The Dynamics of Internal Migration: A New Fact and its Implications*

Greg Howard[†]   Hansen Shao[‡]

September 15, 2023

Abstract

We show a new fact: the t-year interstate migration rate is proportional to the square root of t. This is a puzzle for the standard moving cost model which predicts an approximately linear relationship. We propose a model based on persistent and spatially-correlated preferences that matches the square root fact, as well as the frequency and gravity pattern of internal migration. Compared to the standard model, the new model is better at forecasting individuals' locations, has different implications for welfare effects of migration policies, and makes different predictions for long-run population changes in response to shocks.

Keywords: regional evolution, misallocation, gravity equation, labor mobility, moving costs

JEL Codes: R23, R13, J61

*This paper was previously circulated as "Internal Migration and the Microfoundations of Gravity." We would like to thank Javier Quintana, Kyle Mooney, Yao Wang, Vivek Bhattacharya, Gregor Schubert, Andrii Parkhomenko, Eduardo Morales, Marieke Kleemans, and participants at the Illinois Macro Lunch, Illinois Young Applied Faculty Lunch, the European UEA meetings, Midwest Macro, the AREUEA National Meetings, the Stanford Institute for Theoretical Economics Housing & Urban Economics Conference, the Online Spatial and Urban Seminar, and the North American UEA meetings for constructive feedback. We also would like to thank Jialan Wang, Julia Fonseca, and Peter Han for creating the Gies Consumer and Small Business Credit Panel and the Gies College of Business for supporting this dataset.

[†]University of Illinois Urbana-Champaign, glhoward@illinois.edu.

[‡]University of Illinois Urbana-Champaign.

There are two prominent facts about internal migration: first, a large majority of people in most years do not move; second, if they do move, they are more likely to move to nearby places with large populations. This second fact is often referred to as the "gravity" relationship.[1] Economic modelers typically understand these facts as the consequence of moving costs that are large and increasing in distance.

In this paper, we ask what if these facts are instead due to people's persistent and spatially correlated preferences over where to live? Does the way we model internal migration have consequences for how we answer important spatial economic questions? We are motivated to study these questions based on a new third fact about internal migration: the fraction of people who live in a different state than they did t years ago is proportional to the square root of t. We show that this new fact is a puzzle for the standard moving cost model that is widely used in the literature. In that model, people's location choices are Markov, which when combined with the fact that migration is rare, implies a linear relationship between the t-year migration rate and t.

The main contribution of our paper is a simple but novel model that can match all three facts by assuming individuals' preferences are correlated across time and nearby locations. The model we propose is a dynamic version of a multinomial probit model with preferences that are persistent and spatially-correlated. Because the model features Spatially and Persistently Auto-Correlated Epsilons, we call it the "SPACE" model.[2]

The SPACE model is able to match the three main facts about migration, including the new square root fact. We also show that the SPACE model, when calibrated to match the first two facts of internal migration, can also match a variety of additional facts about the dynamic moments of internal migration, including the conditional probabilities of return migration, and the distribution of the number of moves over longer time periods. In contrast, the moving cost model is unable to match these additional facts.

[1] There are more facts that one could list about internal migration, many of which are covered in Molloy, Smith and Wozniak (2011). These two facts are important to this paper because they motivate moving cost models.

Throughout the paper, we will define the "gravity relationship" to be the empirical correlation of migration between two states with the population of each state (positive) and the distance between the two states (negative). We are not focused on whether this relationship is structural or on suggesting that it follows any exact functional form.

[2] ϵ is the common notation for the random component in a random utility model.

Next, we compare the implications of the SPACE model to those of the moving cost model. For many questions, we show that choosing how to model migration is not an innocuous choice, but leads to important differences in how economists answer central questions about location choice and migration.

First, we show that the SPACE model is better at predicting future locations of individuals. We compare the forecasting performance of each model using simulated log likelihood, and demonstrate that the SPACE model does better at predicting out-of-sample locations, especially at longer-horizons. This is consistent with the idea that the SPACE model is able to match realistic dynamics of location choice.

Second, the SPACE model and the moving cost model have very different perspectives on why people do not move. Moving cost models estimate large moving costs (e.g. Kennan and Walker, 2011; Giannone, Li, Paixao and Pang, 2020; Zerecero, 2021). In contrast, the SPACE model does not need moving costs at all to rationalize observed levels of migration in the data. To the extent that one thinks of moving costs as a friction that causes misallocation and which can be overcome by policy, those potential welfare gains are not present in the SPACE model.[3]

Third, we turn to macroeconomic questions, which typically depend on the elasticity of local populations with respect to local changes in utility. We show that both models feature similar short-run population cross-elasticities, in that the elasticity of population in state i with respect to utility in state j is approximately proportional to the gross migration rate between the two states.[4] In other words, if the purpose of a model with migration is to predict short-run effects on populations, both models deliver similar results.

However, in the long run, population cross-elasticities are quite different across the two models. In the SPACE model, population elasticities are the same in the

[3] In the SPACE model, preferences represent anything that determines individuals' location choice. While we do not model exactly what these are, we are open to the possibility that this may include information frictions or other frictions that lead to misallocation or which are policy-relevant. We are simply making the point that if the planner could magically relocate a person from one place to another without paying moving costs, that would not be welfare-enhancing in the SPACE model as it would be in the moving cost model.

[4] Even though the models have similar elasticities, the rationale for why the elasticity is related to migration is a bit different. In the SPACE model, the rationale is that migrants are close to indifferent between living in each state, so the mass of people that will move in response to a small shock is close to proportional. In the moving cost model, the rationale is that the extreme value function has a functional form such that the number of people on the margin is proportional to the number of people who make that choice.

short-run and the long-run, meaning that long-run elasticities are still proportional to the migration between the two places. But in the moving cost model, the long-run population elasticities are approximately the same as a static discrete choice logit model, i.e. the elasticity is proportional to the population share of the shocked region.[5] So the SPACE and moving cost models have completely different long-run population elasticities (in the data, population shares and gross migration rates have very low correlation). This is a problem if we wish to make predictions about long-run populations and we use the wrong model.

A fourth difference follows naturally from the previous one: the dynamics of regions' population changes are quite different in the two models. In the SPACE model, the dynamics are simple. In response to a permanent utility change, the population adjusts fully, contemporaneous to the utility shock. But in the moving cost model, the dynamics are relatively slow and can be unintuitive. In that model, a new set of people get a sufficiently large enough shock to move each period, and so a permanent utility shock raises the migration rate, and the population adjusts slowly. Furthermore, because the long-run elasticities are related to population shares, not migration shares, faraway states from the shock adjust particularly slowly, while nearby states are going to adjust quickly and may overshoot.

Finally, the SPACE model and the moving cost model interpret the data very differently in terms of which states have become higher-utility over time. In the SPACE model, one can invert the population elasticities matrix to calculate the implied utility changes from the population data. In the moving cost model, changes in utility are identified off of changes in the migration rates between states. When we use data on the recent trends in the U.S. to infer which states are gaining in terms of relative utility, we get very different answers depending on which model we use. This is critical if we want to estimate the role of policy or economic shocks on welfare.

We finish the paper by discussing the importance of the differences between the SPACE and moving cost model in the context of the literature. We show how the differences we highlight are central to some of the questions that are asked in the dynamic spatial literature. We discuss whether various approaches to enrich the moving cost model would deliver similar results to the SPACE model. And finally,

[5] So in the moving cost model, the long-run population elasticity of state i and state j to a shock in state k are the same, no matter how different the migration rates between i and k versus j and k are.

we offer a trick to use the tractability of moving cost models to still approximately match the medium- and long-term elasticities of the SPACE model, which we think will be helpful in quantitative applications.

In order to clarify the contributions of the model, we wish to be specific regarding the difference between moving costs and persistent preferences. While mathematically, it is straightforward to specify them in a model (as we will do here), it is important to understand what each term means when we try to map it onto the real world. A typical moving cost model has a one-time irreversible cost borne by people who leave one area for another. In contrast, persistent preferences across space mean that the change in utility when a person moves from one location to the other is both persistent over time and partially reversible should the person move back to the original location. A moving truck and the psychological cost of throwing a goodbye party clearly are moving costs. But many things described as "costs" in the literature are easily reversible—although it may decay with time—and not one-time. Having to live a long way from your friends or favorite amenity is easily reversible and is borne continuously. So even though those are often called "moving costs" in the literature, we think that that terminology is used because existing models have not been able to distinguish persistent preferences from moving costs. The rest of this paper will give many reasons why this distinction is important.

Literature

How people choose where to live is a classic question in the urban economics literature. Many urban models assume utility is equalized across space in the tradition of Rosen (1979) and Roback (1982). Other more quantitative models assume a discrete choice framework for locations to answer a variety of questions, such as the role of endogenous amenities on location choice (Diamond, 2016) or spatial misallocation on aggregate output (Hsieh and Moretti, 2019).

A part and increasingly large part of this literature has explicitly looked at the dynamics of location choice, i.e. migration. Since at least Blanchard and Katz (1992), migration has been recognized as a key feature to how regions adjust to economic shocks. In this vein, papers studying the rise or decline of regional economies put a significant emphasis on migration (Caliendo, Dvorkin and Parro, 2019; Allen and Donaldson, 2020; Morris-Levenson and Prato, 2022), and especially the speed at which migration happens (Glaeser and Gyourko, 2005; Kleinman, Liu and Redding, 2023; Amior and Manning, 2018; Davis, Fisher and Veracierto, 2021). Similarly, when aggregating up to the macroeconomy, migration plays a critical role in how quickly countries adapt to changing technologies or external shocks (Tombe and Zhu, 2019; Hao, Sun, Tombe and Zhu, 2020; Eckert and Peters, 2018; Giannone, 2017; Heise and Porzio, 2021; Bryan and Morten, 2019). A growing literature has emphasized how migration, including internal migration, plays an important role in adapting to global warming (Rudik, Lyn, Tan and Ortiz-Bobea, 2021; Cruz and Rossi-Hansberg, 2021; Oliveira and Pereda, 2020). Migration is also known to be an important margin when analyzing housing markets in particular (Schubert, 2021).[6] Central to many of these questions is how elastic is the population of a region to various shocks, over various time horizons. One of the contributions of this paper is to examine how robust those conclusions are to alternative ways of modeling migration.

Corresponding to the growth of interesting questions related to migration, there has also been advances in the ways to model migration. Kennan and Walker (2011) wrote down the canonical model of migration using the dynamic logit formulation. Kaplan and Schulhofer-Wohl (2017), Giannone et al. (2020), Porcher (2020), Mangum and Coate (2019), Zerecero (2021), and Monras (2018) have built on this formulation to include additional realistic features of moving, such as richer information frictions, wealth of migrants, home bias, and nested decision making.[7] Other approaches, such as Coen-Pirani (2010) and Davis et al. (2021) do not use the dynamic logit, but have similar discrete choice models that improve the tractability in a way specific to their goals. All of these models use moving costs to explain the low rates of migration, and

[6]Howard and Liebersohn (2021) and Howard, Liebersohn and Ozimek (2023) also study the effects of changing location choice on housing markets, but model location choice in a static discrete choice framework rather than explicitly having a notion of migration.

[7]In particular, Kaplan and Schulhofer-Wohl (2017) argues that changes in information frictions can help explain the decline in interstate migration, along with decreases in the different returns to various occupations across space. Giannone et al. (2020) builds a rich model of migration that incorporates wealth and borrowing, to analyze how credit and savings can affect if and where people choose to move. Porcher (2020) builds a tractable model of rational inattention in the dynamic migration context to argue that information frictions are one of the main reasons people do not move. Mangum and Coate (2019) includes both a bias for living near a birthplace, as well as attachment to a place that grows over time spent there, and uses that to argue that shift of the American population to the West and to the South is responsible for slowing labor mobility. Zerecero (2021) also examines a model that includes a preference for birthplace. Monras (2018) looks at the asymmetric response of inmigration and outmigration to local shocks, and builds a dynamic nested logit model to better understand the phenomenon.

potentially adjust those moving costs to explain the high rates of return migration. In contrast, only one paper to our knowledge uses persistence in preferences to explain low migration rates: Bayer and Juessen (2012). However, the model Bayer and Juessen (2012) is too complicated to extend beyond two regions, limiting its use in many empirical applications.

One type of persistent preferences has been introduced by Zabek (2020), Mangum and Coate (2019) and Zerecero (2021), by adding in a preference for living in one's birthplace.[8] These models have some similarity with the SPACE model, in that they also tend to feature smaller moving costs (Zerecero, 2021) and would intuitively feature more return migration than the standard moving cost model. However, adding birthplace preferences to a moving cost model does not generate the square root fact and does not change many of the implications we highlight as being distinct between the two models.

At the same time that models of internal migration have become more popular, empirical evidence focused on the causes and barriers to migration has also grown. For example, Saks and Wozniak (2011) shows migration is cyclical; Kleemans (2015) studies the income shocks that cause migration; Farrokhi and Jinkins (2021) examines the attachment hypothesis using a policy change amongst Danish refugees; Koşar, Ransom and Van der Klaauw (2021) uses a survey experiment to better understand how people make location choice decisions; and Fujiwara, Morales and Porcher (2022) proposes a methodology for uncovering information frictions in location choice. Our paper also contributes to this literature by establishing the additional stylized fact that the t-year migration rate is proportional to $\sqrt{t}$.

Finally, because we wish to model persistent preferences, we use a discrete choice model with multinomial probit.[9] The reason for this is that the extreme value distributions necessary for specifying a dynamic logit do not easily include persistence. Multinomial probit is known for being flexible in terms of its ability to accommodate different cross-elasticities, but is computationally intensive when the number of choices is large (Butler and Moffitt, 1982; Keane, 1992; Geweke, Keane and Runkle, 1994). We make progress on this latter point by showing how observed migration is an important statistic that, along with the parameters, is sufficient for calculating

[8]The canonical model in Kennan and Walker (2011) also includes a premium for birthplace.

[9]Other papers have tried to model spatially-correlated preferences with logits, i.e. Bhat and Guo (2004), but require either perfect persistence or non-persistence in the personal utility terms, which leads to either no gross migration, or migration rates near 100 percent.

population cross-elasticities.

1 Data

Throughout the paper, we primarily use credit data from one of the leading credit report providers to measure migration (Gies Consumer and Small Business Credit Panel, 2004-2018), although when we can, we verify the data using the IRS migration data, the American Community Survey, or the Panel Survey of Income Dynamics (IRS Migration Data, 2004-2018; Ruggles, Genadek, Goeken, Grover and Sobek, 2015; Panel Survey of Income Dynamics, 1969-1997).[10] The credit data is a 15-year panel of individuals making up a 1 percent sample of the United States. It records the state of residence in each year, allowing us to calculate migration rates at longer horizons. The GCCP has a similar one-year migration rate to the IRS data, which can be seen in Figure 1. One reason the credit data may have a higher migration rate is that not everyone has a credit report. In particular, lower income people tend to not have credit reports, and are also less likely to move.[11]

To calculate distances between states, we use the geographic center of each state from Rogerson (2015) and calculate distances using the formula from Vincenty (1975).

The data is consistent with the two well-known facts we stated at the beginning of the paper.

The first fact is that interstate migration is rare. In Figure 1, we show that the

[10]For other papers using the Gies Consumer and Small Business Credit Panel, see Fonseca (2022), Fonseca and Wang (2022), and Han (2022). DeWaard, Johnson and Whitaker (2019) analyzes a similar credit dataset (the Federal Reserve Bank of New York/Equifax Consumer Credit Panel) on how it can be used to study migration.

[11]While there are some well-known drawbacks to the IRS data, e.g. it is based only on tax filers, it is one of the most comprehensive administrative datasets keeping track of migration. It is not well understood why the migration rate is so low in 2014 or so high in 2016, as these anomalous values did not show up in other datasets measuring migration. See DeWaard, Hauer, Fussell, Curtis, Whitaker, McConnell, Price and Egan-Robertson (2020). Similarly, while credit data are not designed as a dataset to study migration, it does have location information, and the bureau gets the addresses from a person's financial accounts. The biggest concern with credit data is that moves may show up with a lag, as people do not always immediately change their addresses with their financial institutions. For our square root fact, we check the robustness to using the Panel Survey of Income Dynamics (1969-1997). The utility of using credit data to discuss internal migration is discussed in depth in DeWaard et al. (2019). The Gies credit data is an unbalanced panel, with yearly observations occurring in May. For matching migration patterns and rates, we focus on the 2004-2005 period, so we only observe data if they had a credit report in both of those years. For some of the dynamics, we address the unbalanced nature of the panel depending on the moments of the data that we are interested in.

