Internal Migration and the Microfoundations of Gravity*

Greg Howard[†] Hansen Shao[‡]

September 28, 2022

Abstract

We propose a new model of internal migration based on persistent and spatially-correlated location preferences that can match three main facts: migration is rare, it follows a gravity pattern, and a disproportionate share is return migration. We refine this third fact by showing that the t-year migration rate is proportional to $\sqrt{t}$, which our model also matches. We derive tractable expressions for population elasticities and show these are important for a variety of central questions in the literature. Our model has tractability advantages over the typical dynamic logit and starkly different implications for regional dynamics, misallocation, and long-run population adjustments.

Keywords: regional evolution, migratory insurance, misallocation, gravity equation, labor mobility

JEL Codes: R23, R13, J61

*We would like to thank Javier Quintana, Kyle Mooney, Vivek Bhattacharya and participants at the Illinois Macro Lunch, Illinois Young Applied Faculty Lunch, the European UEA meetings, Midwest Macro, the AREUEA National Meetings, and the Stanford Institute for Theoretical Economics Housing & Urban Economics Conference 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. 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 the third most prominent fact about internal migration: a large share of internal migration is return migration, in which people are moving back to a place they previously lived. We refine this fact by showing that the t-year migration rate, the fraction of people who live in a different geography than they did t years ago, is proportional to the square root of t. This is a fact that comes naturally in models of persistent preferences but is a hard fact to match with moving cost models.

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. We show the implications of the model for aggregate dynamics, welfare, and counterfactuals. In particular, we show that migration between two regions is an important statistic for calculating the cross-elasticity of one region's population to the other region's utility. We argue that the model has several important advantages over the current workhorse model. Finally, we show that our model calls into question several central conclusions that have been drawn in the literature about internal migration.

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.[1] We chose a multinomial probit model in order to incorporate persistent preferences in a tractable way.[2]

The SPACE model is able to match the three main facts about migration, including the fact about return migration, which has been a challenge for the literature. We also show that the SPACE model, when calibrated to match the first two facts of internal migration, can also match a variety of 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.

Next, we show that in the SPACE model, gross migration levels provide important infor-

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

[2] We also contribute to the literature on multinomial probits in a geographic setting by showing a specification of the covariance matrix as a function of distance that guarantees positive-definiteness.

mation to calculate population elasticities to utility changes. When the preferences are very persistent, as in our calibration, then gross migration rates are almost a sufficient statistic for calculating these elasticities.[3] Regions with more gross migration, like D.C., have a more elastic population to the baseline utility in their own region. Regions with less gross migration, like California or Texas, have less elastic populations. In addition, the elasticity of population in one state with respect to utility in another state is higher if there is more migration between them. An increase in Florida's utility has a more negative effect on Georgia's population than it does on Arizona's.

These population elasticities are key to answering some of the major questions in the literature. These questions have focused on the causal effects of location specific shocks on population distributions, to which these population elasticities are key.

Another important question regards how quickly a region's population adjusts to a change in utility. In the SPACE model, utility changes immediately affect populations, meaning that migration by itself does not slow down adjustments.

A final application of the population elasticities is that we show that the spatial distribution of a shock is key to estimating how large is the insurance from migration. In general, migration provides insurance by allowing people to move away from relatively declining regions. For any insurance to occur, there must be a spatial component to the shock; there would be no gains if the shock is uniform across all regions. But importantly, if the places that are hit hardest by a negative shock are in one region (e.g. the Rust Belt), there is much less migration and therefore less insurance in response. If the places hit hardest by the shock are far away from each other, then there is more migration to places that are not affected, and therefore insurance. Conversely, for a positive shock, shocks to a whole region (e.g. the Sun Belt) have less of an effect due to a small migration response, and shocks that are in multiple far-away locations around the country will have a larger welfare effect because of a large migration response. This spatial distribution of shocks matters quantitatively. When we apply our model to the changes in the United States from 1970 to 2018, we find that the amount of migration insurance during this time period was only about half as large as in a counterfactual where the shocks had been spatially-allocated randomly.

