berry (1994) contraction mapping to estimate them, further limiting the number of parameters over which I need to search in a Maximum Likelihood routine. The log-likelihood is given by:

$$L(\alpha, \delta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{J} log\Big(\frac{exp(v_t(x_{i,t}, d_{i,t} = k))}{\sum_{j=1}^{J} exp(v_t(x_{i,t}, d_{i,t} = j))}\Big) \cdot I_{[d_{i,t}=k]}$$

[31] I estimate the fixed effects using a Berry (1994) contraction mapping inside the maximum likelihood routine.

where $I_{[d_{i,t}=k]}$ is an indicator equal to one if individual i chooses location k in period t.

The state vector is updated in each period using the estimated transitions from the first step. In practice, I integrate out over the distributions of the stochastic variables by taking 100 draws from their distributions, which are generated using the residuals from the AR(1) regressions.

Finally, a willingness to pay for a marginal change in a particular attribute can be easily calculated as the coefficient on the attribute divided by the coefficient on income, as the variables enter linearly.[32]

5 Results

Table 3 reports the parameter estimates from the second-stage dynamic estimation. Income and housing price are given in tens of thousands of 2000 dollars, population is reported in millions of individuals, distance is measured in thousands of miles, and pollution is given in $\frac{\mu g/m^3}{100}$. Estimates have the predicted sign and most coefficients are significant at the 5% level. Estimated moving costs are large, but consistent with Kennan and Walker (2011).[33]

These results imply a willingness to pay to avoid a one unit (one $\mu g/m^3$) increase in air pollution of \$198.22 in 2000 dollars. Reported in 1982-1984 dollars, this figure is \$114.96. In a static hedonic analysis, Chay and Greenstone (2005) find a comparable willingness to pay of approximately \$22.00 (also in 1982-1984 dollars). However, their model ignores both forward-looking behavior and the costs associated with migration. Each of these omissions

[32]Considering larger changes, one would also need to account for the effects on expected future flow utilities in the calculation of compensating income variations. In this paper, I focus only on willingness to pay for marginal changes in air pollution, dealing with large changes in ongoing work.

[33]The cost of a one-mile move by a twenty-year old individual would be $\$335,714.30$. The comparable figure (converted to 2000 dollars) in Kennan and Walker's analysis is a slightly larger $\$363,085.80$. As discussed in Kennan and Walker, one would expect this cost estimate to be large as it captures the cost an individual would expect to face if forced to move to an arbitrary location in an arbitrary time period. By allowing an individual to choose the best location, but still conditional on being forced to move in an arbitrary time period, would reduce costs by \$128,211 and is calculated as $log(J-1)/\alpha_{inc}$.

Table 3: Results from Dynamic model

Log Likelihood	-3097.1626	
	Coefficient	t-statistic
Moving Cost - Fixed Cost	-3.8971	-40.8531
Moving Cost - Distance	-1.9565	-11.6354
Moving Cost - Distance^2	0.3908	5.1235
Moving Cost - Same Region	0.8588	12.2621
Moving Cost - Age	-0.1019	-24.9054
Moving Cost - Last MSA	0.2111	2.0557
Income	0.1408	3.9983
Living in MSA of Birth	0.3550	18.4645
Living in Region of Birth	0.0001	0.0038
Housing Price	-0.0337	-0.0056
Crime	-0.1853	-0.0389
Pollution	-0.2791	-1.9906
Population	-1.3334	-2.3551

could lead to downward-biased estimates.

It is interesting to note that the coefficient on crime is not significant; this is not surprising, as crime is often extremely localized within an MSA and individuals can choose to avoid it (to a certain degree) based on neighborhood choice. However, this model could be used to derive a willingness to pay for any time-varying attribute, such as crime, in the same manner as pollution. The willingness to pay for a fixed attribute could be similarly estimated by decomposing the location specific fixed effects into observable and unobservable components via OLS regression:

$$\delta_j \ = \ \chi_j' \alpha_\chi \ + \ \xi_j$$

I am unable to do so in the current application only because of the limitations imposed by the current choice set size (*i.e.*, $J = 59$) that results from using NLSY data.

To further examine the consequences of ignoring either dynamics or psychological ties to certain locations, I estimate a static specification and a dynamic specification that ignores ties to birth location using the NLSY79 sample. Results from these specifications are presented in Tables 4 and 5 respectively.