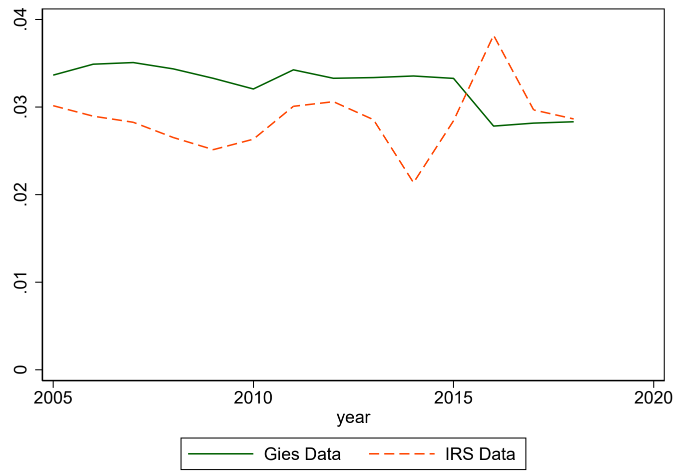


Figure 1: Comparison of 1-year interstate migration rates in IRS and GCCP data.

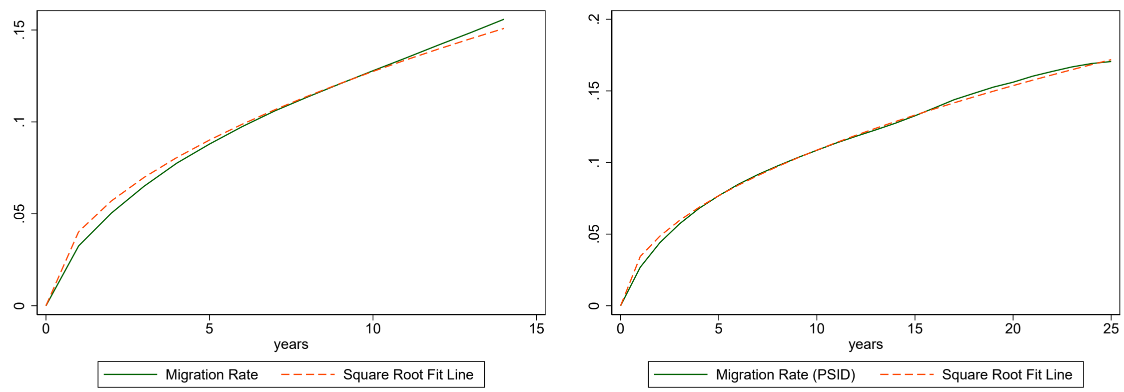
GCCP records slightly more than 3 percent of Americans moving between states in any given year. This is slightly higher than the IRS data, which is also shown in the figure. The vast majority of Americans do not move between states in any given year.

The second fact is that migration follows a gravity pattern, meaning that the amount of migration between two states is increasing in each state's population, and decreasing in the distance between them. Later, in Table 2, we will show the results from a poisson regression in which we regress migration on log distance and the log populations of the origin and destination states (Silva and Tenreyro, 2006; Correia, Guimaraes and Zylkin, 2019):

$$\log m_{i \to j} = \alpha \log p_i + \gamma \log p_j + \beta \log \text{distance}_{ij} + \epsilon_{ij} \tag{1}$$

where p_i is the population of i and $m_{i \to j}$ is the 1-year migration from i to j. We find that both α and γ are positive and β is negative. The coefficients for the IRS data and the GCCP data are similar.[12]

[12] In Table 2, we also show the regressions with origin- and destination- fixed effects. The coefficient on distance is still negative.



(a) Credit Data (GCCP)

(b) PSID

Figure 2: Migration Rates at Different Horizons. Migration rate at year t is calculated as the percentage of people living in a different state than they did t years ago. Both datasets are unbalanced panels and use any observations in which the state of residence is recorded t years apart. Source: GCCP and PSID.

2 New Fact

Define the t-year migration rate to be the number of people living in a different state than they did t years ago. The new fact is that the t-year migration rate is approximately proportional to $\sqrt{t}$. In Figure 2(a), The solid line is the t year migration rate in the GCCP. The dashed line is a constant times the square root of t, with the constant chosen to match the level of migration. As is apparent from the figure, the shape of the migration rate is very similar to the square root line.[13]

Of course, since we cannot measure dynamic migration moments in the IRS data, one might wonder if the square root fact is driven by some sort of mismeasurement in the GCCP. In Figure 2(b), we show that the square root fact is also present in data from the Panel Survey of Income Dynamics (1969-1997). In fact, we extend the horizon to 25 years and show that it holds through that longer time period as well.[14]

This new fact relates to the more-well-known fact that return migration is common. Many papers in the literature show a significant fraction of workers return to their previous location (e.g. Kennan and Walker, 2011; Kaplan and Schulhofer-Wohl, 2017). One consequence of this fact is that the two-year migration rate is significantly

[13]Each point is the mean of a binary variable with millions of observations, so if we tried to put standard errors on the graph, they would not be visible.

[14]Since mismeasurement in the GCCP may be a particular concern for young people, we show in Appendix E.1 that it holds for people over 45, where age is estimated by the credit bureau.

less than twice the one-year migration rate. However, we believe we are the first to document this specific relationship.

2.1 The New Fact is a Puzzle

The square root fact is interesting not only because it is an empirical regularity in need of an explanation, but also because existing models do not provide a satisfactory explanation. This section shows that the most common model of internal migration in fact leads to a linear relationship between the t-year moving rate and t.

First we outline a standard moving cost migration model. There are a continuum of individuals of mass 1, denoted by n, who can choose to live in locations denoted by i, j, or k. An agent that lived in j at $t-1$ has utility:

$$V_{nt}(j) = \max_i \{u_{it} - \delta_{ij} + \epsilon_{int} + \beta \mathbb{E} V_{n,t+1}(i)\} \tag{2}$$

where u_{it} is the common utility for everyone living in i, δ_{ij} is the bilateral moving cost between i and j, and ϵ_{int} is an independent and identically distributed random variable with an extreme-value distribution. We assume ϵ_{int} has a Gumbel distribution with scale parameter 1. If we define $v_{it} \equiv u_{it} + \beta \mathbb{E} V_{n,t+1}(i)$, then the migration probability is given by:

$$\frac{m_{i \to j,t}}{p_{it}} = \frac{e^{v_{jt} - \delta ij}}{\sum_k e^{v_{kt} - \delta_{ik}}} \tag{3}$$

What does this model predict for the dynamics of migration, especially for the shape of the t-year migration rate? In general, not much. However, when moving costs are high—which is required to match the low amounts of interstate migration in the data—then the following proposition shows that the t-year migration rate is approximately linear in t.

To set up the proposition, it is helpful to suppose that moving costs are given by $\delta_{ij} = \delta'_{ij} + \Delta$ when $i \neq j$, and $\delta_{ii} = 0$. For $i \neq j$, migration costs consist of a pair-specific component that governs the relative amount of migration to j, and Δ, a common component which governs the overall amount of migration in the economy. This way, when we change Δ, we are not changing the relative amount of migration from i to j versus i to k.

Proposition 1. *In the steady-state of a moving cost model, as the common component of moving costs go to infinity, the t-year migration rate is proportional to t.*

$$\lim_{\Delta \to \infty} \frac{m^t_{i \to j}}{m^1_{i \to j}} = t$$

where $m^t_{i \to j}$ *is the* t*-year migration rate from* i *to* j*.*

Proofs are collected in Appendix A.

This proposition establishes that the square root fact is not a natural consequence of our standard models. In the standard model, we infer high moving costs based on the fact that migration is low, and this proposition establishes that high moving costs imply a linear relationship between the t-year migration rate and t.

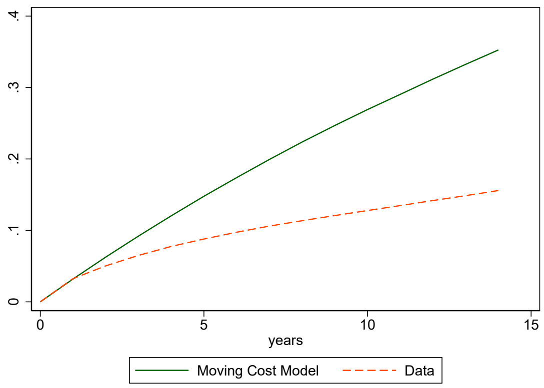
Nonetheless, moving costs are not literally infinity, so we also check that the model estimated to the data would generate a relationship that looks linear.[15] First, using the standard moving cost model, we estimate the moving costs that would correspond to the GCCP migration rates in 2004-2005. Then we calculate the within-model t-year migration rate, by simulating the model for millions of observations over 15 years. Figure 3a shows the results of this simulation. The relationship between the t-year migration rate and t looks very linear, and does not do a good job of approximating the data.

In the simulations, it is also easy to make the migration costs depend on variables that the reader might think could break the linear relationship. For example, many models include as a state variable whether the person moved in the previous year and if so, from where (e.g. Kennan and Walker, 2011). Here, we use the GCCP data to calibrate a model in which the moving costs depend on the state of residence, but also the state of residence from the year before. We show the results in Figure 3b. While this additional state variable adds a kink at $t = 1$, the relationship becomes fairly linear again for larger t's.[16]

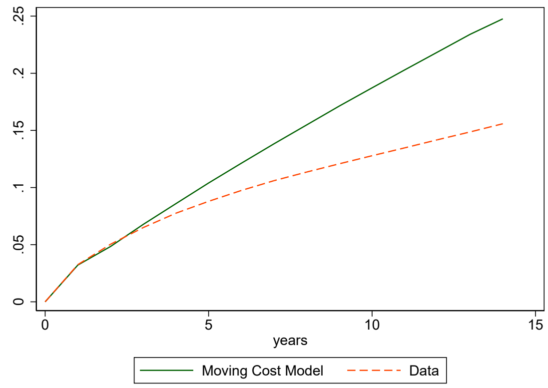
Similarly, one might think that adding age as a state variable could help match the

[15] Alternatively, a reader might wonder if the fact that the world is not in a steady-state could generate a square root pattern. However, as documented in Jia, Molloy, Smith and Wozniak (2023) and other papers, gross migration flows are much larger than net migration flows, often by an order of magnitude, so that cannot explain the difference between moving cost models and the square root fact. In unshown simulations, we estimate the a moving cost model where the moving costs vary year by year to match the migration rates in each year, and they are not appreciably different.

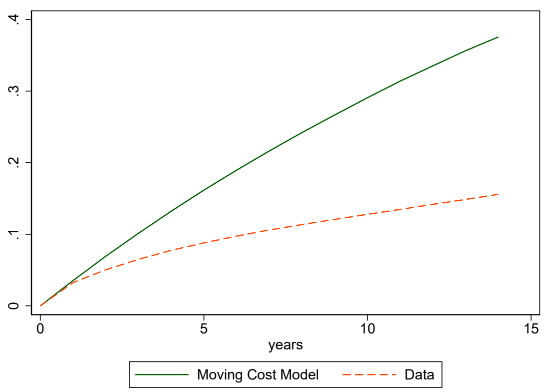
[16] Building on this exercise, one could of course match the square root fact up to 14 years by conditioning the moving costs on the person's location for the last 14 years. Of course, if you only care about using the model for predictions and counterfactuals that occur within 14 years, this would work great. But since the model is likely to start making counterfactual predictions about the dynamics after 14 years, one would still need to proceed with caution beyond that horizon.



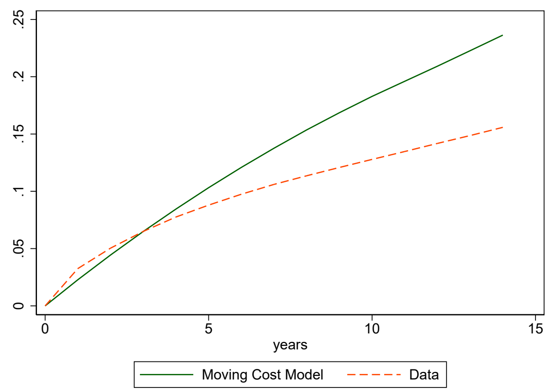
(a) Baseline Moving Cost Model



(b) Previous Year's Location State Variable



(c) Age State Variable



(d) Birthplace State Variable

Figure 3: Do existing moving cost models match the square root fact? Each panel compares the t-year migration rate in simulations of the moving cost model described in Section 2.1 to the data. The model is calibrated by picking migration costs to exactly match interstate migration flows. In Panel (a), migration costs vary by origin-destination pair. In Panel (b), migration costs are allowed to vary by the interaction of location in the year prior, origin, and destination (i.e. someone that had lived in state X then moved to state Y will have different moving costs than someone that lived in state Y for two years). In Panel (c), migration costs vary by the interaction of age, origin, and destination. And in Panel (d), migration costs vary by the interaction of birthplace, origin, and destination. The data is from the GCCP, and for panels (a)-(c), the model is fit on the GCCP data. Panel (d) is fit using ACS data, which has a lower 1-year migration rate, and so the data and model do not match even at $t = 1$.

square root fact since migration rates decline over the life-cycle (through the lens of the standard model, moving costs increase in age). Calibrating the model to depend on the GCCP's age variable still leaves a very linear relationship, as can be seen in Figure 3c.[17]

Finally, we also check whether adding the state of birth is helpful to match the curvature of the t-year migration rate, since moving back home may be a high-enough probability event to make Proposition 1 a bad approximation.[18] For this calculation, we use the American Community Survey (ACS) data, which records the state of birth, as well as interestate migration, to calibrate the t-year migration rate. The migration rate in the ACS is modestly lower, so the migration costs will be calibrated to be higher. However, what we are primarily interested in is the linear relationship, which does not change with this additional state variable.

3 New model that can hit the new fact

This section introduces a new model of internal migration which can resolve the square root puzzle from the previous section. Rather than depend on moving costs, it assumes that the match-specific utility (the ϵ's) are spatially-correlated and persistent. In fact, because the model features Spatially and Persistently Auto-Correlated Epsilons, we call it the SPACE model.

As in the moving cost model, there is a continuum of individuals with mass 1, a finite number of discrete locations, and discrete time. We keep the same notation where n denotes the individual, i the location, and t the year. And as in the standard model, individuals pick their location to maximize utility:

$$V_{nt}(\vec{\epsilon}_{nt}) = \max_i\{u_{it} + \epsilon_{nit}\} + \beta V_{n,t+1}(\vec{\epsilon}_{t+1}) \tag{4}$$

where u_{it} is a common utility for location i and ϵ_{nit} (the ith element of vector $\vec{\epsilon}_{nt}$)

[17]The GCCP imputes age, so that may introduce errors into the calibration. However, Figure A2 confirms that the GCCP does pick up the lifecycle decline in migration.

[18]The reader might wonder if other demographics would matter. For example, college educated workers are known to move more (Molloy et al., 2011). However, for immutable characteristics, the aggregate migration rate would just be the average of the different groups, so if the groups have linear t-year migration rates, then the aggregate would also be linear. So as long as migration costs are high for every group, then the t-year migration rate will still be linear.

is a person-location match specific utility.[19] Note that the choice of location i does not affect the continuation value because there are no moving costs, so the choice is made sequentially each period, and it has no effect on future choices.

We further assume that the ϵ vector is auto- and spatially-correlated:

$$\vec{\epsilon}_{nt} = \rho \vec{\epsilon}_{n,t-1} + \sqrt{1-\rho^2}\vec{\eta}_{nt} \qquad \text{and} \qquad \vec{\eta}_{nt} \sim N(0, \Sigma) \tag{5}$$

Note that $\vec{\epsilon}_{nt}$ also has variance Σ.

We assume that states that are closer to one another have a higher correlation of personal utility by using the following functional form for the covariance matrix Σ:

$$\Sigma_{ij} = \exp(-a \text{ distance}_{ij}) \tag{6}$$

where a is a parameter and distance_{ij} is the distance between i and j.[20]

A reader may wonder if the specific features of the SPACE model are "cooked up" to match facts but lack a basis in reality. Rather, we think that the SPACE model features adds two very realistic features of preferences. First is that the personal preferences for location are persistent over time. People mostly cite family and work in surveys about why they move (Jia et al., 2023). People's feelings about these things are surely correlated over time, and it is an empirical fact that each of these things

[19] We do not take a stand on where the u_i's originate, so the reader can think of the SPACE model as being the migration block of a spatial model, and that the u_i's would originate in the housing, production, and amenities blocks.

Note that because there are no moving costs, the continuation value $v_{it} = u_{it} + \beta V_{n,t+1}(\vec{\epsilon}_{t+1})$ differs from u_i by just a constant.

[20] A reader may wonder why we use this specific functional form.

Observation 1. Σ *is positive-definite.*

This observation establishes that creating a covariance matrix based on this function of distance is always going to create a positive definite matrix, so it can be applied to other countries or geographic divisions of the United States. The proof can be seen in Appendix A, but the key to the proof is to show that $\exp(-ax)$ is completely monotone and then to apply the Schoenberg Interpolation Theorem (Schoenberg, 1938). Complete monotonicity is a fairly restrictive requirement. However, there are other functions besides $\exp(-ax)$ that would work, such as $1/(1+ax)$. We went with $\exp(-ax)$ since it has a nice parallel to the covariance over time, which also decays exponentially in this model.