Next, we discuss the SPACE model in comparison to the workhorse model used in the literature: the dynamic logit. We argue that when focused on aggregate outcomes, the SPACE model has tractability advantages over the standard model. This is important because one of the main reasons the dynamic logit is so popular is because of its easiness-to-use. One advantage of the SPACE model is that it requires zero state variables, as compared to the

[3]The reason we say "almost" sufficient is that these elasticities are also related to two parameters of the model and the distance between regions. Migration explains most of the variation, however.

dynamic logit which requires keeping track of last periods' populations. This advantage grows as the modeler tries matching more of the dynamic moments because the number of state variables in a dynamic logit quickly explodes. Since the persistent spatially-correlated model already matches dynamic moments, there are no additional state variables required. Similarly, if a modeler wishes to match the fact that people are more likely to move to the last state they lived in, or even the state before that, the dynamic logit requires keeping track of how many people lived in that sequence of states, meaning that the number of state variables grows with the square or the cube of the number of locations. Since the SPACE model also naturally matches that fact, no additional state variables are required. Finally, our model makes different predictions regarding migration elasticities than the dynamic logit. In our model, they decrease in distance, while the dynamic logit assumes they are constant. Using variation in migration induced by Hurricane Katrina, we check that our prediction holds in the data. While logits are flexible enough to accommodate all these features, they lose their tractability advantage, while the SPACE model does not have to be augmented to match all these moments.

Finally, we show that our model has different implications for quantitative conclusions that might be drawn if using the dynamic logit model. Perhaps trivially, the SPACE model implies that moving costs do not need to be large to match observed patterns of migration. This provides a different perspective on aggregate misallocation and policy approaches aimed at reducing misallocation. Second, the aggregate dynamics of the two models are quite different. In a dynamic logit model, permanent changes in a region's utility change the migration rate leading to very slow population adjustments. In the SPACE model, utility changes the population size, causing a temporary increase in migration and very quick population adjustments. Third, different moments of the data identify utility changes across space: the SPACE model focuses on changes in populations, while the dynamic logit is identified by looking at changes in migration rates. Finally, the SPACE model has very different implications for the long-run population elasticities across locations. We show numerically that the dynamic logit's long-run population elasticities are well-approximated by the static logit, which does not depend on gross migration or distance. In contrast, the level of gross migration is a key statistic in calculating population elasticities in both the short-run and the long-run in the SPACE model.

In order to clarify the contributions of the model, it is helpful to be specific regarding the difference between moving costs and persistent preferences. While mathematically, it is straightforward to specify them in a model, it is important to understand what each term means when we try to map it onto the real world. A typical dynamic logit model has a one-time irreversible cost borne by people who leave one area for another, which economists call a "moving cost." In contrast, persistent preferences across space mean that the decrease in utility when a person moves from one location to the other is both persistent over time and 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 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 argue 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

Migration is central to a wide variety of questions in the spatial literature. Since at least Blanchard and Katz (1992), it has been recognized as a key feature to how regions adjust to economic shocks.[4] 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; Liu, Klieman and Redding, 2021; 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). More generally, location choice has important aggregate and distributional consequences for economies (e.g. Hsieh and Moretti, 2019; Diamond, 2016, among many others). 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). And migration is known to be an important margin when analyzing housing markets in particular (Schubert, 2021; Howard and Liebersohn, 2021). Central to many of these questions is how elastic is the population of a region to various shocks, over various time horizons.

Corresponding to the growth of interesting questions related to migration, there has also been advances in the ways to model migration.[5] Kennan and Walker (2011) wrote down the

[4] Rosen (1979) and Roback (1982) developed the most widely used model of where people live, but with no dynamics, there is not a notion of migration per se.

[5] There is also a large literature on the microfoundations of the gravity equation for trade. See, for example, Bergstrand (1985), Helpman, Melitz and Rubinstein (2008), and Chaney (2018). There are several

canonical model of migration using the dynamic logit formulation. Kaplan and Schulhofer-Wohl (2017); Giannone, Li, Paixao and Pang (2020); Porcher (2020); Mangum and Coate (2019); Zerecero (2021); 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.[6] 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 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 Mangum and Coate (2019) and Zerecero (2021), by adding in a preference for living in one's birthplace. 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 dynamic logit. However, adding birthplace preferences to a dynamic logit model would not generate the square root fact, and many of the implications that I highlight are different for the SPACE model would still be there in a dynamic logit model with birthplace preference.

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; 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

reasons that the microfoundations of gravity for migration and for goods might be different. At the most basic level, the decision-makers for migration are what is being measured, whereas decision-makers for trade are importing or exporting, and it is the goods being measured. Migration also has a dynamic element to it, as people make repeated decisions on where to live.