Table 4: Results from Static model

Log Likelihood	-3118.0622	
	Coefficient	t-statistic
Moving Cost - Fixed Cost	-1.0863	-28.7229
Moving Cost - Distance	-2.0204	-12.0548
Moving Cost - Distance^2	0.5622	7.3940
Moving Cost - Same Region	0.6562	9.4096
Moving Cost - Age	-0.0927	-20.325
Moving Cost - Last MSA	-0.9613	-4.4994
Income	0.3167	5.1757
Living in MSA of Birth	3.0761	84.2343
Living in Region of Birth	0.0023	0.0354
Housing Price	-0.0017	-0.2552
Crime	-0.9525	-0.1378
Pollution	-0.2401	-1.8793
Population	-1.9054	-1.5985

In the static specification, agents choose a location based on the current level of attributes. Thus, the value function is comprised of flow utility and a choice-specific shock only; agents ignore the continuation value associated with current choice. Note that the static specification does, however, still allow individuals to have psychological ties to their birth locations.

The results from the static model imply that individuals sort on the basis of both current and expected future levels of pollution (in particular, choosing locations where pollution levels are expected to fall). This leads to estimates of willingness to pay that are downward-biased in the static framework. In particular, this model yields a willingness to pay for air quality that is 62% lower than in the dynamic specification.

Finally, it is also interesting to note that in the static model, the sign on "Last MSA" has flipped; without dynamic concerns, the importance of a prior location to agents is unclear, as opposed to the dynamic model, where agents consider the possibility of future return migration.

Table 5: Results from Dynamic model with no ties to Birth Location

Log Likelihood	-3127.3720	
	Coefficient	t-statistic
Moving Cost - Fixed Cost	-4.1609	-0.6533
Moving Cost - Distance	-1.9292	-0.4152
Moving Cost - Distance^2	0.3995	0.0298
Moving Cost - Same Region	0.9616	0.2322
Moving Cost - Age	-0.1038	0.0066
Moving Cost - Last MSA	0.0020	0.0002
Income	0.3781	0.0041
Housing Price	-0.0337	-0.0056
Crime	0.7036	3.3274
Pollution	-0.3355	-0.1188
Population	-4.8837	-2.6464

In the next specification, agents are both forward-looking and experience costs associated with migration in the traditional sense (*i.e.*, there is a fixed cost moving and costs increase with age and distance between locations). However, often over-looked are the emotional ties to one's "home," or birth location. This specification ignores individual's preference for living in the MSA or region of their birth. Bayer, Keohane, and Timmins (2009) show using 2000 Census data, that more than sixty percent of household heads are currently living in the Census division of their birth, with a higher number for high-school graduates, such as those used in this application.

If individuals have ties to locations with relatively high levels of pollution because of birth, a naïve model that ignores these ties would, by omission, interpret them as a preference for birth location attributes. This would lead estimates of willingness to pay for air quality to be biased downwards. There is a positive correlation between birth location and pollution (the correlation coefficient is 0.22, measured in the first year of my panel, 1986), implying that this source of bias may be a concern.

The results presented in Table 5 imply that individuals have ties to birth locations with relatively worsening air quality, as identification comes off of variation through time. These willingness to pay results are again substantially lower than in the base model; $52.21 versus $114.96 in 1982-1984 dollars. In addition, the omission of these ties leads many of the other coefficients to take on the wrong signs.

The goal of the preceding exercise was to illustrate the important role played by dynamics (both costly migration and forward-looking behavior) in non-market valuation. In addition, since I have recovered the parameters of the utility function that determine migration behavior, I can also use the estimated model to perform counterfactual simulations of individual migration decisions. Of particular interest is the role of amenities in the migration patterns observed over the last twenty years in the United States. Isolating the relative weights on income prospects and on non-pecuniary amenity values in migration trends, simulations show that amenities play a significant role (as opposed to falling wages in declining industries alone) in the recent flows out of Northeast and Midwest and into the South, Southwest, and West of the United States. Driving this finding is the result that income and amenity values are positively correlated in metropolitan areas.

In addition, these simulations can be used to examine the ex post distributions of the individual- and location-specific match components of income. Following the basic insights of Roy (1951), observed income distributions will have higher means than offered income distributions, as individuals choose locations that offer the highest wages. Although I capture only the partial equilibrium effects while holding location-specific mean incomes fixed, results show that the ex post distribution of location- and individual-specific match components has both a larger mean and a larger variance in the counterfactual world in which amenities do not enter utility. This implies that inter-location convergence of incomes could be in part explained by geographic differences in amenities.