A caveat to this exercise is that the Schoenberg Interpolation Theorem applies to distances in $\mathbb{R}^n$ but not necessarily to distances on a sphere. However, the correlation between distances measured using Vincenty (1975)—which we use in this paper—and distances calculated using the Pythagorean Theorem based on longitude and latitude is 0.98. In practice, we do not encounter any non-invertible matrices regardless of how we parametrize a.

is persistent in terms of location. Second is that personal preferences for location are spatially-correlated. Again, when you think about people's preferences, the ability to live near family is highly-correlated across space. If state i is close to family, then states near i are also close to family. Industrial composition, i.e. the types of jobs people can get, also tend to be geographically concentrated. Natural amenities or regional cultures—other possible sources of personal preferences—are also spatially correlated. The functional forms of equations (5) and (6) are indeed convenient mathematically and are likely not precisely true, but it should be hard to argue that spatially and auto-correlated location preferences are somehow less realistic than the i.i.d. personal preferences of a moving cost model.

Migration occurs when the locations i that maximize $u_{it} + \epsilon_{nit}$ and $u_{i,t+1} + \epsilon_{ni,t+1}$ are different.

Without moving costs, the model relies on persistence and spatial correlation to generate migration that is rare and follows a gravity pattern. The intuition behind why the model generates rare migration is that if preferences are persistent, then the state that people like most today is still going to be liked a lot by them tomorrow, so they are unlikely to move. The intuition for the gravity pattern is that people are more likely to like states that are nearby. Since they like their own state more than every other state, they are much more likely to have high ϵ_{in} for nearby states, as well, making it more likely to get a shock that makes them choose one of those states than a shock that makes them choose one of the faraway ones.

Having setup the model, we now present the analogous proposition from the moving cost model, to see whether the SPACE model can match the square root fact. In this case, the limit we consider that makes it so that there is only a little migration is for the persistence parameter ρ to approach 1.

Proposition 2. *In steady-state of the SPACE model, as preferences become perfectly persistent, the t-year migration rate is proportional to* $\sqrt{t}$*, i.e.*

$$\lim_{\rho \to 1} \frac{m^t_{i \to j}}{m^1_{i \to j}} = \sqrt{t} \tag{7}$$

where $m^t_{i \to j}$ *is the t-year migration rate from* i *to* j*.*

Proposition 2 shows that the new model can match the new fact, resolving the puzzle.

Table 1: Parameterization

Parameter	(1) Value	Moment	(2) Data	(3) Model
a	0.000279	Distance coefficient	-0.7436	-0.7436
ρ	0.999621	Average migration	0.0336	0.0336
v_i		Population of each state		

One intuition for why the model follows the precise square root pattern is that the standard deviation of how much your relative preference for i versus j changes in t periods is proportional to the square root of t. If the area near the indifference cutoff is relatively uniform, then the number of people who cross the cutoff in t years will be proportional to the standard deviation.

3.1 Calibration

In this section, we calibrate the model to match the rarity of migration, and the gravity relationship of internal migration. We then use the calibrated model to assess the model's performance on dynamic moments related to migration.

In light of Proposition 2, this section serves two purposes: first, the proposition assumes $\rho \to 1$, so we can see if a calibration with $\rho < 1$ can also match the square root pattern; and second, simulating a calibrated model allows us to look at many other dynamic moments for which we do not have propositions. We generally find that the SPACE model is good at matching dynamic moments of migration, at least compared to the standard moving cost model.

We parametrize ρ, a, and u_i in order to match the probability of migration, the coefficient on distance in a gravity regression, and the population in each region. For the gravity moment, we run a Poisson regression of migration on log population of the origin state, log population of the destination state, and log distance. We match the coefficient on log distance. We simulate ten million workers for two periods.[21]

ρ, which represents how persistent people's taste shocks are, is helpful to match the average amount of migration. The best match involves a very persistent taste

[21] Even though there are ten million individuals in the simulation, the outcomes are discrete, and so the typical techniques based on differentiation are not helpful to do the simulated method of moments. See Appendix B for details on the calibration.

shock of more that 0.999.[22] a, which is about how correlated tastes are across space, is used to match the gravity coefficient. Here, distance is measured in kilometers, so a is not particularly interpretable, but the value is .000279 meaning that correlation of shocks for two states 1000 kilometers apart—i.e. approximately the distance from North Carolina to Massachusetts—is about 0.76. The u_i's are picked to match the population of each state.

Table 2 shows the results of the Poisson gravity regression using IRS data (Column 1), the GCCP (Columns 2 and 4), and simulated data from the calibrated model (Column 3 and 5). The first three columns are the estimated coefficients from the typical Poisson regression, specification (1). Since we target the coefficient on distance in a gravity equation, the estimated coefficients are similar by construction. The coefficient from simulated data is not significantly different from the results from IRS data. Also, the coefficients on origin population and destination population are very similar across the two datasets and the simulation, even though we are not targeting these coefficients.[23] The last two columns in Table 2 use a different specification. We replace the population terms with state-level fixed effects. Again, the estimated coefficients are comparable.[24]

Our calibration targeted the migration rate and the gravity relationship, so we test the model by seeing how it does at matching a variety of untargeted dynamic moments. We start by looking at the moving probability over different horizons, to see if the calibration can capture the square root fact.

In Figure 4a, we show the t-year migration rate does follow a square root pattern in both the data and in a simulation of the SPACE model. For the data, we include any observations for which we have credit reports t years apart, so it should be noted

[22]This is helpful since many of our theorems will only be applicable as $\rho \to 1$.

[23]A related fact is that places with high inmigration also have high outmigration (Coen-Pirani, 2010). For example, D.C. has much higher gross migration than average and California and Texas have much lower gross migration than average. In this model, there is a statistically-significant correlation between population and gross migration rates, as there is in the data.

[24]We also compare the relationship of distance and migration non-linearly. Appendix Figure A3 shows a binscatter plot of migration adjusted for populations and distance. First, we divide the number of migrants by the product of origin and destination population, for both the IRS data and the simulation. Since the units are not particularly interpretable, we normalize this measure to have mean 1. Then we separate the state-pairs into bins based on the distance between states and plot the mean within each bin. Even though our parametrization did not target this pattern, the simulated data and the IRS data look similar. In the bin with the closest states, population-adjusted migration is about five times higher than average for both the data and the simulation. For distance bins over 2000 kilometers, migration is about half of the average.

Table 2: Gravity Equations

	(1) Migration (IRS)	(2) Migration (Credit)	(3) Sim. Migration	(4) Migration (Credit)	(5) Sim. Migration
Log Distance	-0.736***	-0.744***	-0.744***	-1.063***	-0.978***
	(0.0572)	(0.0515)	(0.0396)	(0.0672)	(0.0552)
Log Origin Population	0.900***	0.923***	0.892***		
	(0.0832)	(0.0797)	(0.0486)		
Log Destination Population	0.822***	0.893***	0.889***		
	(0.0976)	(0.0799)	(0.0501)		
Observations	2550	2550	2550	2550	2550
Origin and Destination FEs				Yes	Yes

Standard Errors are two-way clustered by origin and destination states
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

that the sample changes slightly depending on t.[25] The model does not match the data perfectly, with the lines diverging over time. In part, this is because the model is calibrated to match the one-year migration rate, and the fourteen-year migration rate, which in the simulation is about $\sqrt{14}$ times the one-year migration rate, is going to be sensitive to that choice. There would be much less divergence between the model and the data if we had tried to match the five-year migration rate instead. The figure also presents the same exercise for a simulation of the moving cost model, which as expected is much more linear and diverges much more from the data.

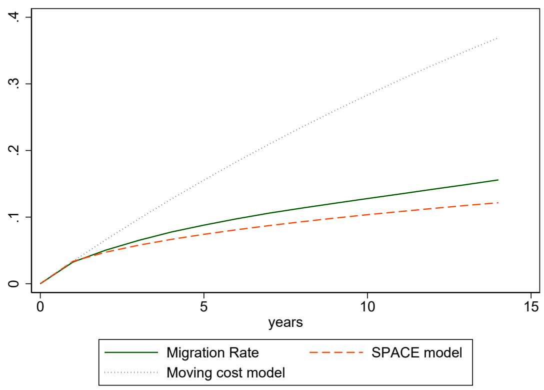
Of course, the t-year migration rate is not the typical way the dynamic moments of migration are presented in the data, so it is interesting whether the model is able to capture the more-commonly-examined moments as well. A natural moment is the conditional probability of moving given a previous move. This is sometimes split up between onward and return migration. Return migrants are those who move back to their original state after an interstate move. Onward migrants move to a new state after an interstate move, and the new state is not the state where they came from.

Figure 4b shows the probability of return migration and onward migration at different time horizons after an interstate move.[26] Since we do not target these statistics in the calibration, the simulated statistics do not match the data perfectly. The re-

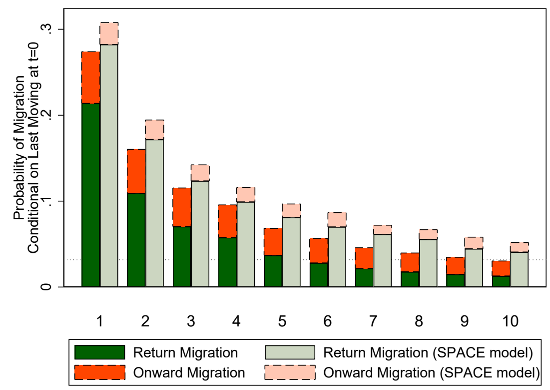
[25] Focusing only on a balanced panel of individuals gives an indistinguishable pattern, but raises concerns about excluding younger people who are most likely to move. Each of the dynamic moments that we look at in Figure 4 selects a slightly different sample of people, so the fact that the model is still a good match across the different moments shows the sample does not matter that much.

[26] To be included in this analysis, a person must show up for the number of years that would be necessary to calculate the statistic, but we do not use a balanced panel.

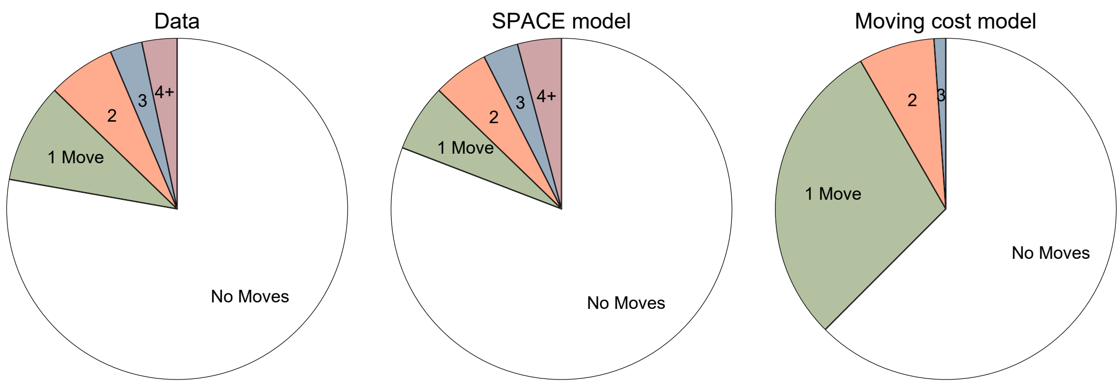
For this figure, we do not include bars for the moving cost model. However, since migration is Markov in that model, the probability of migration is the same regardless of when the person last moved. Hence, the bars would all be about 0.03, and mostly orange.



(a) Migration rate in the data and models



(b) Return and onward migration



(c) Number of moves in 14 years

Figure 4: Dynamic moments. In panel (a), the t-year migration rate is calculated as the percent of people living in a different state than they were t years ago. Data is from an unbalanced panel, and included any observations from 2004-2018 for which the state of residence is observed t years apart. In panel (b), the conditional probability of migration is plotted. For the value at x years, the probability of migration is conditional on the person having migrated x years previously and remained in the same state ever since. It is broken up into return migration, which is when the person moves back to the original state, and onward migration if they move to a third state. In panel (c), the number of moves in 14 years is calculated for people whose state is observed in every year from 2004-2018.

turn migration rate is generally higher in the simulated data, and the onward rate is lower. But the general pattern is similar, especially its decay as the person has lived in the state for longer.

The intuition for the return migration decreasing over time is simple. Conditional on having moved recently, the agents are likely relatively indifferent between the two regions, and are likely to move back. The longer they have stayed in one region, the more likely that their accumulated preference shocks have drawn them further away from being indifferent, so the probability of return migration decreases over time. The literature has typically focused on the concept of "attachment" to explain this phenomenon (Mangum and Coate, 2019; Farrokhi and Jinkins, 2021). In the SPACE model, people who have lived in a location for longer are more attached, but it is because their repeated decision not to move has revealed that they like the location, not due to an economic force that increases their utility by staying there longer.[27]

Another easy-to-measure moment is the distribution of the number of interstate moves over time. Figure 4c looks at how many moves are made over a 14 year period. In the data, a large majority of people make zero moves, but some people make many moves. Here, we include in this chart only people for whom we have data in all 15 years (for up to 14 possible moves). The model is able to capture the large fraction of people that never move, as well as come close to the data on the number of people that move once or twice. Importantly, it captures the fact that a few percent of people move four or more times over the fourteen years. The figure also includes similar statistics for a moving cost model, which does a much worse job.

3.2 Extensions

We wrote down the SPACE model to be able to capture the rich dynamics of internal migration. A reader may wonder whether the model is flexible to incorporate additional features of migration. After all, one of the advantages of the moving cost model is its flexibility in terms of being able to add realistic features tractably, so we ought to ask the same of the SPACE model.

In this section, we consider three extensions: other geographies, demographic differences, and a factor that might affect migration between two places other than

[27] Adding attachment to the SPACE model might be able to improve the match of return migration in the model and the data, but would come at the cost of the model's tractability.

distance. These are not meant to be an exhaustive list, but we hope it gives a flavor of the flexibility of the model.

Parameters are not geography-specific

In Appendix D.1, we see how the model performs when using a smaller geographic unit: commuting zones. We keep the same parameterization of the persistence, ρ, and spatial correlation, a, to see how the model does at matching cross-commuting-zone migration rates and gravity patterns. This is not a trivial challenge for the model. In the data, the migration rate is more than a percentage point higher and the gravity relationship is a bit steeper when measured with commuting zones. Nonetheless, the model is able to match both these features, suggesting that the parameters of the model are not specific to the geography of the model.

Importantly, this implies that we can be more comfortable using the model to estimate population elasticities—as we will do in the next section—in settings where the geography is not at the state-level, as well.

Demographics

One aspect of migration that we have abstracted from is that migration rates are highly heterogenous along demographic lines. For example, young people and college-educated people move at higher rates than other people (Molloy et al., 2011). Having a demographic-dependent ρ_d would be able to match different migration rates without changing the rest of the economics. In fact, because migration is almost directly proportional to $\sqrt{1-\rho^2}$ when ρ is close to 1, it is fairly straightforward to calculate the appropriate ρ_d's for different demographics, and the model would still approximately aggregate to $\sqrt{1-\rho_{\text{agg}}^2} = \sum_d \frac{L_d}{L}\sqrt{1-\rho_d^2}$, where L_d/L is the fraction of the population of demographic d.[28]

Generalizing "Distance"

Another way to extend the model is to consider other similarities besides distance that could drive migration. For example, there is more migration between states with

[28] A model with this type of heterogeneity would have slightly different implications for some of the dynamic moments of migration. For example, it might be able to better predict the increased likelihood of onward migration, conditional on past moves. However, the square root fact which motivated the model would still be the same.

similar levels of education. While the proof of Proposition 1 does impose requirements on the functional form, including that distance_{ij} be a norm in $\mathbb{R}^n$, we can include other measures in our definition of distance that allow us to match other deteminants of migration. For example, we could imagine a distance in $\mathbb{R}^3$, where the three dimensions are latitude, longitude, and education levels.

We carry out this exercise in Appendix D.2. It leads to small improvements in matching migration patterns. In theory, the covariance matrix Σ has many degrees of freedom with which to match bilateral migration, so one could extend the model in many dimensions.[29] However, this may come at a large computational cost, as there is not a closed-form relationship between Σ and bilateral migration between regions.

4 Does the new model matter?

The previous section introduced a new model and showed that it did a better job at matching dynamic moments in the micro data. In this section, we explore the implications of that model, especially by comparing it to the workhorse model. For some questions, we find that the differences are minimal, while for others we think there are substantial differences.