Even within trade, Lind and Ramondo (2018) has emphasized the gains from thinking about spatial correlation in quantitative models.

[6] 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.

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 $\sqrt{t}$.

The industrial organization literature has emphasized the difficulty of distinguishing switching costs and persistent preferences for a long time (Heckman, 1981). Here, I argue that the new $\sqrt{t}$ fact is consistent with a simple model of persistent preferences.

Finally, because I wish to model persistent preferences, I use a discrete choice model with multinomial probit.[7] 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 the important population cross-elasticities.

1 Three facts about internal migration

We use two main sources of data: IRS migration data and credit data from one of the leading credit report providers (IRS Migration Data, 2004-2018; Gies Consumer and Small Business Credit Panel, 2004-2018).[8] The IRS data has aggregated data on state-to-state migration, useful for estimating a gravity equation, while 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. The two datasets have similar one-year gross migration rates, 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. The two datasets also exhibit similar gravity patterns, which we show later in Table 2.[9]

[7]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.

[8]For other papers using the Gies Consumer and Small Business Credit Panel, see Fonseca (2022), Fonseca and Wang (2022), and Han (2022).

[9]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

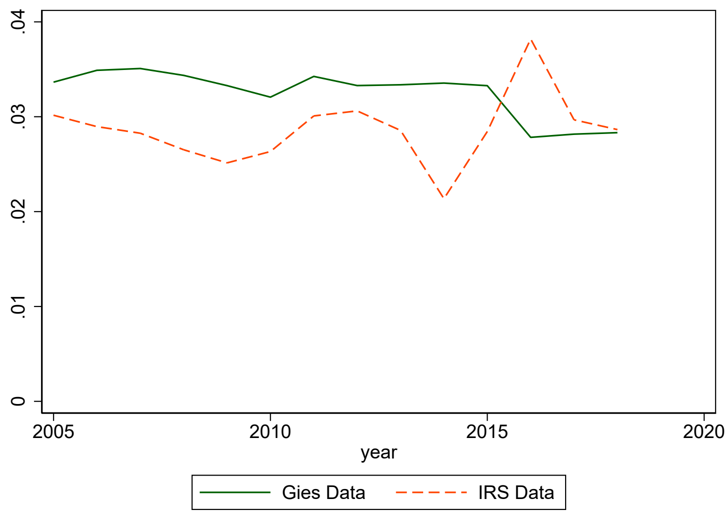


Figure 1: Comparison of interstate migration rates in IRS and GCCP data.

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).

Using this data, we present three facts about internal migration.

Fact 1: Migration is rare. The average interstate migration rate is less than 3 percent in the IRS data and just over 3 percent in the GCCP (see Figure 1).

Fact 2: Migration follows a gravity pattern. In both datasets, migrants move to nearby and more populous states. 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 migration from i to j. We find that both α and γ are positive and β is negative.

Fact 3: The t-year migration rate is proportional to $\sqrt{t}$. Our final fact relates to the 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,

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 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.

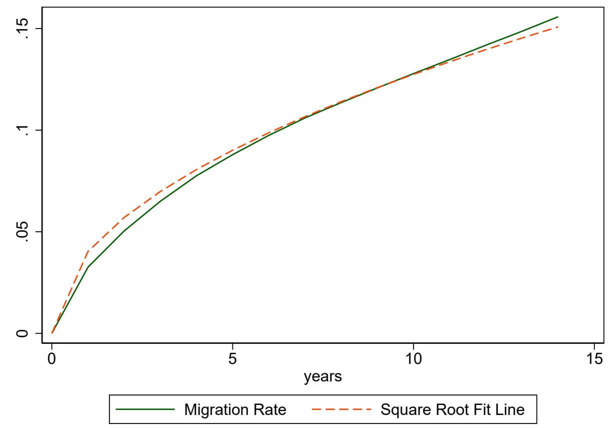


Figure 2: Migration Rates at Different Horizons. Migration rate at year t is calculated as the percentage of people living in a different state than they did t years ago. Data is from an unbalanced panel and uses any observations in which the state of residence is recorded t years apart. Source: GCCP.