6 Conclusion

In this paper, I estimate a dynamic model of location choice where the choice set is defined at the level of the metropolitan area. I employ the two-step, Hotz and Miller -style estimation routine developed in Arcidiacono and Miller (2011), which permits me to include a rich descriptive set of local attributes, including a set of location fixed effects. The inclusion of these amenity values not only allows for a more realistic specification of individual utility, but also allows for the use of a dynamic migration model in a non-market valuation exercise.

In the first step of the estimation routine, I flexibly estimate conditional choice probabilities, transition probabilities of the state variables, and individual incomes. In the second step, I estimate the structural parameters of individuals' utility functions. The overall computational burden of the estimator is low, with the second step estimated as a linear-in-parameters Logit. I estimate the model using a homogenous sample of white, high-school educated males from the NLSY79 panel, over the period 1986 to 2004.

My particular application recovers the willingness to pay for clean air. Using a panel of annual air pollution data for each metropolitan area, I find the willingness to pay (in 2000 dollars) to be \$198.22 for a one unit decrease in annual pollution (a one $\mu g/m^3$ decrease in particulate matter). When calculated at mean income and mean pollution, this corresponds to a willingness to pay elasticity with respect to air quality of approximately 0.17. I find evidence of a downward bias in estimates produced using either a static sorting model or a dynamic model that ignores the emotional ties to individuals' birth locations. This indicates that individuals are both forward-looking with respect to location amenities and have ties to relatively polluted areas.

References

Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75(1), 1–54.

Arcidiacono, P., and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," 79(6), 1823–1867.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

Bajari, P., and M. Kahn (2005): "Estimating Housing Demand With An Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics*, 23(1), 20–33.

Barro, R., and X. Sala-i-Martin (1992): "Convergence," *The Journal of Political Economy*, 100(2), 223–251.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, 58(1), 1–14.

Bayer, P., S. Khan, and C. Timmins (2011): "Nonparametric Identification and Estimation in a Roy Model with Common Nonpecuniary Returns," *Journal of Business & Economic Statistics*, 29(2), 201–215.

Bayer, P., R. McMillan, and K. Rueben (2004): "An Equilibrium Model of Sorting in an Urban Housing Market," NBER Working Paper #10865.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differation," *Rand Journal of Economics*, 25, 242–262.

Blomquist, G., M. Berger, and J. Hoehn (1988): "New Estimates of Quality of Life in Urban Areas," *The American Economic Review*, 78(1), 89–107.

Chay, K., and M. Greenstone (2005): "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113(2), 376–424.

Dahl, G. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367–2420.

Epple, D., and H. Sieg (1999): "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107, 645–681.

Gemici, A. (2011): "Family Migration and Labor Market Outcomes," mimeo, New York University.

Gottschalk, P., and R. Moffitt (1994): "The Growth of Earnings Instability in the US Labor Market," *Brookings Papers on Economic Activity*, 1994(2), 217–272.

Hicks, J. (1932): "The Theory of Wages," London: MacMillan.

Hoehn, J., M. Berger, and G. Blomquist (1987): "A Hedonic Model of Interregional Wages, Rents, and Amenity Values," *Journal of Regional Science*, 27(4), 605–620.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Kennan, J. (2004): "A Note on Approximating Distribution Functions," mimeo, University of Wisconsin-Madison.

Kennan, J., and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

Kuminoff, N. V. (2012): "Partial Identification of Preferences in a Dual-Market Sorting Equilibrium," mimeo, Arizona State University.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Samples)," *RAND Journal of Economics*, 38(2), 373–399.

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *The Review of Economic Studies*, 75(3), 901–928.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *The Journal of Political Economy*, 90(6), 1257–1278.

Rosen, S. (1979): “Wage-Based Indexes of Urban Quality of Life,” *Current Issues in Urban Economics, ed. P. Mieszkowski and M. Straszheim*, pp. 74–104, Baltimore, MD: Johns Hopkins University Press.

Roy, A. (1951): “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, 3(2), 135–146.

Rust, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55(5), 999–1033.