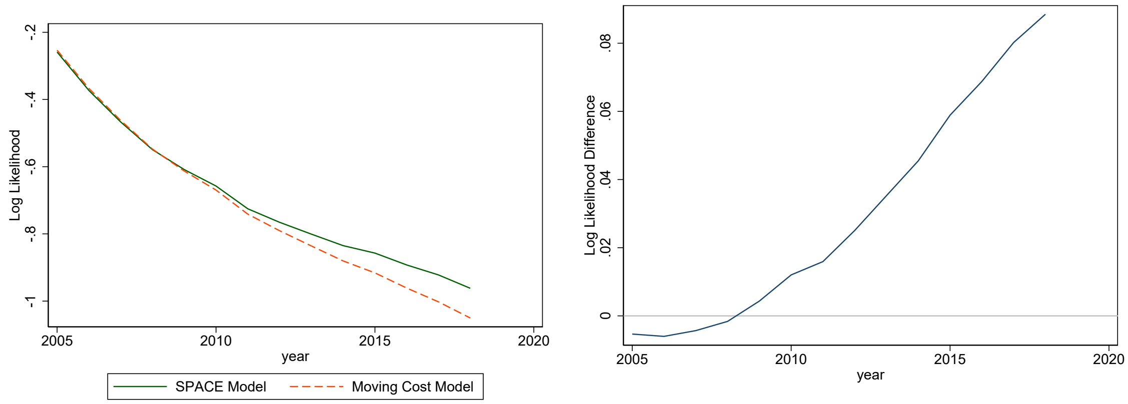
4.1 Micro Forecasting

The first reason we might care about the difference between the two models is for purposes of forecasting the location of an individual agent. Suppose we observe the agent's location in 2004, and wish to forecast where they will live in every subsequent year until 2018. We use the calibrated versions of each model to do the forecasting. We judge the performance of the models using the mean log likelihood: we simulate each model for millions of people, and then for each initial state, we calculate the simulated probability that a person who was in state i in 2004 ends up in state j in year y. Then using people's true locations in the data, we calculate the mean log likelihood,

$$L_y = \frac{1}{N} \sum_n \log \mathbb{P}(\text{lives in } j \text{ in } y | \text{lived in } i \text{ in } 2004)$$

for each year. We plot L_y in Figure 5.

[29] For example, including richer notions of geography, such as road networks, would be an appealing direction for future research.



(a) Mean Log Likelihoods

(b) Difference in Mean Log Likelihoods (SPACE minus moving cost)

Figure 5: Log Likelihood of SPACE and moving cost models

In Panel (a), we show the overall likelihood for both models. Both likelihoods decrease over time, as it becomes harder to forecast where people live. In Panel (b), we plot the difference in likelihood between the two models in order to focus on model comparison. While the moving cost model does a little bit better in the first few years,[30] the SPACE model has a higher likelihood in later years, and has an average log likelihood that is more than 8 percent larger per observation by 2018.

The SPACE model does better over time because it can match the dynamic moments. In particular, many more people end up moving away from their initial state in the moving cost model because moving probabilities are independent over time, whereas the SPACE model is better able to match the total number of people who leave.

4.2 Moving costs need not be large

Kennan and Walker (2011) estimate an average moving cost of $312,146 (in 2010 dollars).[31] This is more than six times the median household income in that year, which was $49,445 (Census, 2011).[32] Such a large cost is one of the main reasons that

[30]This is because the Markov model is calibrated to exactly match migration for every bilateral state pair, whereas the SPACE model is only calibrated to match the gravity relationship. The moving cost model calibrates thousands parameters, whereas the SPACE model has only fifty-three.

[31]They also include an analysis of moving costs conditional on moving, but they include the payoff shocks in the moving costs, so find that the average moving cost is actually very negative.

[32]Most papers focusing on the United States estimate moving costs of a similar magnitude (Mon-

most people do not move, in their model. Economists can argue about whether that number is reasonable, and even within Kennan and Walker (2011), there is substantial heterogeneity in moving costs.

In contrast, the persistent spatially-correlated preferences model can match the main facts about internal migration without any moving costs.[33] In other words, the fact that most people do not move is not sufficient evidence to conclude that moving costs are large.

A common counterfactual in the literature is to consider changes in moving costs (Kennan and Walker, 2011; Schubert, 2021; Zerecero, 2021), which is also an actual policy used by some localities.[34] For example, Kennan and Walker (2011) finds that a moving subsidy could substantially increase the gross migration rate. In a moving cost model, a temporary incentive to move to location i has a very persistent impact on the population of i. In contrast, in our model, a moving subsidy would encourage people to relocate, but only for as long as the subsidy lasts. After the subsidy expires, they are no more likely to remain in the place they moved to, than they would be to live there had the subsidy never occurred.[35]

One particular way of lowering "moving costs" may be improving infrastructure such as roads, which increases migration (Morten and Oliveira, 2018). The SPACE model would need to be modified to account for this fact, along the lines of the extensions from Section 3.2. Such a modification would assume that roads increase the correlation of the idiosyncratic shocks between two places, which raises the migration between them, even though moving costs have not gone down. In the moving cost model, the increased migration must reflect lower moving costs, and so welfare would have increased, net the cost of the roads. In the modified SPACE model, though, the welfare effects are much more muted.[36]

---

ras, 2018; Bartik and Rinz, 2018). Tombe and Zhu (2019) also estimates very large moving costs in China, on the order of 50 percent of annual income within province, and 90 percent of annual income across provinces. However, their measure of migration is a flow cost, borne by the person every year. Hence, in distinguishing between moving costs and persistent preferences, it maps more naturally onto a persistent preference to not live outside the origin location.

[33]Including birthplace as a state variable in a moving cost model, which includes some persistence in preferences, lowers estimates of moving costs by about 10 percent (Zerecero, 2021).

[34]A handful of cities around the United States offer monetary incentives to relocate (Cornerstone Home Lending, 2021).

[35]Of course, one could imagine other economic reasons that populations remain higher after a population expansion, such as the accumulation of housing capital or agglomeration in that place.

[36]Of course, other welfare benefits, such as those that come through increased trade, do not depend on how migration is modeled.

4.3 Macro Population Elasticities

Another common use for migration models is to calculate population elasticities to changes in a location's utility, v_i, both in the short-run or the long-run.

In the short-run, the elasticities are relatively similar, which we formalize in the following two propositions. In each, we consider the cross-elasticity of population in i to v_j. In both the SPACE model and the moving cost model, we show that when there is little migration, then the interstate migration rate is related to the elasticity.

Proposition 3. *In the steady-state of the SPACE model, as $\rho \to 1$, the elasticity of population with respect to the utility of another state is proportional to the migration rate between the two states, adjusted for the covariance of the match-specific utility:*

$$\frac{\partial \log p_i}{\partial v_j} = -\lim_{\rho \to 1} \frac{m_{i \to j}}{p_i} \frac{1}{\sqrt{1 - \Sigma_{ij}}} \sqrt{\frac{\pi}{1 - \rho^2}} \tag{8}$$

when $i \neq j$.[37]

This proposition says that the cross-utility of population to a shock in another state is related to the amount of migration between the two states.[38] One intuition is that the amount of migration between two states represents how many people are roughly indifferent between the two states. So if a shock hits one state and not the other, migration is roughly proportional to the number of people that might be induced to move in response to the shock.

The reason the expression has to adjust for $(1 - \Sigma_{ij})^{-1/2}$ is that the individual-level shocks that induce steady-state migration—the η's—are spatially correlated, and so when calculating the effect of a shock to only one state, you have to adjust the gross migration rate to account for that correlation.

[37] π is not new notation; it is the common mathematical constant representing the ratio of a circle's circumference to its diameter.

Note we could also write the left-hand side of equation (8) as $\partial \log p_i / \partial u_j$ since v_j and u_j differ only by a constant. However, writing it with respect to v_j provides a closer analog to our next proposition which concerns the moving cost model.

[38] Traditionally, one of the issues with the multinomial probit model is that these elasticities of interest have no closed form solution. However, we overcome this hurdle in the migration context by considering the limit when there is little migration, and by showing that the elasticity is related to a sufficient statistic: the observed amount of migration between places.

Note that the limit in equation (8) does not diverge to infinity because as ρ gets close to 1, the migration between the two states falls proportionally to $\sqrt{1 - \rho^2}$.

Proposition 4. *In the steady-state of the moving cost model, as $\Delta \to \infty$, the elasticity of population in state i to the common utility in state j is proportional to the migration rate between the two states:*

$$\lim_{\Delta\to\infty} \frac{\partial \log p_i}{\partial v_j} \exp(\Delta) = - \lim_{\Delta\to\infty} \frac{m_{i\to j} + m_{j\to i}}{p_i} \exp(\Delta) \tag{9}$$

when $i \neq j$.

There expression includes terms for both inmigration and outmigration because they are not necessarily the same in the steady-state of the moving cost model. In the SPACE model, these two terms are the same in steady-state, so only one shows up in the expression.[39]

The elasticities in the two models are similar in that they are approximately proportional to the migration rate when migration rates are low. Given that the scale of v_j is not specified in either model, the constant terms are ignorable without loss of generality. So the short-term elasticities in the two models only really differ by the term $(1 - \Sigma_{ij})^{-1/2}$. The variance in this term is not that large compared to the variance in migration rates, so we can interpret both propositions as saying that the short-term elasticity is roughly proportional to the migration rate.[40]

However, in the long-run, the similarities of population elasticities between the SPACE model and the moving cost model break down. This is illustrated in the following two propositions:

Proposition 5. *In steady-state, the long-run elasticity $\frac{\partial \log p_i}{\partial v_j}$ of the SPACE model is the same as in the short-run.*

The intuition for this proposition is that even though individuals have evolving preferences over time, the distribution of match-specific utility is stationary. So the number of people choosing a location does not change except when the v_j's change.

In Appendix E.3, we verify in the calibrated version of the SPACE model, where $\rho < 1$, that the long-run population elasticities are similar to those of Propositions 3 and 5. For that exercise, we calibrate the SPACE model in 1990 and in 2018 to

[39] Also note that just as in the previous proposition, the limit does not diverge to infinity when $\Delta \to \infty$ because the amount of migration is falling at the same rate.

[40] In Appendix C, we use variation from Hurricane Katrina to show empirical evidence that the SPACE model approximation—with the adjustment for $(1 - \Sigma_{ij})^{-1/2}$—is closer to the data than the moving cost model, but we do not think this is a first-order difference between the two models.

match the population shares of all the states in both years. We compare the utility changes implied by that model to the utility changes implied by equation (8).[41]

Proposition 6. *In steady-state, the long-run elasticity* $\frac{\partial \log p_i}{\partial v_j}$ *of a moving cost model is not the same as in the short-run. Rather, as* $\Delta \to \infty$*,*

$$\lim_{\Delta \to \infty} \frac{\partial \log p_i}{\partial v_j} = -2p_j \tag{10}$$

when $i \neq j$*.*

Note that since the total population is mass 1, p_i is both the population of i and its population share.

Interestingly, these steady-state elasticities are the same as a static logit, and a key difference from the SPACE model is that they have no relationship to migration data. In Appendix E.4, we show equation (10) is a good numerical approximation to the long-run of a calibrated moving cost model, when moving costs lead to realistic migration rates, rather than being infinite.

In the long-run, the moving cost model has no notion that closer states are better substitutes or that states with higher migration are likely to be more impacted by a change in the other state. The moving cost model would not predict that a state with a high migration rate has a more long-run elastic population in response to a policy change than a state with a low migration rate. Rather, the only thing that you need to know is the population share of the state receiving the shock to calculate all the relevant elasticities (approximately).[42]

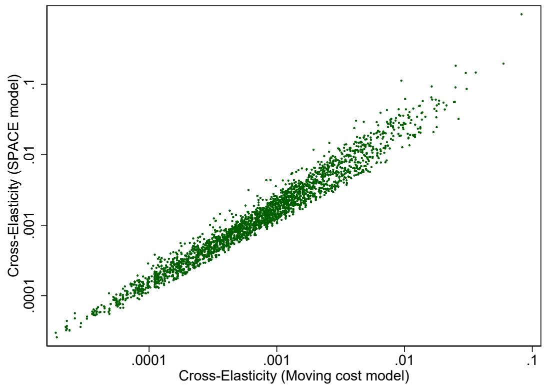
Figure 6 summarizes the conclusions of this section. In the short-run (Panel a), the SPACE model and the moving cost model predict practically the same cross-elasticities of population. But the in the long-run (Panel b), there is almost no relationship between the two.

[41]We construct the full-matrix of population elasticities to utilities. By inverting that matrix, we can calculate the implied utility changes were for any vector of population changes, up to a normalization.

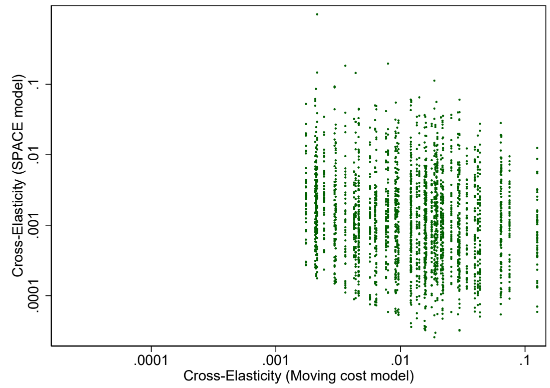
[42]In the moving cost model, the migration rates govern the speed of adjustment (Kleinman et al., 2023), but not the long-run effects.

Including birthplace as a state variable in the moving cost model would mean that adjustments would depend on the population shares of people born in different places. To this extent, migration between states would be correlated to the population cross-elasticities (Zabek, 2020).

Other features of the model can also change the population elasticities. For example, Monte, Redding and Rossi-Hansberg (2018) adds commuting to a static model of location choice to generate variation in population elasticities.



(a) Short-run



(b) Long-run

Figure 6: A comparison of the population cross-elasticities between the SPACE and moving cost models. For both figures, each dot represents a pair of states. The point is located at the population cross-elasticity between the two states in each of the two models, as given by Propositions 3-6. The constant multiplicative terms are ignored, since each model is subject to a normalization of utility. All four axes have log scales.

4.4 Macro dynamics

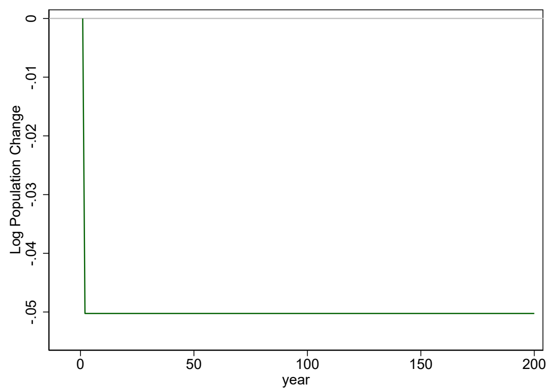
Given the differences in long-run population elasticities, it follows that the intermediate dynamics must also be different across the two models. We illustrate by considering a one-time permanent shock to Louisiana utility, $v_{\text{Louisiana}}$, to see how populations respond over time.

In Figure 7(a) and (b), we show that the population dynamics after a one-time permanent shock are starkly different. In both, we consider a one-time permanent change to the baseline utility of Louisiana, leaving all other states' utilities constant, and we simulate both models for many periods.[43]

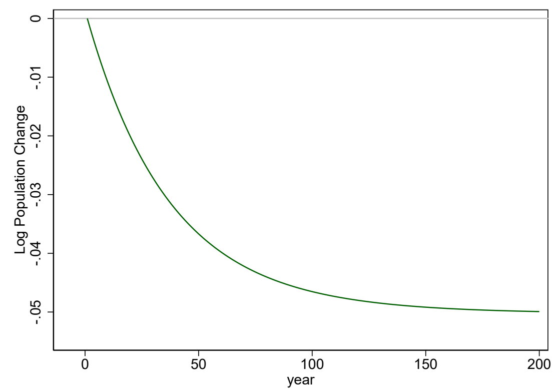
In the SPACE model (Panel a), the population adjustment in Louisiana is immediate, and the population stops adjusting after the first period. In contrast, in the moving cost model (Panel b), the population adjustment takes many years, with the model finally getting close a steady-state after almost 200 years.

In Panels (c) and (d) we illustrate the dynamics for other states in response to the same shock to Louisiana's utility. In the SPACE model (Panel c), there is a bigger

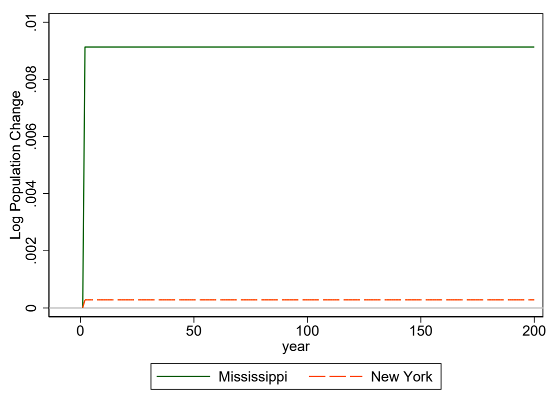
[43]We chose Louisiana because we do an exercise about Hurricane Katrina in Append C, but any state would work as well. The size of the shocks in each model is normalized to have a long-term effect of about 5 percent of the population for Louisiana. However, the scales are not particularly important, as the focus of this exercise is the dynamics.