2011; Kaplan and Schulhofer-Wohl, 2017). One consequence of this fact is that the two-year migration rate is significantly less than twice the one-year migration rate. In Figure 2, we show a new twist on this fact, that the t-year migration rate is approximately proportional to $\sqrt{t}$, up to 14 years, the length of our panel. 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.

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 Appendix A, we show that the square root fact is also present in data from the Panel Survey of Income Dynamics (1969-1997). And since mismeasurement in the GCCP may be a particular concern for young people, we also show that it holds for people over 45.

In Appendix A, we also show that the square root fact is not something easily explained by a Markov model of migration. We first consider Markov processes that depend on the state of residence, which leads to a very linear relationship between t and the t-year migration rate. Adding age to the state space does not change the linear relationship noticeably. And while adding the state of residence the previous year helps add a kink at $t = 1$, the relationship beyond $t = 2$ is still fairly linear, and does not do a good job generating the curvature of the square root. So while the square root fact is related to the well-known fact that there is a lot of return migration, it is not a simple consequence of that well-known fact.

2 The SPACE model

This section introduces a new model of internal migration which is designed to match the three moments from the previous section. Notationwise, n denotes the individual, i the location, and t the year. Individuals pick their location to maximize utility:[10]

$$u_{nt} = \max_i u_{nit} = \max_i v_{it} + \epsilon_{nit} \tag{2}$$

where v_{it} is a baseline utility for location i and ϵ_{nit} is a person-location match specific utility.[11] To denote it using vectors of size $I \times 1$,

$$u_{nt} = \max \vec{u}_{nt} = \max \vec{v}_t + \vec{\epsilon}_{nt} \tag{3}$$

where

$$\vec{\epsilon}_{nt} = \rho \vec{\epsilon}_{n,t-1} + \sqrt{1-\rho^2}\vec{\eta}_{nt} \qquad \text{and} \qquad \vec{\eta}_{nt} \sim N(0, \Sigma) \tag{4}$$

Note that ϵ_{nt} also has variance equal to the matrix 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:

$$\Sigma_{ij} = \exp(-a \text{ distance}_{ij}) \tag{5}$$

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

Recall that we named the model the "SPACE" model because its distinctive feature is its Spatially and Persistently Auto-Correlated Epsilons.

Proposition 1. *Σ is positive-definite.*

Proposition 1 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. Proofs are collected in Appendix B,

[10] Equation (2) assumes that agents are myopic. This is more restrictive than necessary, as the agents can costlessly move every period. In other words, the agent's location is not a state variable. However, it is restrictive as it rules out the possibility of accumulating capital, either through savings or acquiring skills. This is a common assumption in the literature (Caliendo et al., 2019; Liu et al., 2021), as keeping track of wealth distributions across many locations is computationally difficult. In sum, the model needs to impose that agents cannot save or invest, but the fact that agents are myopic is easily relaxed without changing any implications.

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

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).[12]

Migration occurs when the location j that maximizes utility u in time t and time $t-1$ are different.

2.1 Matching the facts

This model can match all three facts from Section 1.

Immobility. The model can match the fact that most people do not move because their preferences are persistent. As ρ gets closer to 1, the migration rate will decrease to zero.

Gravity. The model can also match the gravity pattern of migration. This feature is the product of both the persistence and the spatial-correlation of individual preferences. Why does having a high correlation increase the migration between two regions relative to other regions? The effect is subtle. It is not because the two regions have more people who prefer one in the past period and now prefer the other. For example, if there are only two regions with equal baseline utility, then the correlation does not matter: a fraction independent of the correlation will migrate between the two regions each period. With more regions, it is still true that the fraction for which i becomes more preferable to j is a constant fraction for any i and j regardless of the correlation. But in this case, the number of movers is also dependent on having a high value for both i and j, which is more likely to occur when i and j are highly correlated. As a concrete example, consider two locations with a joint normal distribution where the mean utility is zero and the variance is 1 for both locations, with correlation Σ_{ij}. Regardless of Σ_{ij}, the number of people for which $\epsilon_{it} > \epsilon_{jt}$ and $\epsilon_{it+1} < \epsilon_{jt+1}$ is determined only by ρ, the persistence of the preferences. However, the distribution of the values of ϵ_{it} for which switchers, who prefers i to j in period t, but prefers j to i in period $t+1$, as $\rho \to 1$ is distributed $N(0, \frac{1}{2} + \frac{1}{2}\Sigma_{ij})$. If Σ_{ij} is not large, then there are very few switchers in the far right-tail, meaning they probably did not live in i or j to begin with, and therefore do not migrate from i to j.[13]

[12]Complete monotonicity is a fairly restrictive requirement. However, there are other functions besides $\exp(-ax)$ that would work, such as $1/(1+ax)$.