Savageau, D., and R. D’Agostino (2000): “Places Rated Almanac,” CA: IDG Books Worldwide, Inc.

Sieg, H., V. Smith, H. Banzhaf, and R. Walsh (2004): “Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods,” *International Economic Review*, 45(4), 1047–1077.

Sinnott, R. (1984): “Virtues of the Haversine,” *Sky and Telescope*, 68(2), 158.

Sjaastad, L. (1962): “The Costs and Returns of Human Migration,” *The Journal of Political Economy*, 70(5), 80–93.

Smith, V., and J. Huang (1995): “Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models,” *The Journal of Political Economy*, 103(1), 209–227.

Tiebout, C. (1956): “A Pure Theory of Local Government Expenditures,” *Journal of Political Economy*, 64(5), 414–424.

Appendix – Model and Estimation Details

Beginning with the value function:

$$v_t(x_{i,t}, d_{i,t}) \;=\; u_t(x_{i,t}, d_{i,t}) \;+\; \beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = j)\Bigg)\Bigg]\; q(x_{i,t+1}|x_{i,t}, d_{i,t})$$

it is possible to multiply and divide the inclusive value term by the value of choosing a particular location h in period $t+1$, given that $d_{i,t} = k$:

$$\begin{aligned} v_t(x_{i,t}, d_{i,t} = k) \;&=\; u_t(x_{i,t}, d_{i,t} = k) \\ +\beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = j)\Bigg) & \frac{exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg)}{exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg)}\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \end{aligned}$$

Separating out the $exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg)$ in the numerator and dividing through by the
$exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg)$ in the denominator yields:

$$\begin{aligned} v_t(x_{i,t}, d_{i,t} = k) \;&=\; u_t(x_{i,t}, d_{i,t} = k) \\ +\, \beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1} = j) - & v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\ &+ \beta \sum_{x_{i,t+1}} \Bigg[v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg] q(i, x_{t+1}|x_{i,t}, d_{i,t} = k) \end{aligned}$$

The third right-hand side term in the above equation,

$$\beta \sum_{x_{i,t+1}} \Bigg[v_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \;\;,$$

or the expected value of choosing location h in period $t+1$, can be written as the sum of a period $t+1$ flow utility and the associated continuation value. Similar to the previous normalization, it is possible to normalize this continuation value relative to choosing some location g in period $t+2$:

$$\begin{aligned}
v_t(x_{i,t}, d_{i,t}=k) &= u_t(x_{i,t}, d_{i,t}=k) \\
&+ \beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1}=j) - v_{t+1}(x_{i,t+1}, d_{i,t+1}=h)\Bigg)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \beta \sum_{x_{i,t+1}} \Bigg[u_{t+1}(x_{i,t+1}, d_{i,t+1}=h)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+2}(x_{i,t+2}, d_{i,t+2}=j) - v_{t+2}(x_{i,t+2}, d_{i,t+2}=g)\Bigg)\Bigg] \\
&\quad \cdot q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \Bigg[v_{t+2}(x_{i,t+2}, d_{i,t+2}=g)\Bigg] q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k)
\end{aligned}$$

Finally, it is possible to again expand the final right-hand side value function into a flow utility and a continuation value, and normalize the continuation value by the value of choosing some location m in period $t+3$:

$$
\begin{aligned}
v_t(x_{i,t}, d_{i,t}=k) \quad &= \quad u_t(x_{i,t}, d_{i,t}=k) \\
&+ \ \beta \sum_{x_{i,t+1}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+1}(x_{i,t+1}, d_{i,t+1}=j) - v_{t+1}(x_{i,t+1}, d_{i,t+1}=h)\Bigg)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \ \beta \sum_{x_{i,t+1}} \Bigg[u_{t+1}(x_{i,t+1}, d_{i,t+1}=h)\Bigg] q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \ \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+2}(x_{i,t+2}, d_{i,t+2}=j) - v_{t+2}(x_{i,t+2}, d_{i,t+2}=g)\Bigg)\Bigg] \\
&\quad \cdot q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \ \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \Bigg[u_{t+2}(x_{i,t+2}, d_{i,t+2}=g)\Bigg] \\
&\quad \cdot q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \ \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} ln\Bigg[\sum_{j=1}^{J} exp\Bigg(v_{t+3}(x_{i,t+3}, d_{i,t+3}=j) - v_{t+3}(x_{i,t+3}, d_{i,t+3}=m)\Bigg)\Bigg] \\
&\quad \cdot q(x_{i,t+3}|x_{i,t+2}, d_{i,t+2}=g) q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k) \\
&+ \ \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} \Bigg[v_{t+3}(x_{i,t+3}, d_{i,t+3}=m)\Bigg] \\
&\quad \cdot q(x_{i,t+3}|x_{i,t+2}, d_{i,t+2}=g) q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1}=h) q(x_{i,t+1}|x_{i,t}, d_{i,t}=k)
\end{aligned}
$$