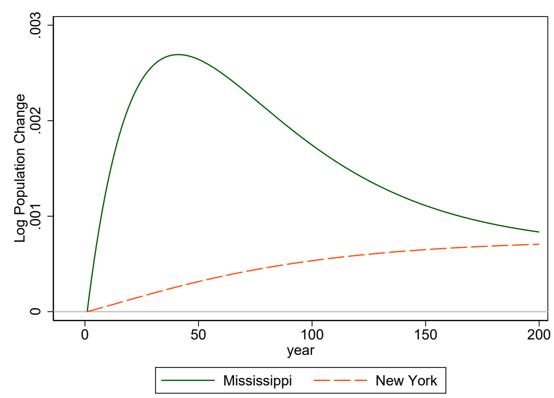
(a) Louisiana, SPACE model



(b) Louisiana, moving cost model



(c) Mississippi and New York, SPACE model



(d) Mississippi and New York, moving cost model

Figure 7: Population Dynamics after a one-time permanent change in $v_{\text{Louisiana}}$, in the SPACE model and the moving cost model, for Louisiana, Mississippi, and New York. Mississippi and New York were chosen to represent two states for which there is high gross migration with Louisiana, and low gross migration with Louisiana, respectively.

population effect on Mississippi than there is on New York, as one would expect due to the geography. Again the dynamics are immediate. But in Panel (d), the dynamics follow interesting and perhaps unintuitive patterns. In New York, the population adjustment is particularly slow because of low migration between Louisiana and New York. In contrast, for Mississippi, the population dramatically overshoots its long-run steady state because there is so much migration between Louisiana and Mississippi.

This exercise is not necessarily helpful for distinguishing between the two models, but it does mean that the models will interpret empirical facts differently. We discuss this more in the next section.

4.5 Implied Utility Changes

Given the differences in population elasticities, it must be the case that the models will imply different things about changes in utility over time. This is important for papers that wish to estimate the welfare effect of some policy or event that varies across space. For example, Diamond (2016) asks what determines why people live in different places over time.

In the SPACE model population responds according to the equation: $dp = \frac{\partial p}{\partial v} dv$.[44] So relative utility changes can be solved by

$$dv^* = \left(\frac{\partial p}{\partial v}^* \right)^{-1} dp^* \tag{11}$$

where dp^* is the change in populations when dropping one state, and $\frac{\partial p}{\partial v}^*$ is the matrix of population responses to utility changes, when dropping the corresponding row and column, and dv^* is the change in utilities relative to the dropped state.

In contrast, for the moving cost model, populations are not the right statistic to look at to calculate utility. Instead, economists can look at interstate migration, which is given by:

$$\frac{\partial \log m_{i \to j}}{\partial v_k} = \begin{cases} m_{i \to k} & \text{if } j \neq k \\ 1 - m_{i \to j} & \text{if } j = k \end{cases}$$

[44]To write down the matrix $\frac{\partial p}{\partial v}$, one also needs the population elasticity to the state's own utility. This is given by

$$\frac{\partial \log p_i}{\partial v_i} = - \sum_j \frac{\partial \log p_i}{\partial v_j}$$

where the right-hand side can be found using equation (8).

Since migration rates are close to zero, the elasticities are close to 0 and 1. In other words, the change in utility is close to the percent increase in migration to the city. Since there are in principle 51^2 changes in migration to calculate 51 changes in utility, the model is overidentified.

A typical way to estimate the changes in utility would be to run the regression

$$d \log m_{i \to j} = dv_j + \alpha_i + \epsilon_{ij} \tag{12}$$

where α_i is a origin-state fixed effect, and ϵ_{ij} is some unmodeled error term.

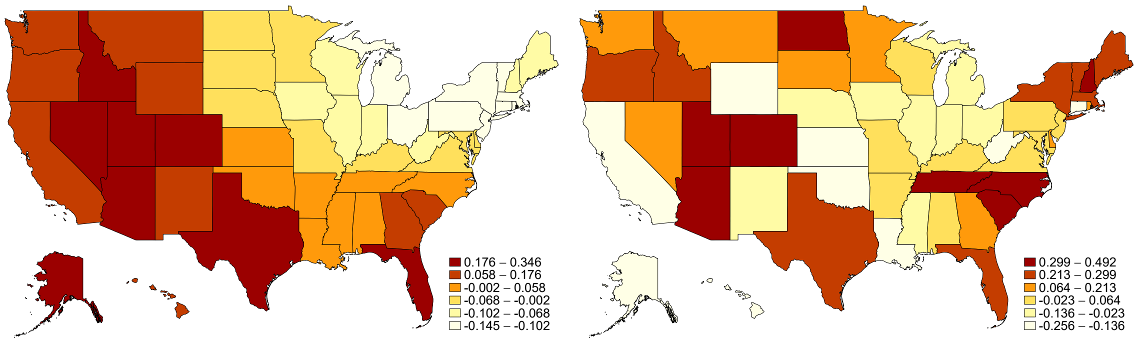
Fundamentally, there is a difference in what data identifies the utility of a location. In the SPACE model, it is the populations, and in the moving cost it is the migration rates. In the SPACE model, a place has gotten better to live if people have moved there. In the moving cost model, a place has gotten better to live if the rate at which people move there has gone up.

To illustrate this, we consider the utility changes implied by the SPACE model and moving cost model from 1990-2018.[45] In the SPACE model, the places that have the biggest increase in utility are in the South and West. New England and the Rust Belt have some of the largest decreases. In the moving cost model, the utility changes are less spatially concentrated. New York and New England have increased in relative utility, while California has actually declined in relative utility. Overall, there is only a 0.16 correlation between the utility changes implied by the two models.

This has important implications for estimating spatial models. For example, if one wanted to estimate the effects of a wage, rent, or amenity change on utility, you would get very different answers using the implied utilities from the SPACE model versus a moving cost model.

It should be noted that there are differences between the implied utility changes in the SPACE model and the static logit model as well. In the static logit model, there is a linear mapping between relative utility changes and log population growth. In Appendix E.3, we show that while there is a high correlation, there are significant deviations from a linear relationship of population growth and utility changes implied by the SPACE model.

[45] 1990 is the first year in which we have full state-to-state migration flows in the IRS data.



(a) SPACE model (b) Moving cost model

Figure 8: Change in utilities v_j, 1990-2018, implied by the SPACE model and the moving cost model.

5 Discussion

5.1 Relation to Literature

Given the major differences between the two models on several of these questions, it is worth emphasizing why these differences matter. For the micro forecasting and for the interpretation, we think the reasons to care about differences are obvious, but for the population elasticities and the dynamics, it is important to consider the context of the literature.

One big question in the literature is to what extent does population adjust to shocks? For example, if one particular location has a shock that permanently increases the utility of living there, how will that affect the distribution of the population around the country? This is a question that is asked by Caliendo et al. (2019) with respect to the China shock of Autor, Dorn and Hanson (2013), by Giannone (2017) with respect to skill-biased technological change, and by Cruz and Rossi-Hansberg (2021), Oliveira and Pereda (2020), and Rudik et al. (2021) with respect to climate change.[46]

[46] A reader may wonder why we do not replicate one of these papers to highlight the differences. However, doing such a replication would erroneously indicate that the SPACE model is less good at hitting the medium-run dynamics of population adjustment. The reason for this is that even if these models are not explicitly targeting the medium-run dynamics, they do get to choose what features of the world to add and can choose to include or not include features that will get the dynamics right. For example, Glaeser and Gyourko (2005) argues that the reason declining cities decline slowly is because the housing stock is slow to depreciate. Many of the quantitative papers do not have this feature. So of course, subbing out the migration block from those models would lead to unrealistic dynamics, if we did not also add in a feature like the one in Glaeser and Gyourko (2005). Sometimes, people tell us that they think of the moving costs as representing these other

Based on the previous propositions, both models agree that places with high gross migration, such as D.C., have very elastic population to local shocks in the short-run, compared to places with little migration. However, in the long-run, the SPACE model continues to make this prediction, while the moving cost model predicts similar elasticities for all locations. Similarly, in the short-run, both models agree that the population effects are felt in states that have lots of migration between them and the state with the shock. A shock to D.C. will affect Maryland and Virginia more than it will affect Arizona. This is consistent with the "donut" phenomenon during the recent COVID-19 crisis, as areas around major cities have experienced population and house price growth in recent years (Ramani and Bloom, 2021). Again, this holds in the long-run too for the SPACE model, but migration would not generate a long-run donut phenomenon in the moving cost model.

Another key question in this literature is how quickly the migration adjustment takes place (Kleinman et al., 2023; Amior and Manning, 2018; Caliendo et al., 2019). Proposition 3 answers this question in that population adjustment occurs as quickly as the baseline utility of a place changes.

In the data, migration is usually quite persistent. For example, the Rust Belt has had low inmigration for decades, and the Sun Belt has had high inmigration for decades. In the moving cost model, much of this persistence is due to the fact that migration is inherently persistent (Kleinman et al., 2023), i.e. the Rust Belt had a large negative utility shock a long time ago, and the process of moving out has been very slow.

The SPACE model interprets this fact as being about the utility of a location adjusting slowly. It could be that the underlying shocks to utility are slow. Or there was still a big initial shock, but some equilibrium force makes utility fall slowly. For example, housing is durable, and so housing becomes cheap as people move out, keeping utility from falling too quickly (Glaeser and Gyourko, 2005). Similarly, there may be similar mechanisms through the labor market that make structural transformation slow.[47] Or it could be that initial shocks are small, but then as people move in, the

features, such as housing depreciation or labor market frictions. In that case, we think it is better to explicitly model them. However, we do not think such a model is within the scope of this paper. Writing a completely new model, while it might add to the point about macro elasticities, would distract from highlighting the important differences in micro forecasting and interpretation.

[47] Kleinman et al. (2023) discuss how having location specific durable capital can keep wages high after a negative productivity shock.

effects on utility become amplified with a delay (Howard, 2020). The SPACE model would emphasize these various forces as reasons for persistence in migration, whereas the moving cost model would attribute the persistence to an inherent property of migration itself, and find less explanatory power for these forces.

Other papers are concerned with the effects of location-specific shocks on aggregate outcomes such as welfare or output (Tombe and Zhu, 2019; Eckert and Peters, 2018; Hsieh and Moretti, 2019). The degree to which people are able to move is an important factor for these outcomes. Because the second order effects are determined by these population elasticities, it shows that a shock that affects a higher-gross-migration place will have larger total effects on welfare if the shock is positive, and smaller effects if the shock is negative. Again, this holds in the short-run for both models, but in the long-run only for the SPACE model.

Finally, the spatial correlation of a shock is an important determinant of its welfare consequences. If a negative shock is extremely localized, it may be easy to move away from it, and there will be lots of insurance. If shocks are correlated across space, then the welfare effects may be much less insurable. Of course, in the long-run of a moving cost model, this effect will no longer hold.

5.2 More complex moving cost models

In Section 4, we compared the SPACE model to a very simple version of a moving cost model. Yet as mentioned in the literature review, there is significant research that enriches the moving cost model to match a variety of facts (Kaplan and Schulhofer-Wohl, 2017; Giannone et al., 2020; Porcher, 2020; Mangum and Coate, 2019; Zerecero, 2021; Monras, 2018). Even Kennan and Walker (2011) includes features to increase home bias and return migration.

As we showed in Section 2.1, return migration, home bias, nor age are sufficient features to hit the square root fact. So the puzzle that motivated the model is not dependent on us having considered a simple version of the moving cost model. But what about evaluating the differences between the moving cost model and the SPACE model? Are the conclusions that the choice of model is important robust to considering richer versions of the moving cost model? Here, we argue that the answer is yes. In the rest of this section, we consider each of the differences we highlighted before.

For the prediction of individuals' locations, the reason that the SPACE model outperformed the moving cost model was because it could hit the square root fact. So if extensions of the moving cost model still do not hit the square root fact, they might be an improvement at predicting locations, but are not going to make the same predictions as the SPACE model.

For the interpretation of why people rarely move, some of the additional features imply lower estimated moving costs (Zerecero, 2021; Giannone et al., 2020), but never by orders of magnitude. So the difference between the SPACE model—which has no moving costs—and any moving cost model will remain large.

For population elasticities, more complex moving cost models feature long-run elasticities may not necessarily be the same as a static logit model. For example, models with home bias have more similar elasticities to the SPACE model than the baseline moving cost model does.[48] However, except by coincidence, none of the additional features would generate the feature that the long-run and short-run population elasticities are the same. So the dramatic difference between the SPACE model and the moving cost model will remain, both for long-run elasticities and for dynamics.

For the implied utility changes, the exact implied utilities will obviously change with a richer model. However, it doesn't change the fundamental fact that the implicit utility is a function of migration in the moving cost model, but populations in the SPACE model.

5.3 A simpler moving cost approximation

Of course, even if thought the SPACE model was the true model, there are advantages to an economic modeler from the standard moving cost model, in particular its closed-form solutions.[49] However, the medium- and long-run population elasticities that such a model produces are not robust to the method of modeling why people rarely move. If the true model were the SPACE model, then the standard moving cost model would predict incorrect counterfactuals and welfare effects.

[48]With home bias, the elasticities are given by $\lim_{\Delta\to\infty} \partial \log p_j / \partial v_k = -\sum_i w_{ij} p_{ik}$, where $w_{ij} = \frac{p_{ij}}{p_j}$ is the share of people in j who are from i, and p_{ik} is the share of people from i living in j. These population shares are likely correlated to the amount of bilateral migration. However, it is a different formula, and the population share levels are likely different than the migration rates, so the dynamics in the two models will still be different.

[49]Recall that the closed-form solutions for the SPACE model were only approximate in that they relied on $\rho \to 1$.

The good news is that a it is straightforward to write down a model with elasticities similar to the SPACE model but tractability similar to the moving cost model. In fact, the moving cost model from Section 2 only needs one modification to it.

The following equation is the same as the standard model (equation 2), but the i in parentheses at the end has been replaced with a j.

$$V_{nt}(j) = \max_i \{u_{it} - \delta_{ij} + \epsilon_{int} + \beta \mathbb{E} V_{n,t+1}(j)\} \tag{13}$$

The way to interpret the model is as if the moving cost is paid every single period, rather than just once, giving it a similar flavor to a persistent preference.[50]

The population elasticities are given by the short-run population elasticities for the moving cost model, which we previously showed were similar to the short- and long-run population elasticities of the SPACE model. But this model still has a closed form solution, so the hat algebra techniques that are common in this literature (Caliendo et al., 2019) can still be applied.[51]

This trick works well for many questions in the literature, especially those discussed in Section 4.3. However, it is not a cure-all for any question. For one thing, the interpretation of the model is not straightforward, and so we think it is best thought of a trick to make the SPACE model a bit more tractable when modeling population changes. For another thing, it will not match any of the micro dynamics that the first part of the paper focused on, so if the researcher is interested in those questions, this trick cannot be used. Nonetheless, we want to emphasize that the logit model is commonly used for a reason, and the tractability advantages when this trick works are likely to be sizable.[52]

[50]If we wanted to, we could move the continuation value term outside of the maximization since it does not depend on i, as in the SPACE model.

Whether ϵ_{int} is drawn once and kept forever or drawn repeatedly makes no difference for the macro population elasticities. In neither case will the model lead to realistic micro dynamics.

[51]Of course, if a modeler wanted to linearize the model as in Kleinman et al. (2023), then they could just use equation (8) for the relevant elasticities, without any need for this trick.

[52]One interpretation of this result is that we are arguing to bring back static logit models as a way to model location choice (e.g. Rosen, 1979; Roback, 1982), rather than the dynamic models that the literature has largely adopted. We have two comments on this. First, our model still has a role for gross migration in determining elasticities, which old static models did not typically feature (recall that the moving costs in equation (13) are calibrated to match the migration in the data). Second, this model is motivated as a trick to approximate the SPACE model, which is primarily focused on more realistically matching the micro-dynamics. So saying that this model is somehow abandoning the dynamics of migration is ignoring the fact that the SPACE model does a better job of matching the micro dynamics.

6 Conclusion

We use a dataset of credit reports to document a new fact: the t-year migration rate is proportional to the square root of t. We propose a new model to match this fact, which has different implications for many economic questions than the standard moving cost model.

References

Allen, Treb and Dave Donaldson, "Persistence and path dependence in the spatial economy," 2020. National Bureau of Economic Research Working Paper.

Amior, Michael and Alan Manning, "The persistence of local joblessness," *American Economic Review*, 2018, *108* (7), 1942–70.

Autor, David, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Bartik, Alexander W and Kevin Rinz, "Moving costs and worker adjustment to changes in labor demand: Evidence from longitudinal census data," 2018.