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.

[13]Having a more correlated shock also reduces the population of the two locations. For example, if the correlation is close to 1, the two locations only get one effective draw to live in the two locations, whereas if they are uncorrelated, there are two chances to pick one of the locations. This also increases the migration rate, once it is normalized by population.

Return Migration. Intuitively, the reason this model exhibits significant return migration is that a person who has just moved from location i to location j is close to indifferent between the two locations, and is likely to move back to location i if they get a shock in the opposite direction of the previous shock.

Persistence of preferences is also important to match the square root fact. In fact, as ρ gets close to 1, we can show a theorem that the model will match our fact.

Proposition 2. *Holding the v_i's constant over time, as $\rho \to 1$, the t-year migration rate is proportional to $\sqrt{t}$, i.e.*

$$\lim_{\rho \to 1} \frac{m_{i \to j,t}}{m_{i \to j,1}} = \sqrt{t} \tag{6}$$

where $m_{i \to j,t}$ is the t-year migration rate.

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.

2.2 Calibration

So far, we have outlined theoretical reasons that the model can match the main three facts about migration, but it is also important to check that they can match the facts quantitatively.

Throughout the section, we use a calibration of the model calibrated to match the first two facts. We parametrize ρ, a, and v_i in order to match the probability of migration, the coefficient on distance in a gravity equation, 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.[14]

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 shock of more that 0.999.[15] 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

[14]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 C for details.

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

Table 1: Parameterization

Parameter	(1) Value	Moment	(2) Data	(3) Model
a	0.000299	Distance coefficient	-0.7376	-0.7376
ρ	0.999627	Average migration	0.0334	0.0334
v_i		Population of each state		

interpretable, but the value is .000299 meaning that correlation of shocks for two states 1000 kilometers apart is about 0.74. The v_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), Credit data (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 three different datasets even though we are not targeting these coefficients.[16]

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.

We also compare the relationship of distance and migration non-linearly. Figure 3 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.

Our calibration also matches a variety of dynamic moments. We start by looking at the

[16] One possible reason for the coefficients on population being less than one is that we model the personal preferences as normal. Normal distributions have an increasing hazard function, so the probability of being near a cutoff, conditional on being above that cutoff, increases as the cutoff gets bigger. In this context, people who live in low-utility states—and therefore must have very high draws of personal utility for that state—are more likely to be close to indifferent with another state.

A highly 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.

Table 2: Gravity Equations

	(1) Migration (IRS)	(2) Migration (Credit)	(3) Simulated Migration	(4) Migration (Credit)	(5) Simulated 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
R^2					
Pseudo R^2	0.725	0.719	0.903	0.847	0.949
Origin and Destination FEs				Yes	Yes

Standard Errors are two-way clustered by origin and destination states

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the moving probability at 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 that the sample changes slightly depending on t.[17] 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.[18]

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 important that the model is able to capture the more-commonly-examined moments as well. A natural moment is the distribution of the number of interstate moves over time.

Figure 4b 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

[17] 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.

[18] The figure also presents the same exercise for a simulation of the workhorse dynamic logit model, which we discuss later in Section 3.2.

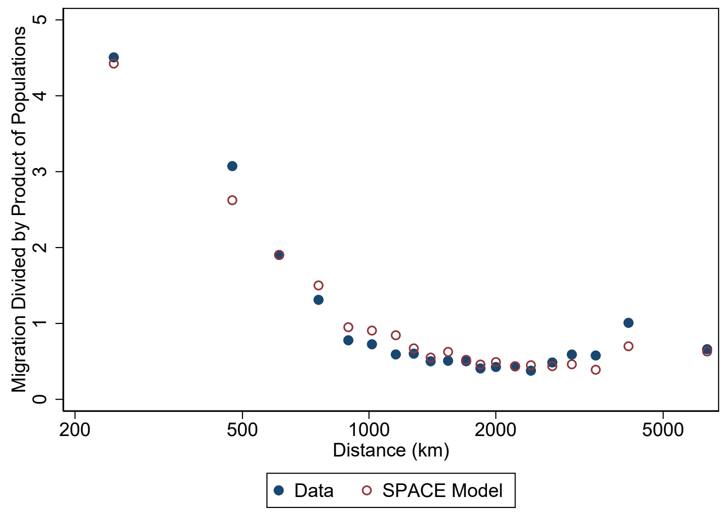


Figure 3: Migration and distance. For each state-pair, the value on the y-axis is calculated by dividing the migration between the two states by the product of the two states' populations. We then normalize that value so that the average across all state pairs is 1. The value on the x-axis is the distance 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.