While these expansions appear to have complicated the estimation equation, they actually allow for an important simplification. Consider the probability of choosing a particular location c in a given time τ. Given the Logit framework, this probability can be written as:

$$
\begin{aligned}
Pr(d_{i,\tau}=c|x_{i,\tau}) \quad &= \quad \frac{exp\Bigg(v_\tau(x_{i,\tau}, d_{i,\tau}=c)\Bigg)}{\sum_{j=1}^{J} exp\Bigg(v_\tau(x_{i,\tau}, d_{i,\tau}=j)\Bigg)} \\
&= \quad \frac{1}{\sum_{j=1}^{J} exp\Bigg(v_\tau(x_{i,\tau}, d_{i,\tau}=j) - v_\tau(x_{i,\tau}, d_{i,\tau}=c)\Bigg)}
\end{aligned}
$$

or as:

$$Pr(d_{i,\tau} = c|x_{i,\tau})^{-1} = \sum_{j=1}^{J} exp\Big(v_\tau(x_{i,\tau}, d_{i,\tau} = j) - v_\tau(x_{i,\tau}, d_{i,\tau} = c)\Big)$$

The right-hand side of the above equation is equivalent to the right-hand side normalized continuation values in the choice-specific value function. Thus, it is possible to write the choice-specific value function as:

$$\begin{aligned}
v_t(x_{i,t}, d_{i,t} = k) \;\; &= \;\; u_t(x_{i,t}, d_{i,t} = k) \\
&+ \; \beta \sum_{x_{i,t+1}} ln\Big[Pr(d_{i,t+1} = h|x_{i,t+1})^{-1}\Big] q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&+ \; \beta \sum_{x_{i,t+1}} \Big[u_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Big] q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&+ \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\Big[Pr(d_{i,t+2} = g|x_{i,t+2})^{-1}\Big] \\
&\quad \cdot q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h) q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&+ \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \Big[u_{t+2}(x_{i,t+2}, d_{i,t+2} = g)\Big] \\
&\quad \cdot q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h) q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&+ \; \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} ln\Big[Pr(d_{i,t+3} = m|x_{i,t+3})^{-1}\Big] \\
&\quad \cdot q(x_{i,t+3}|x_{i,t+2}, d_{i,t+2} = g) q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h) q(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&+ \; \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} \Big[v_{t+3}(x_{i,t+3}, d_{i,t+3} = m)\Big] \\
&\quad \cdot q(x_{i,t+3}|x_{i,t+2}, d_{i,t+2} = g) q(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h) q(x_{i,t+1}|x_{i,t}, d_{i,t} = k)
\end{aligned}$$

These future conditional choice probabilities, $Pr(d_{i,t}|x_{i,t})$, are estimated in a separate first-step. They can therefore be thought of as data in the above equation and replaced by

$\widehat{Pr}(d_{i,t}|x_{i,t})$. The transition probabilities, $q(x_{i,t+1}|x_{i,t}, d_{i,t})$, can be similarly recovered in a preliminary procedure and replaced by $\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t})$:

$$
\begin{aligned}
v_t(x_{i,t}, d_{i,t} = k) \;\; = \;\; & u_t(x_{i,t}, d_{i,t} = k) \\
& + \; \beta \sum_{x_{i,t+1}} ln\left[\widehat{Pr}(d_{i,t+1} = h|x_{i,t+1})^{-1}\right]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
& + \; \beta \sum_{x_{i,t+1}} \left[u_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\right]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
& + \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\left[\widehat{Pr}(d_{i,t+2} = g|x_{i,t+2})^{-1}\right] \\
& \quad \cdot \widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
& + \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \left[u_{t+2}(x_{i,t+2}, d_{i,t+2} = g)\right]\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
& + \; \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} ln\left[\widehat{Pr}(d_{i,t+3} = m|x_{i,t+3})^{-1}\right] \\
& \quad \cdot \widehat{q}(x_{i,t+3}|x_{i,t+2}, d_{i,t+2} = g)\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
& + \; \beta^3 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \sum_{x_{i,t+3}} \left[v_{t+3}(x_{i,t+3}, d_{i,t+3} = m)\right] \\
& \quad \cdot \widehat{q}(x_{i,t+3}|x_{i,t+2}, d_{i,t+2} = g)\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k)
\end{aligned}
$$