Bayer, Christian and Falko Juessen, "On the dynamics of interstate migration: Migration costs and self-selection," *Review of Economic Dynamics*, 2012, *15* (3), 377–401.

Bhat, Chandra R and Jessica Guo, "A mixed spatially correlated logit model: formulation and application to residential choice modeling," *Transportation Research Part B: Methodological*, 2004, *38* (2), 147–168.

Blanchard, Olivier Jean and Lawrence F Katz, "Regional evolutions," *Brookings Papers on Economic Activity*, 1992, *1992* (1), 1–75.

Bryan, Gharad and Melanie Morten, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Butler, John S and Robert Moffitt, "A computationally efficient quadrature procedure for the one-factor multinomial probit model," *Econometrica*, 1982, pp. 761–764.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Census, U.S., "Income, Poverty and Health Insurance Coverage in the United States: 2010," 2011.

Coen-Pirani, Daniele, "Understanding gross worker flows across US states," *Journal of Monetary Economics*, 2010, *57* (7), 769–784.

Correia, Sergio, Paulo Guimaraes, and Thomas Zylkin, "PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects," 2019.

Cruz, José-Luis and Esteban Rossi-Hansberg, "The economic geography of global warming," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-130).

Davis, Morris A, Jonas DM Fisher, and Marcelo Veracierto, "Migration and urban economic dynamics," *Journal of Economic Dynamics and Control*, 2021, *133*, 104234.

DeWaard, Jack, Janna Johnson, and Stephan Whitaker, "Internal migration in the United States: A comprehensive comparative assessment of the Consumer Credit Panel," *Demographic research*, 2019, *41*, 953.

_ , Mathew Hauer, Elizabeth Fussell, Katherine J Curtis, Stephan Whitaker, Kathryn McConnell, Kobie Price, and David Egan-Robertson, "User Beware: Concerning Findings from Recent US Internal Revenue Service Migration Data," 2020. Minnesota Population Center Working Paper No. 2020-02.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Eckert, Fabian and Michael Peters, "Spatial Structural Change," 2018.

Farrokhi, Farid and David Jinkins, "Dynamic amenities and path dependence in location choice: Evidence from danish refugee placement," 2021.

Fonseca, Julia, "Less Mainstream Credit, More Payday Borrowing? Evidence from Debt Collection Restrictions," *Journal of Finance*, 2022.

_ and Jialan Wang, "How Much do Small Businesses Rely on Personal Credit?," 2022.

Fujiwara, Thomas, Eduardo Morales, and Charly Porcher, "A Revealed Preference Approach to Measuring Information Frictions in Migration Decisions," 2022.

Geweke, John, Michael Keane, and David Runkle, "Alternative computational approaches to inference in the multinomial probit model," *The Review of Economics and Statistics*, 1994, pp. 609–632.

Giannone, Elisa, "Skill-Biased Technical Change and Regional Convergence," 2017.

_ , Qi Li, Nuno Paixao, and Xinle Pang, "Unpacking Moving," 2020.

Gies Consumer and Small Business Credit Panel, 2004-2018. Accessed July 20, 2021.

Glaeser, Edward L and Joseph Gyourko, "Urban decline and durable housing," *Journal of Political Economy*, 2005, *113* (2), 345–375.

Han, Peter, "The Bank Premium and Market Concentration in the Mortgage Market," 2022.

Hao, Tongtong, Ruiqi Sun, Trevor Tombe, and Xiaodong Zhu, "The effect of migration policy on growth, structural change, and regional inequality in China," *Journal of Monetary Economics*, 2020, *113*, 112–134.

Heise, Sebastian and Tommaso Porzio, "The aggregate and distributional effects of spatial frictions," 2021. National Bureau of Economic Research Working Paper.

Howard, Greg, "The migration accelerator: Labor mobility, housing, and demand," *American Economic Journal: Macroeconomics*, 2020, *12* (4), 147–79.

_ and Jack Liebersohn, "Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities," *Journal of Urban Economics*, 2021, *124*, 103369.

_ , _ , and Adam Ozimek, "The Short-and Long-Run Effects of Remote Work on US Housing Markets," *Journal of Financial Economics*, 2023.

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

IRS Migration Data, "Internal Revenue Service Statistics of Income Tax Statistics - Migration Data," 2004-2018. https://www.irs.gov/statistics/soi-tax-stats-migration-data.

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak, "The economics of internal migration: Advances and policy questions," *Journal of Economic Literature*, 2023, *61* (1), 144–180.

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the long-run decline in interstate migration," *International Economic Review*, 2017, *58* (1), 57–94.

Keane, Michael P, "A note on identification in the multinomial probit model," *Journal of Business & Economic Statistics*, 1992, *10* (2), 193–200.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kleemans, Marieke, "Migration choice under risk and liquidity constraints," 2015.

Kleinman, Benny, Ernest Liu, and Stephen Redding, "Dynamic spatial general equilibrium," *Econometrica*, 2023.

Koşar, Gizem, Tyler Ransom, and Wilbert Van der Klaauw, "Understanding migration aversion using elicited counterfactual choice probabilities," *Journal of Econometrics*, 2021.

Lending, Inc. Cornerstone Home, "These 11 U.S. Cities Will Pay You to Live There," 2021. Accessed August 9, 2022. https://www.houseloanblog.net/get-paid-to-relocate/.

Mangum, Kyle and Patrick Coate, "Fast locations and slowing labor mobility," 2019.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles, "IPUMS National Historical Geographic Information System," 2021.

Molloy, Raven, Christopher L Smith, and Abigail Wozniak, "Internal migration in the United States," *Journal of Economic Perspectives*, 2011, *25* (3), 173–96.

Monras, Joan, "Economic shocks and internal migration," 2018.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg, "Commuting, migration, and local employment elasticities," *American Economic Review*, 2018, *108* (12), 3855–90.

Morris-Levenson, Joshua and Marta Prato, "The Origins of Regional Specialization." PhD dissertation, The University of Chicago 2022.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *NBER Working Paper*, 2018, *22158*, 1–64.

Oliveira, Jaqueline and Paula Pereda, "The impact of climate change on internal migration in Brazil," *Journal of Environmental Economics and Management*, 2020, *103*, 102340.

Panel Survey of Income Dynamics, 1969-1997. https://psidonline.isr.umich.edu/.

Porcher, Charly, "Migration with costly information," 2020.

Ramani, Arjun and Nicholas Bloom, "The Donut effect of COVID-19 on cities," 2021. National Bureau of Economic Research Working Paper.

Roback, Jennifer, "Wages, rents, and the quality of life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rogerson, Peter A, "A new method for finding geographic centers, with application to US States," *The Professional Geographer*, 2015, *67* (4), 686–694.

Rosen, Sherwin, "Wage-based indexes of urban quality of life," *Current Issues in Urban Economics*, 1979, pp. 74–104.

Rudik, Ivan, Gary Lyn, Weiliang Tan, and Ariel Ortiz-Bobea, "The Economic Effects of Climate Change in Dynamic Spatial Equilibrium," 2021.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 6.0 [dataset]," http://doi.org/10.18128/D010.V6.0 2015. Minneapolis: University of Minnesota.

Saks, Raven E and Abigail Wozniak, "Labor reallocation over the business cycle: New evidence from internal migration," *Journal of Labor Economics*, 2011, *29* (4), 697–739.

Sandev, T. and Z. Tomovski, *Fractional Equations and Models*, Developments in Mathematics, 2019.

Schoenberg, Isaac J, "Metric spaces and positive definite functions," *Transactions of the American Mathematical Society*, 1938, *44* (3), 522–536.

Schubert, Gregor, "House Price Contagion and US City Migration Networks," 2021.

Silva, JMC Santos and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 2019, *109* (5), 1843–72.

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge University Press, 2009.

Vincenty, Thaddeus, "Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations," *Survey review*, 1975, *23* (176), 88–93.

Zabek, Michael A, "Local ties in spatial equilibrium," 2020.

Zerecero, M, "The Birthplace Premium," 2021.

Online Appendix

A Proofs

A.1 Proof of Proposition 1

Proof: In the steady-state of the moving cost model, consider the probability of someone living in i living in j after t years. This probability is given by:

$$\frac{m_{i\to j}^t}{p_i} = \sum_{k_1,k_2,k_3,\ldots,k_{t-1}\in I} \frac{m_{i\to k_1}}{p_i}\left(\prod_{s=1}^{t-2}\frac{m_{k_s\to k_{s+1}}}{p_{k_s}}\right)\frac{m_{k_{t-1}\to j}}{p_{k_{t-1}}}$$

Consider

$$\lim_{\Delta\to\infty}\frac{m_{i\to j}^t}{m_{i\to j}} = \lim_{\Delta\to\infty}\sum_{k_1,k_2,k_3,\ldots,k_{t-1}\in I}\frac{\frac{m_{i\to k_1}}{p_i}\left(\prod_{s=1}^{t-2}\frac{m_{k_s\to k_{s+1}}}{p_{k_s}}\right)\frac{m_{k_{t-1}\to j}}{p_{k_{t-1}}}}{\frac{m_{i\to j}}{p_i}}$$

We can calculate each term in the summation. First, consider all the summations such that there exists a cutoff T such that for all $s < T$, $k_s = i$ and for all $s \geq T$, $k_s = j$. There are t such combinations. For each combination, the product in the numerator is a lot of migration probabilities from i to i (non-migration), and one migration probability from i to j. For the non-migration probabilities, the limit as $\delta\to\infty$ of $\frac{m_{i\to i}}{p_i}$ or $\frac{m_{j\to j}}{p_j}$ is 1. The remaining term, $\frac{m_{i\to j}}{p_i}$ cancels with the denominator, so the limit is 1.

Next consider all other terms. For each, there is one year in which the person moves away from i and another year in which they move from their next location to somewhere else. Put these two terms first in the product:

$$\frac{m_{i\to\ell}/p_i}{m_{i\to j}/p_i}\cdot\frac{m_{\ell\to k_s}}{p_\ell}\cdot\ldots$$

The first fraction is $\frac{\exp(v_{\ell t}-\delta_{i\ell})}{\exp(v_{jt}-\delta_{ij})}$ regardless of Δ, which is a constant. The second term converges to zero when $\Delta\to\infty$. All the following terms are between 1 and 0, so the whole product converges to 0.

Therefore, the sum is over t 1's and a lot of zeros. Hence,

$$\lim_{\Delta \to \infty} \frac{m_{i \to j,t}}{m_{i \to j}} = t$$

$\square$

A.2 Proof of Observation 1

If $f(x) = \exp(-a\sqrt{x})$ is completely monotone and non-constant, then Σ is positive definite by the Schoenberg Interpolation Theorem (Schoenberg, 1938).[53] f is clearly non-constant, so it remains to show that $f(x)$ is completely monotone. Note that $\exp(-ax)$ is completely monotone. Note also that $\sqrt{x}$ is a Bernstein function, meaning that its derivative is completely monotone. The composition $g(h(x))$ of a completely monotone function g and a Bernstein function h is completely monotone (Sandev and Tomovski, 2019). Therefore, $\exp(-a\sqrt{x})$ is completely monotone. $\square$

A.3 Proof of Proposition 2

Consider people who are living in i, but close to indifferent between living in i and j (As $\rho \to 1$, the number of people that might move between three places, as a percentage of those that might move to two places, diminishes to zero). A person moves from i in year 0 to j in year t if $u_i + \epsilon_{in0} > u_j + \epsilon_{jn0}$ but $u_j + \epsilon jnt > u_i + \epsilon_{int}$. By the AR(1) nature of the utility shocks,

$$u_i + \epsilon_{int} - u_j - \epsilon_{jnt} = u_i - u_j + \rho^t(\epsilon_{in0} - \epsilon_{jn0}) + \sum_{s=1}^{t} \rho^{t-s}\sqrt{1-\rho^2}(\eta_{ins} - \eta_{jns}) \quad (14)$$

The distribution of the cumulative shocks over t periods are given by

$$\sum_{s=1}^{t} \rho^{t-s}\sqrt{1-\rho^2}\eta_{ns} \sim N\left(0, (1-\rho^{2t})\Sigma\right) \quad (15)$$

[53] A function $f : [0,\infty) \to [0,\infty)$ is completely monotone if it is smooth and $(-1)^k f^{(k)}(x) \geq 0$ for all $x > 0$ and $k \in \mathbb{N}$.

The Schoenberg Interpolation Theorem states that if f is completely monotone and not constant, then for any distinct points $x_1, ...x_n$ in any real inner product space the $n \times n$ matrix A defined by $A_{ij} = f(||x_i - x_j||^2)$ is positive definite.

So for a given u_i, u_j, ϵ_{in0}, and ϵ_{jn0}, the probability that the person lives in j in t is

$$1 - \Phi\left(\frac{u_i - u_j + \rho^t(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}\right) \tag{16}$$

where Φ is the standard normal cumulative density function.

Define $F_{ij}(\epsilon)$ to be the mass of people that have $\epsilon_{in0} - \epsilon_{jn0} = \epsilon$. Note that this is continuous in ϵ and bounded because the ϵ_{int} have a multivariate normal distribution. Then the total number of migrants over t years is

$$\int_{u_j-u_i}^{\infty} F_{ij}(\epsilon)\left(1 - \Phi\left(\frac{u_i - u_j + \rho^t\epsilon}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}\right)\right) d\epsilon \tag{17}$$

or with a simple u substitution,

$$\frac{1}{\rho^t}\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{\frac{(1-\rho^t)(u_j-u_i)}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho^t}(w\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}} - (u_j-u_i))\right)\Phi(w)\,dw \tag{18}$$

Dividing by when $t = 1$, this quantity is

$$\frac{1}{\rho^{t-1}}\sqrt{\frac{1-\rho^{2t}}{1-\rho^2}}\frac{\int_{-\infty}^{\frac{(1-\rho^t)(u_j-u_i)}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho^t}(w\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}} - (u_j-u_i))\right)\Phi(w)\,dw}{\int_{-\infty}^{\frac{(1-\rho)(u_j-u_i)}{\sqrt{1-\rho^{2}}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^{2}}\sqrt{2-2\Sigma_{ij}} - (u_j-u_i))\right)\Phi(w)\,dw} \tag{19}$$

To take the limit as $\rho \to 1$, we can evaluate various terms separately:

$$\begin{aligned} \lim_{\rho\to1}\frac{1}{\rho^{t-1}} &= 1 \\ \lim_{\rho\to1}\sqrt{\frac{1-\rho^{2t}}{1-\rho^2}} &= \sqrt{t} \\ \lim_{\rho\to1}\frac{(1-\rho^t)(u_j-u_i)}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}} &= 0 \\ \lim_{\rho\to1}\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}} - (u_j-u_i)) &= u_j - u_i \end{aligned}$$

where the second line is by applying L'Hôpital's rule to the interior of the square root, and the third line can be obtained similarly, by moving the numerator into the square root, and applying L'Hôpital's rule. So the whole limit is

$$\sqrt{t}\frac{F_{ij}(u_j - u_i)\int_{-\infty}^{0}\Phi(w)dw}{F_{ij}(u_j - u_i)\int_{-\infty}^{0}\Phi(w)dw} = \sqrt{t} \tag{20}$$

The F term can be pulled out because F_{ij} is continuous and bounded and $\lim_{w\to-\infty}\Phi(w) = 0$. Therefore, the ratio of migration over t periods to migration over 1 period converges to $\sqrt{t}$. □

A.4 Proof of Proposition 3

From equation (18) in the proof of Proposition 2,

$$m_{ij} = \frac{1}{\rho}\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{\frac{(1-\rho)(u_j-u_i)}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}} - (u_j - u_i))\right)\Phi\left(w\right)dw \tag{21}$$

Taking the limit as $\rho \to 1$ on both sides,

$$\lim_{\rho\to 1} m_{ij}\frac{1}{\sqrt{1-\rho^2}} = \rho\sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{\frac{(1-\rho)(u_j-u_i)}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}} - (u_j - u_i))\right)\Phi\left(w\right)dw$$

As in the Proof of Proposition 2, the integrand goes to zero, and the term inside of F_{ij} goes to $v_{i0} - v_{j0}$.

$$\lim_{\rho\to 1} m_{ij}\frac{1}{\sqrt{1-\rho^2}} = \sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{0} F_{ij}\left(u_i - u_j\right)\Phi\left(w\right)dw$$

By the definition of F, the change in population is

$$\frac{\partial p_i}{\partial v_j} = -F_{ij}(u_j - u_i) \tag{22}$$

which can be pulled out of the integral. So

$$\lim_{\rho\to 1} m_{ij}\frac{1}{\sqrt{1-\rho^2}} = -\sqrt{2-2\Sigma_{ij}}\frac{\partial p_i}{\partial v_j}\int_{-\infty}^{0}\Phi\left(w\right)dw$$

Simplifying,

$$\frac{\partial p_i}{\partial v_j} = -\lim_{\rho \to 1} \frac{m_{ij}}{\sqrt{1-\Sigma_{ij}}} \sqrt{\frac{\pi}{1-\rho^2}}$$