the fact that a few percent of people move four or more times over the fourteen years.[19]

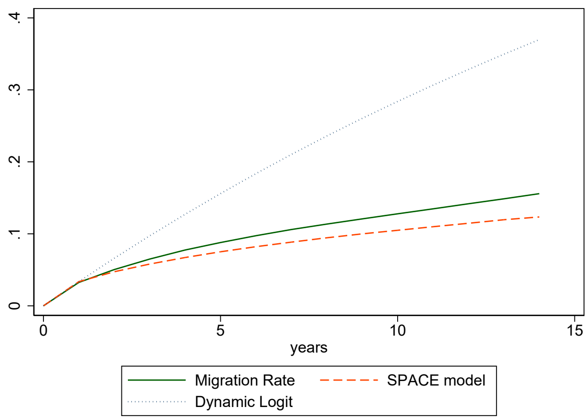
Another common moment is to look at the conditional probability of moving given a previous move. This is sometimes split up between onward and return migration. Return migrants are those who move back to their original state after an interstate move. Onward migrants move to a new state after an interstate move, and the new state is not the state where they came from.

Figure 4c shows the probability of return migration and onward migration at different time horizons after an interstate move.[20] In the first year after a move in the credit data, around 21 percent of migrants move back to their origin state one year after their initial move, and another 6 percent move to a new state.

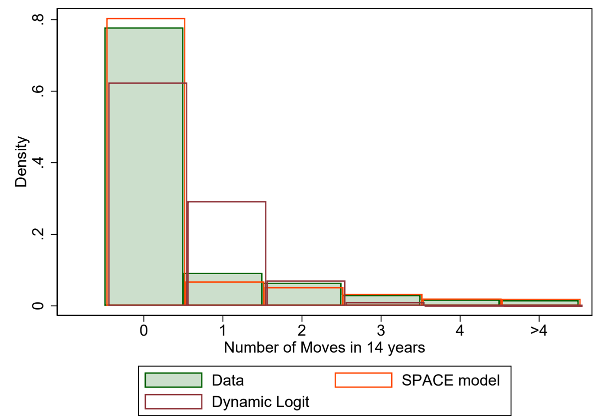
Since we do not target these statistics in the calibration, the simulated statistics do not match the data perfectly. In the simulated data, around 28 percent of migrants move back to the origin state one year after the inter-state migration, and 2 percent move to a new state. The return migration rate is generally higher in the simulated data, and the onward

[19] Again, the figure also includes similar statistics for a dynamic logit model, which we will discuss in Section 3.2.

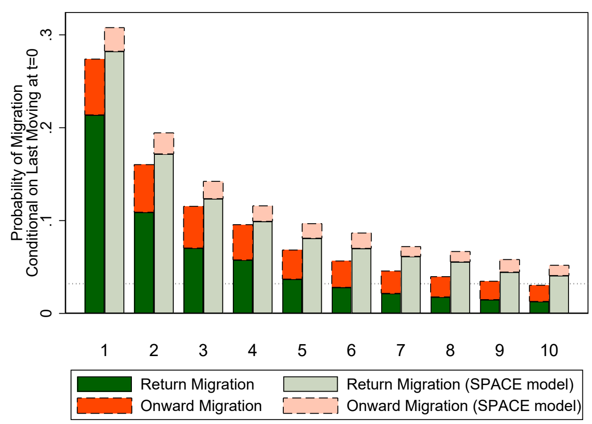
[20] To be included in this analysis, a person must show up for the number of years that would be necessary to calculate the statistic, but we do not use a balanced panel.



(a) Migration rate in the data and models



(b) Number of moves in 14 years



(c) Return and onward migration

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 number of moves in 14 years is calculated for people whose state is observed in every year from 2004-2018. In panel (c), 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.

rate is lower.