Now, the main estimating equation has been re-written to eliminate $J-1$ of the original J right-hand side value functions. Exploiting the "finite dependence" simplification of Arcidiacono and Miller will eliminate the remaining one. Before describing this final step, however, it is useful to consider the economic meaning of the remaining right-hand side value function ($v_{t+3}(x_{i,t+3}, d_{i,t+3} = m)$) – *i.e.*, the value of choosing location m in period $t+3$, when the individual's current ($t+2$ choice) location is g and prior ($t+1$ choice) location is h. The fact that the agent chose location k in period t does not affect the period $t+3$ decision, as "memory" only extends to the prior location. In this case, the dependence on initial choice can be broken after two periods, while in a model with no "memory," dependence could be broken after only one period.[34]

Using the fact that the location parameter in a Logit model is not identified, *i.e.*, that only relative values matter, it is possible to difference the above value function with respect to the value function associated with another choice (*i.e.*, estimate $(v_t(x_{i,t}, d_{i,t} = k) - v_t(x_{i,t}, d_{i,t} = c))$, where the future component of $v_t(x_{i,t}, d_{i,t} = c)$ has been expanded in the manner just described). With dependence on initial choice broken, the final right-hand side term has the same value in both cases. In addition, as the value of choosing and particular location in period $t + 3$ is independent of the period t choice, the probability of choosing a particular location m in period $t+3$ is also independent of the period t choice. Thus, the final two right-hand side terms will drop out when differenced. In general, the model must be expanded to one period further than "memory" allows; if an agent can "remember" current location and prior location, the value function must be extended three periods, as above.

The estimating equation therefore becomes:

[34]This would allow the methodology to be used on publicly available IPUMS data, which provides information on the individual's birth location and two other observed locations (*e.g.*, location in 1995 and in 2000). The latter information can be used to model one migration decision. The current specification requires data on birth location and at least three other observed locations, or two migration decisions.

$$
\begin{aligned}
v_t(x_{i,t}, d_{i,t} = k) - v_t(x_{i,t}, d_{i,t} = c) \;\; &= \;\; u_t(x_{i,t}, d_{i,t} = k) - u_t(x_{i,t}, d_{i,t} = c) \\
&+ \; \beta \sum_{x_{i,t+1}} ln\Big[\widehat{Pr}(d_{i,t+1} = h|x_{i,t+1})^{-1}\Big]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&- \; \beta \sum_{x_{i,t+1}} ln\Big[\widehat{Pr}(d_{i,t+1} = h|x_{i,t+1})^{-1}\Big]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = c) \\
&+ \; \beta \sum_{x_{i,t+1}} \Big[u_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Big]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&- \; \beta \sum_{x_{i,t+1}} \Big[u_{t+1}(x_{i,t+1}, d_{i,t+1} = h)\Big]\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = c) \\
&+ \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\Big[\widehat{Pr}(d_{i,t+2} = g|x_{i,t+2})^{-1}\Big]\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&- \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} ln\Big[\widehat{Pr}(d_{i,t+2} = g|x_{i,t+2})^{-1}\Big]\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = c) \\
&+ \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \Big[u_{t+2}(x_{i,t+2}, d_{i,t+2} = g)\Big]\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = k) \\
&- \; \beta^2 \sum_{x_{i,t+1}} \sum_{x_{i,t+2}} \Big[u_{t+2}(x_{i,t+2}, d_{i,t+2} = g)\Big]\widehat{q}(x_{i,t+2}|x_{i,t+1}, d_{i,t+1} = h)\widehat{q}(x_{i,t+1}|x_{i,t}, d_{i,t} = c)
\end{aligned}
$$

While it is still quite complicated, note that this equation is linear in the structural parameters. These parameters can therefore be estimated using a straight-forward Logit procedure without recursion, after integrating out over the transition of the state.