□

A.5 Proof of Proposition 4

The population of i at time t is given by:

$$p_{it} = \sum_j p_{j,t-1} \frac{\exp(v_{it} - \delta_{ij})}{\sum_k \exp(v_{kt} - \delta_{kj})}$$

Taking the partial derivative with respect to v_j where $j \neq i$ gives us:

$$\frac{\partial p_{it}}{\partial v_{\ell t}} = -\sum_j p_{j,t-1} \frac{\exp(v_{it} - \delta_{ij}) \exp(v_{\ell t} - \delta_{\ell j})}{(\sum_k \exp(v_{kt} - \delta_{kj}))^2}$$

$$\frac{\partial \log p_{it}}{\partial v_{\ell t}} = -\frac{1}{p_{it}} \sum_j p_{j,t-1} \frac{\exp(v_{it} - \delta_{ij}) \exp(v_{\ell t} - \delta_{\ell j})}{(\sum_k \exp(v_{kt} - \delta_{kj}))^2}$$

Recall that $\delta_{ij} = \delta'_{ij} + \Delta$ if $i \neq j$. Plugging that in,

$$\begin{aligned}\frac{\partial \log p_{it}}{\partial v_{\ell t}} = -\frac{1}{p_{it}} \Bigg(& \sum_{j \neq i,\ell} p_{j,t-1} \frac{\exp(v_{it} - \delta'_{ij} - \Delta) \exp(v_{\ell t} - \delta_{\ell j} - \Delta)}{(\exp(v_{jt}) + \sum_{k \neq j} \exp(v_{kt} - \delta'_{kj} - \Delta))^2} \\ & + p_{i,t-1} \frac{\exp(v_{it}) \exp(v_{\ell t} - \delta'_{\ell i} - \Delta)}{(\exp(v_{it}) + \sum_{k \neq i} \exp(v_{kt} - \delta'_{ki} - \Delta))^2} \\ & + p_{\ell,t-1} \frac{\exp(v_{it} - \delta'_{i\ell} - \Delta) \exp(v_{\ell t})}{(\exp(v_{\ell t}) + \sum_{k \neq \ell} \exp(v_{kt} - \delta'_{\ell k} - \Delta))^2} \Bigg)\end{aligned}$$

Multiply both sides by $\exp(\Delta)$, and take the limit as $\Delta \to \infty$:

$$\lim_{\Delta \to \infty} \frac{\partial \log p_{it}}{\partial v_{\ell t}} \exp(\Delta) = -\frac{1}{p_{it}} \left(p_{i,t-1} \frac{\exp(v_{\ell t} - \delta'_{\ell i})}{\exp(v_{it})} + p_{\ell,t-1} \frac{\exp(v_{it} - \delta'_{i\ell})}{\exp(v_{\ell t})} \right)$$

But the terms on the right are the migration rates times $\exp(\Delta)$ when $\Delta \to \infty$. So that means:

$$\lim_{\Delta \to \infty} \frac{\partial \log p_{it}}{\partial v_{\ell t}} \exp(\Delta) = \lim_{\Delta \to \infty} \left(\frac{m_{i \to \ell}}{p_{it}} + \frac{m_{\ell \to i}}{p_{it}} \right) \exp(\Delta)$$

$\square$

A.6 Proof of Proposition 5

The proof of this proposition is identical to the proof of Proposition 3 because, in steady-state, the ϵ_{int}'s in the SPACE model have a stationary distribution.

A.7 Proof of Proposition 6

When $\Delta \to \infty$, the migration rate is given by:

$$\frac{m_{i \to j}}{p_i} = \exp(v_j - v_i - \delta_{ij})$$

So

$$\frac{\partial \log m_{i \to k}}{\partial v_k} - \frac{\partial \log p_i}{v_k} = 1$$

and

$$\frac{\partial \log m_{k \to i}}{\partial v_k} - \frac{\partial \log p_k}{v_k} = -1$$

In any steady-state, it must be the case that total inmigration and total outmigration are equal, which means that

$$\sum_{i \neq k} \left(\frac{\partial m_{i \to k}}{\partial v_k} - \frac{\partial m_{k \to i}}{\partial v_k} \right) = 0$$

For all $k \neq j$, the change in $\partial \log p_j / \partial v_k$ must be the same. This is because the migration rate between i and j does not change with v_k when $\Delta \to \infty$. So if their relative populations change, then the migration levels will change and we will not be in steady-state.

Substituting in,

$$\left(\sum_{i \neq k} m_{i \to k} \right) \left(\frac{\partial p_i}{\partial v_k} + 1 - \left(\frac{\partial p_k}{\partial v_k} - 1 \right) \right) = 0$$

where we can pull out the $\partial p_i / \partial v_k$ since it is the same for all i. This simplifies to:

$$\frac{\partial \log p_k}{\partial v_k} - \frac{\partial \log p_j}{\partial v_k} = 2$$

The total population is fixed:

$$p_k \frac{\partial \log p_k}{\partial v_k} + (1 - p_k) \frac{\partial \log p_j}{\partial v_k} = 0$$

So we need to solve for the population elasticities with two equations and two unknowns. The solution is:

$$\frac{\partial \log p_k}{\partial v_k} = 2 - 2p_k$$
$$\frac{\partial \log p_j}{\partial v_k} = -2p_k$$

□

A.8 Proof of Proposition 7

Note that this proposition is in Appendix C. However, we still include its proof here so that it is with the other proofs.

As in the proof of proposition 2, for a given utility difference $\epsilon_{in0} - \epsilon_{jn0}$, the probability that person moves to j in $t = 1$ is

$$1 - \Phi\left(\frac{u_{i0} - u_{j0} + \rho(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right) \tag{23}$$

where Φ is the standard normal cumulative density function.

If u_{j0} changes to u'_{j1}, then

$$1 - \Phi\left(\frac{u_{i0} - u'_{j1} + \rho(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right) \tag{24}$$

Define $F_{ij}(\epsilon)$ to be the mass of people that have $\epsilon_{in0} - \epsilon_{jn0} = \epsilon$. Note that this is continuous and bounded. Then the total migration from i to j is

$$\int_{v_{j0} - v_{i0}}^{\infty} F_{ij}(\epsilon) \left(1 - \Phi\left(\frac{u_{i0} - u'_{j1} + \rho\epsilon}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right)\right) d\epsilon \tag{25}$$

Taking a derivative with respect to u'_{j1},

$$\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{1}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}} \int_{u_{j0}-u_{i0}}^{\infty} F_{ij}(\epsilon)\phi\left(\frac{u_{i0}-u'_{j1}+\rho\epsilon}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}\right) d\epsilon \tag{26}$$

Evaluating it at $u'_{j1} = u_{j0}$,

$$\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{1}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}} \int_{u_{j0}-u_{i0}}^{\infty} F_{ij}(\epsilon)\phi\left(\frac{u_{i0}-u_{j0}+\rho\epsilon}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}\right) d\epsilon \tag{27}$$

Making a u-substitution,

$$\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{1}{\rho} \int_{-\infty}^{\frac{(1-\rho)(u_{j0}-u_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(u_{j0}-u_{i0}))\right)\phi\left(w\right)dw \tag{28}$$

And dividing by the overall migration (with the same u-substitution),

$$\frac{1}{m_{ij}}\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{1}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}} \frac{\int_{-\infty}^{\frac{(1-\rho)(u_{j0}-u_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(u_{j0}-u_{i0}))\right)\phi\left(w\right)dw}{\int_{-\infty}^{\frac{(1-\rho)(u_{j0}-u_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(u_{j0}-u_{i0}))\right)\Phi\left(w\right)dw} \tag{29}$$

Taking the limit as $\rho \to 1$,

$$\lim_{\rho\to 1}\sqrt{1-\rho^2}\frac{1}{m_{ij}}\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{1}{\sqrt{2-2\Sigma_{ij}}}\frac{F_{ij}\left(u_{i0}-u_{j0}\right)\int_{-\infty}^{0}\phi\left(w\right)dw}{F_{ij}\left(u_{i0}-u_{j0}\right)\int_{-\infty}^{0}\Phi\left(w\right)dw} \tag{30}$$

Simplifying,

$$\lim_{\rho\to 1}\sqrt{1-\rho^2}\frac{1}{m_{ij}}\frac{\partial m_{ij}}{\partial u'_{j1}} = \frac{\sqrt{\pi}}{\sqrt{1-\Sigma_{ij}}} \tag{31}$$

□

B Parametrization Details

In this appendix, we go over the procedure used to parametrize the model. As explained in the main text, the goal is to match the distance coefficient in a standard gravity equation, the population of each state, and the overall migration rate. In the baseline model, the parameters are a, which governs the spatial correlation of preferences; ρ which governs the persistence of preferences; and the u_i's which govern the desirability of each location.

Here is an overview of the procedure, with details below.

1. Guess an a.

2. Given that a, find the u_i that generate the populations in period 1.

3. Given that a and the u_i, find ρ that matches the migration rate.

4. Run a gravity regression and update the guess of a, and go back to step 2.

For step 2, we rely on an approximation in the proof of Proposition 3.

$$\frac{\partial p_i}{\partial u_j} \approx \frac{1}{\sqrt{1-\rho^2}} \frac{\sqrt{\pi}}{\sqrt{1-\Sigma_{ij}}} m_{ij}$$

We start with an initial guess for the vector of u_i's and use any number close to 1 for ρ,[54] simulate two periods of the model with 10 million people,[55] add up the m_{ij}'s in the simulation, and then calculate the approximate cross-partials according to this formula. With the cross-partials, we can calculate the change in the u_i's needed to hit the target population if their relationship was linear. Of course, the matrix of cross-partials is not invertible since the v_i's are only meaningful relative to one another. So we normalize $u_1 = 0$. We update the other u_i's based on inverting an $I-1$ by $I-1$ matrix of the $\partial p_i / \partial u_i$ times the vector of how far the populations in the simulation were from the populations in the data. We repeat this until each state's population is within one-tenth of one percent of the data, which typically takes a couple of iterations.

[54] In practice, we use zeros for u_i in the first two iterations, and for future iterations, we use the previous iteration's solution for u_i.

[55] To do a simulation, we simply draw 10 million random multivariate normals, with covariance Σ, add the u_i's, find the maximum for each, then simulate the second period by drawing a new 10 million random numbers, and adding them together based on ρ. Migration is calculated by counting the number of the draws that ended up with a person living in i in period 1 and j in period 2.

For step 3, we use the fact that migration is approximately proportional to $\sqrt{1-\rho^2}$. So we guess a ρ, simulate the model, find the migration rate in the simulation, and then scale ρ in order to hit the true migration rate.[56] We iterate until the migration rates is within one-hundredth of a percent. This typically takes 1 or 2 iterations.

For step 4, we use a bisection procedure. We store a "too-high" guess for a which generates too little of a relationship between distance and migration, and a "too-low" guess for a, which generates too strong of a relationship. Our next guess is the geometric mean of the two guesses, and depending on the gravity coefficient at the end, we replace either our "too-high" or "too-low" guess with the previous guess. We repeat this loop until we match the gravity distance coefficient to four digits.

[56]The formula for this is $\rho_{\text{new}} = \sqrt{1 - (1 - \rho_{\text{old}}^2)\frac{m_{\text{data}}}{m_{\text{simulation}}}^2}$.

C Comparing Migration Elasticities Using Hurricane Katrina

One related feature to Proposition 3—which gave an expression for population cross-elasticities—is that migration elasticities are decreasing in distance. To formalize this, we propose the following proposition.

Proposition 7. *In steady-state, as $\rho \to 1$, the short-run local semi-elasticity of migration from i to j to utility in i is*

$$\lim_{\rho \to 1} \sqrt{1-\rho^2} \frac{\partial \log m_{i \to j,t}}{\partial u_{j,t}} = \frac{\sqrt{\pi}}{\sqrt{1-\Sigma_{ij}}} \tag{32}$$

Recall, that $\Sigma_{ij} = \exp(-a \text{ distance}_{ij})$. So the magnitude of elasticity is decreasing in distance. The difference in elasticities is sizable but not enormous. A pair of states 220 kilometers apart (roughly the first percentile) have an elasticity that is 4 times greater than a pair of states infinitely far away from each other. The 75th percentile of elasticity across all state pairs is about 50 percent larger than the 25th percentile.

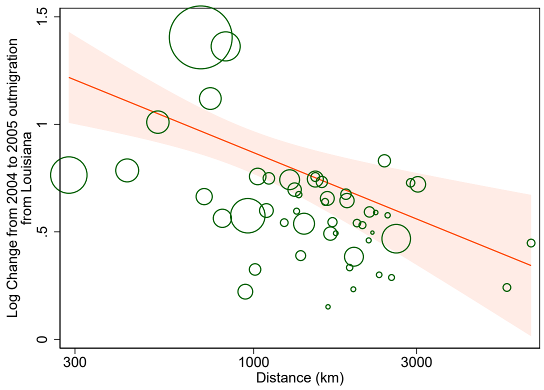
This contrasts to the baseline moving cost model, in which migration elasticities are governed by a single parameter and do not vary with distance.

In principle, this is a testable implication. If there were an exogenous shock to a specific state that had no equilibrium impact on the utility of nearby states, we could see if migration increased more between nearby states than far away states. However, it is not easy to test the elasticities because many changes in location utility are also correlated across space, either because of correlated underlying shocks or because of equilibrium effects that come through migration or trade.

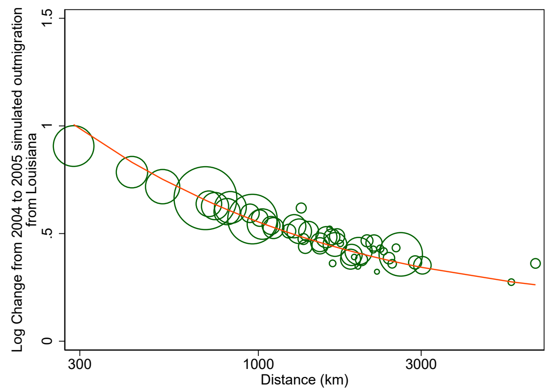
The best we can do is explore the Hurricane Katrina shock, which is spatially concentrated in Louisiana and large compared to its potential general equilibrium effects. Arguably, the change in the utility of living in Louisiana after the hurricane was much larger than the spillover effects to nearby states.[57]

We show the log-change in the amount of migration from Louisiana to other states after the hurricane in Figure A1. The left subfigure is the observed data, and the right

[57]We considered other shocks as well, but could not identify large shocks that affected migration significantly in one specific state without spillovers to nearby states. For example, the inmigration to North Dakota during the fracking boom coincided with significant inmigration to South Dakota, and to a less extent, other nearby states as well. Oddly, the abrupt change in migration in Louisiana in 2005 was only on the outmigration side; inmigration responded comparatively little.



(a) Data, with line of best fit



(b) Simulated data from the SPACE model, with theoretical approximation line

Figure A1: The log change in outmigration from Louisiana to the other 50 states after Hurricane Katrina. Panel (a) shows the log change in outmigration from Louisiana from 2004-2005 to 2005-2006. Size of dots is proportional to 2004-2005 outmigration. Distance is on a log scale. A line of best fit with 95 percent confidence interval is also shown. Data is from the IRS Migration Data. Panel (b) shows the same, but from a simulation of the SPACE model. The orange line is the theoretical approximation that the outmigration elasticity should be proportional to $1/\sqrt{1-\Sigma_{ij}}$.

is the model simulation with a theoretical approximation line. For this exercise, we use the IRS migration data, since the absolute amount of migration between some state pairs is quite small. The log-change in the observed data was larger in closer states, as predicted by the SPACE model. It is well-publicized that the government aided in moving some displaced migrants to Texas (the top point), but the trend is more general than that one point. In addition, the hurricane also hit parts of Mississippi which may explain that outmigration to Mississippi (the furthest left point) was not as high as might be predicted.[58]

In the simulated subfigure, we see that the approximation given by theoretical elasticities $1/\sqrt{1-\Sigma_{ij}}$ is not perfect, but the approximation error is small. To cut down on simulation error, we increase the number of simulated agents to 300 million. Any systemic deviations from the theoretical prediction line are likely due to the fact that ρ is a tiny bit less than one in our calibration.

[58]Of course, one could think of other explanations for the migration elasticity to vary, based on the public policies affecting the displaced residents or the demographics of the neighborhoods that were affected. We are not claiming that this evidence alone is enough to prefer the SPACE model, but we do view the data as consistent with the model's predictions.

Even though the baseline moving cost model would not predict migration elasticities that vary with distance, one could modify the logit model to match this fact, for example with a mixed logit (Train, 2009). The point of this exercise is not to falsify the whole class of moving cost models. Rather, the point of this exercise is to show that the additional $(1 - \Sigma_{ij})^{-1/2}$ term that shows up in Proposition 3 is not a reason to prefer the moving cost model to the SPACE model.

D Extensions of the SPACE model

D.1 Commuting Zones

In this appendix, we demonstrate that the parameters estimated in Section 3.1 work reasonably well when we use commuting zones instead of states. While another approach would have been to recalibrate the entire model for commuting zones, we would rather highlight the fact that the parameterization of ρ and a are more general than the specific setting in which they were calibrated, and can therefore be extended to other settings.

We focus on the 683 commuting zones for which the IRS records some migration in or out of the commuting zone.

There are two concerns with this exercise. The first is that we do not have a complete tabulation of commuting zone-to-commuting zone migration. The IRS data provides county-to-county migration, which can be aggregated to commuting zones, but they censor the data at 10 returns, so migration between many commuting zones is not recorded, or is biased.

The second concern is that we have to recalibrate the v_i's in order to estimate a gravity equation on simulated data.

To address the first concern, we use additional information in the IRS data to infer missing county-to-county migration. The IRS reports total domestic inmigration and total domestic outmigration from each county. Our methodology is as follows:

1. Run a truncated poisson regression of migration (measured by returns) on log distance and log origin and destination populations. The truncation parameter is 9, since all returns below 10 are censored.

2. Use the results of the regression to predict bilateral county migration for all missing values.

3. For origin counties, add up all predicted migration, and compare to the difference between measured migration in the bilateral county migration and the total reported migration by county. Adjust predicted values proportionally to match.

4. Repeat step 3, but for destination counties.

5. Repeat steps 3 and 4 five times (at which point they have reached a fixed point).
6. Replace any predicted migration that is above 10 with 10.
7. Repeat steps 3, 4, and 6 ten times (at which point they have again reached a fixed point).

This procedure guarantees that the inferred migrations are less than or equal to 10, and is much more likely between nearby and large counties.

We then aggregate migration to commuting zones, using both the data and the predictions that come from this procedure.

To address the second concern, we rerun the part of our algorithm to calibrate the u_i using commuting zones instead of states. We do not recalibrate either ρ or a.

	(1) Migrants (IRS data)	(2) Simulated Migrants
Log Distance	-1.085***	-0.997***
	(0.0495)	(0.0448)
Log Destination Population	0.832***	0.864***
	(0.0369)	(0.0266)
Log Origin Population	0.869***	0.860***
	(0.0338)	(0.0272)
Observations	465806	465806

Standard errors two-way clustered by origin and destination commuting zones.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A1: Gravity equation, commuting zones

At this point, we compare the data (including the predicted data from the procedure above), and the simulated data from the model. We focus on two dimensions: the migration rate and the gravity equation. The commuting zone 1-year migration rate in the data is 4.39 percent. In the simulation, it is 4.77 percent. Recall that the simulation is parametrized based on Gies data, whereas the data is based on IRS data. At the state level, the Gies data migration rate is about 10 percent higher than the IRS data, which can account for the difference between the simulation for commuting zones and the data.

The estimated gravity equations are presented in Table A1. The estimated coefficient on distance, in both the data and the simulated data is a lot higher than it was for states, and are fairly similar to one another (the same coefficient was -0.74 for states). Obviously it is not exact, but it reflects the fact that the a which was calibrated on states can still be in the right ballpark for commuting zones.

D.2 Education as a distance between states

There are other determinants of migration flows besides distance and population. In our baseline model, we only try to match the gravity coefficient on distance, but in this appendix, we demonstrate the versatility of the model.

One fact about migration is that people move more between similarly-educated regions. One reason for this is that people may have correlated preferences over the amenities that are present in educated areas.[59]

In fact, if we run a gravity equation with one extra term, the "distance" in terms of education matters significantly. We run the following Poisson regression:

$$\log m_{i\to j} = \alpha \log p_i + \gamma \log p_j + \beta \log \text{distance}_{ij} - \delta |\text{Bachelor's share}_i - \text{Bachelor's share}_j| + \epsilon_{ij} \tag{33}$$

where Bachelor's share$_i$ is the share for the 25 and older with a bachelor's degree or higher in the 2000 Census (Manson, Schroeder, Riper, Kugler and Ruggles, 2021). The results of this regression are in column (1) of Table A2. The coefficient on the absolute difference in the bachelor's share is negative and statistically significant.

	(1) Migration (Credit)	(2) Simulated Migration (with Education)	(3) Simulated Migration
Log Distance	-0.741***	-0.741***	-0.741***
	(0.0503)	(0.0382)	(0.0388)
Abs. Diff. in Bachelor's Deg. Share'	-2.485**	-2.485***	-2.375***
	(0.895)	(0.506)	(0.533)
Log Destination Population	0.879***	0.879***	0.876***
	(0.0794)	(0.0491)	(0.0496)
Log Origin Population	0.909***	0.878***	0.879***
	(0.0796)	(0.0479)	(0.0481)
Observations	2550	2550	2550
Pseudo R^2	0.722	0.907	0.906

Standard Errors are two-way clustered by origin and destination states
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: An extension to the gravity equation

Can our model capture this determinant of migration? The challenge is to make locations with similar bachelor's degree shares more correlated while continuing to ensure that the covariance matrix is positive definite. By proposition 1, we know that

[59] Another possibility is that the high-skill and low-skill workers have different preferences over where to live. This is totally plausible, but it is straightforward to incorporate into the model by explicitly modeling the heterogeneity, so there is no need to do the exercise we carry out in this section.

as long as we are measuring a distance in $\mathbb{R}^n$, then the matrix is still positive definite. The solution is to make the education level another dimension:

$$\text{distance}_{ij} = \sqrt{\text{physical distance}_{ij}^2 + b(\text{education level}_i - \text{education level}_j)^2} \qquad (34)$$

where b is another parameter to estimate. We will aim to also match δ from the Poisson regression coefficient, in addition to the gross migration rate and β, which we were already aiming to match.

The parameters we calibrate are $\rho = .999623$, $a = -.025044$ and $\sqrt{b} = 908.4$.[60] We show $\sqrt{b}$ because it has a geometric interpretation: it is like imagining that for every 1 percent higher share of bachelor's degrees, the state's location is displaced 9.04 kilometers vertically.

We show the Poisson regression of the model including b in column (2) of Table A2. We are able to match the targeted coefficients to three decimal places. The population coefficients, which are not targeted, are again fairly close.

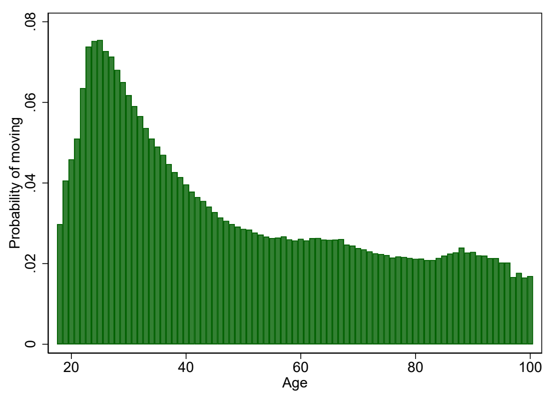
Interestingly, in column (3), the original parametrization from Table 1, where $b = 0$, also does a fairly good job of matching the education coefficient. Likely, this is because much of what the education coefficient is capturing is the non-linear effects of distance (e.g. excess movement between California and Massachusetts). And we already know from Figure A3 that the baseline model can capture that non-linearity.

[60] In the code, we calibrate a^2/b^2 for computational purposes, and then calculate $\sqrt{b}$ from that.

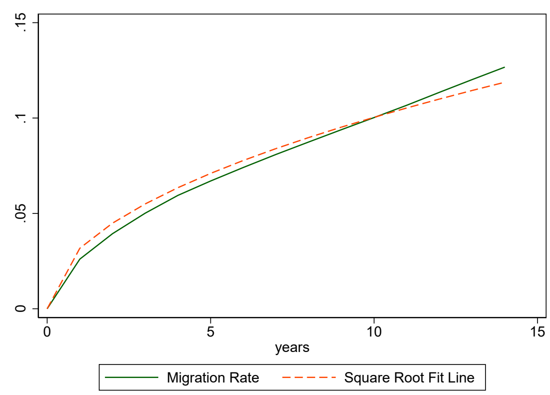
E Appendix Tables and Figures

E.1 Age and the Square Root Fact

A reader might worry that migration in the GCCP is not measured well for young people. For example, college students may not be measured accurately. In the GCCP, the data provider includes a proxy for age, which we use to exclude people that are young. As you can see in Figure A2a, younger people are much more likely to move, and so it is a reasonable concern that the square root fact may be specific to young people. However, in Figure A2b, we can exclude people below the age of 45, and the square root fact still holds.



(a) Moving probability by age in the GCCP



(b) Square root fact for people over 45 in the GCCP

Figure A2: Robustness to considering whether age explains the square root fact

E.2 The Non-Linear Relationship between Migration and Distance

Figure A3 shows the non-linear relationship between the distance and the migration between two states, as described in Section 3.1.

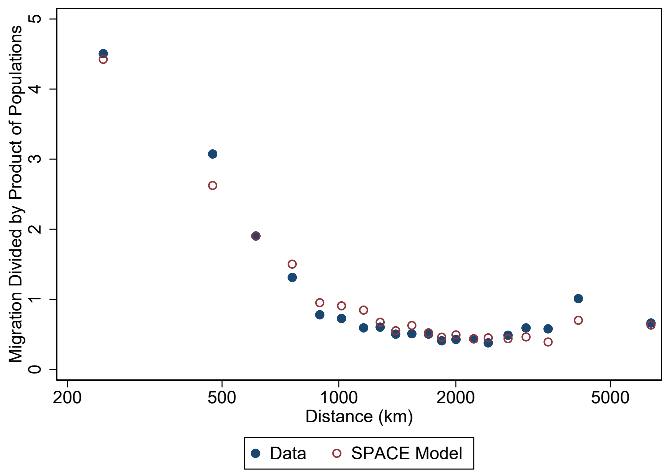


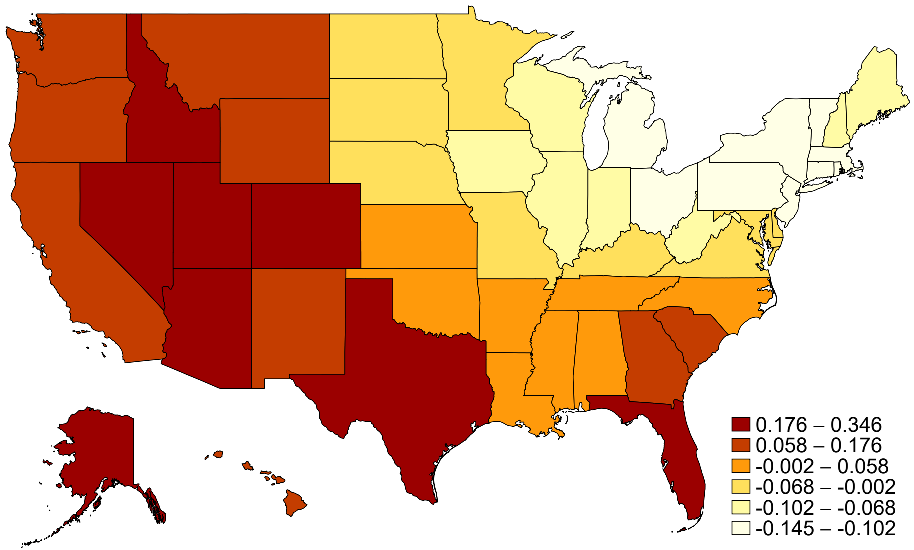
Figure A3: Migration and distance. For each state-pair, the value on the y-axis is calculated by dividing the migration between the two states by the product of the two states' populations. We then normalize that value so that the average across all state pairs is 1. The value on the x-axis is the distance between the two states and is presented on a log scale. Each point represents a bin of about 128 state-pairs, sorted by distance, with the average value of distance and population plotted. The blue dots are the data in 2004-2005, and the hollow red dots are the simulated data based on the calibration in the text.

E.3 The Change in States' Utility, 1980-2019

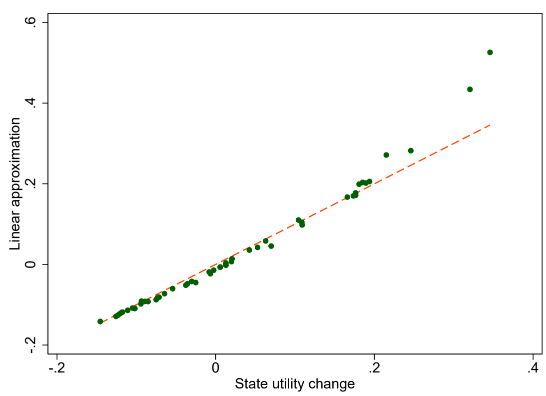
Figure A4 explores how accurate the approximation in Propositions 3 and 5 is. We calibrate the u_i in the SPACE model to match population shares in 1980 and in 2019. We do not recalibrate a or ρ. The changes in u_i are shown on the map in Panel (a).

In Panel (b), we compare the calibrated changes to the changes that we would have calculated based on the elasticities in Proposition 3. To calculate these changes, we construct the full matrix of population cross-elasticities. Then we delete the row and column corresponding to one state (the choice is arbitrary, since we can only calculate identify relative utility changes, not absolute changes). Then we can invert that matrix and multiply it by a vector of population changes in order to see what this proposition would imply for the utility changes of each state. We plot those against the non-linearly calibrated changes in Panel (b). The fit is quite good except for some of the states that experienced a lot of population growth.

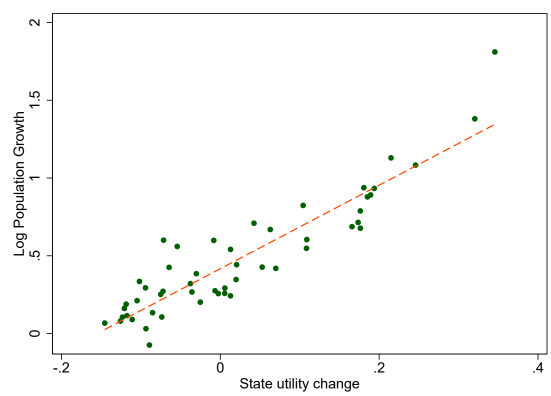
In Panel (c), we compare the calibrated changes to population growth of the state, to show that there are meaningful deviations. In a static logit model, we would have a perfectly linear relationship between log population change and the change in utility. However, in the SPACE model, we expect deviations from a perfect relationship because the elasticities of population to own utility depend on gross migration rates and because the changes in other states' utility affect population differentially across space.



(a) Utility changes as implied by the model



(b) Utility changes as implied by the model and the first-order approximation



(c) Utility changes implied by the model versus population growth

Figure A4: Utility Changes from 1980-2019. Panel (a) shows a map of utility changes, by state. Dark red states had the largest relative increases in utility, and the white states had the smallest relative increases. In Panel (b), each point is a state, and the dashed line is a 45 degree line. The x-axis is the change in utility from simulations of the model that match the population shares in 1980 and 2019, and the y-axis is the first-order approximation from equation (8). The state with the largest utility increase is Nevada, followed by Arizona.

E.4 Long-run population elasticities in the moving cost model

In this appendix, we compare the elasticities in the calibrated moving cost model to the approximation from Proposition 6, which stated that the long-run population elasticities were proportional to those of a standard static logit model.

Calculating these elasticities for the moving cost model is not straightforward analytically, but easy to do numerically. For each state, we change the utility by a small amount, calculate the new migration matrix from equation (3), and simulate 500 periods to see how much population in each state changes.

The overall correlation between the long-run elasticities of the moving cost model and the static logit is 0.99996. Splitting it between same-state and cross-state elasticities, the correlation is 0.948 and 0.9994, respectively. Hence, it is a reasonable approximation to say that the moving cost model approaches a static logit in the long-run.

We show a plot of the cross-state elasticities in Figure A5. The relationship is quite strong.

Recall that, in contrast, the SPACE model does not approach a static logit model, but has a much richer set of cross-elasticities, given by Proposition 3.

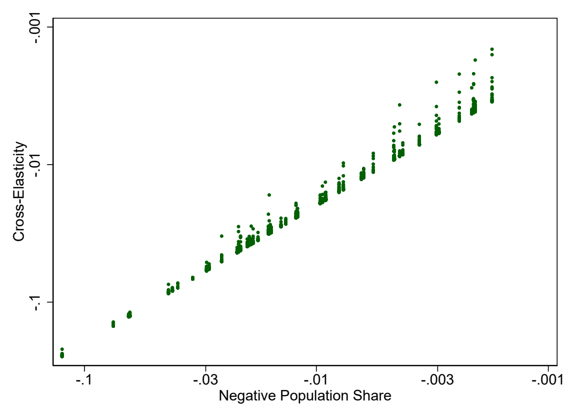


Figure A5: Population cross-elasticities in simulations of the moving cost model versus the theoretical approximation (in a static logit, the cross-elasticity is proportional to the population share)