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 an economic force that increases their utility by staying there longer.[21]

2.3 Implications of the Model

In this section, we show how this model behaves in response to shocks or policies that affect the location specific utilities, v_i. Many important economic questions in this literature can be answered only if we know how population will respond to changes in local utility levels. 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 showing that the elasticity is related to the observed amount of migration between places.

Our main result is that the cross-elasticity of the population with respect to the utility in another state is based on the migration rate to that state, adjusted for the covariance of the utility shocks:

Proposition 3. *As $\rho \to 1$, the derivative of population with respect to the utility of another state is proportional to the migration level adjusted for the covariance of shocks:*

$$\frac{\partial p_i}{\partial v_j} = -\lim_{\rho \to 1} m_{i \to j} \frac{1}{\sqrt{1-\Sigma_{ij}}} \sqrt{\frac{\pi}{1-\rho^2}} \tag{7}$$

when $i \neq j$.[22] *In elasticity terms,*

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

[21] 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.

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

Recall that $\Sigma_{ij} = \exp(-a \text{ distance}_{ij})$, meaning that migration, along with the parameters of the model are a sufficient statistic for calculating the elasticities. Given that the utility is normalized such that the diagonal of the covariance matrix Σ is 1, the constant term $\sqrt{\pi/(1-\rho^2}$ does not have economic meaning.

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. 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 $\sqrt{\frac{1}{1-\Sigma_{ij}}}$ is that the shocks that induce migration are also correlated to one another, and so when calculating the effect of a shock that is uncorrelated, you have to adjust for that correlation.

An implication of this result is that a state's own population response to utility is the additive inverse of the sum over all other states:

$$\frac{\partial p_i}{\partial v_i} = \lim_{\rho \to 1} \sum_{j \neq i} \left[m_{i \to j} \frac{1}{\sqrt{1-\Sigma_{ij}}} \sqrt{\frac{\pi}{1-\rho^2}} \right]$$

So places with higher steady-state migration rates also have more elastic populations.

Knowing $\frac{\partial p}{\partial v}$ is useful for counterfactuals, but it is also important for welfare. The first order effect of a change to v_i on utility is: $\frac{\partial \mathbb{E}u}{\partial v_i} = p_i$. And so the second derivative is: $\frac{\partial^2 \mathbb{E}u}{\partial v_i \partial v_j} = \frac{\partial p_i}{\partial v_j}$. We can use that to approximate the change in utility to change in v, to second order:[23]

$$d\mathbb{E}u \approx p \cdot dv + \frac{1}{2} dv^T \frac{\partial p}{\partial v} dv \tag{9}$$

The first-term is straightforward. If the baseline utility of living in state i gets worse, that affects welfare in proportion to the number of people living in state i. The second term is more interesting because it encompasses the "insurance" aspect of migration.

If only state i is shocked, then this insurance term is $\frac{1}{2}\frac{\partial p_i}{\partial v_i} dv_i^2$. So in response to a negative shock (i.e. $dv_i < 0$), the second-order term is still positive, and so it acts as insurance for the shock because some people have moved out. Places with more migration and that are closer to other states, i.e. have a higher $\frac{\partial p_i}{\partial v_i}$, have more insurance.

When states i and j are shocked, the second-order welfare effect is:

$$\frac{1}{2}\frac{\partial p_i}{\partial v_i} dv_i^2 + \frac{1}{2}\frac{\partial p_j}{\partial v_j} dv_j^2 + \frac{\partial p_i}{\partial v_j} dv_i dv_j \tag{10}$$

[23]This formula is true of any discrete choice model, but highlights the importance of $\frac{\partial p}{\partial v}$.

The first two terms were discussed above, and the new term is the third term, involving the interaction of the shocks. Remember that $\frac{\partial p_i}{\partial v_j}$ is negative. So if two regions that have high migration and are close to one another are both shocked, that reduces the amount of insurance which comes from migration.

Discussion

Several key questions in the literature are directly related to Proposition 3.

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 Cruz and Rossi-Hansberg (2021), Oliveira and Pereda (2020), and Rudik et al. (2021) with respect to climate change. Proposition 3 shows that the answer is informed by the distribution of the shocks and the patterns of migration. Places with high gross migration, like D.C. have very elastic population to local shocks, compared to places with little migration. This is true in both the short-run and the long-run. Similarly, the effects are felt in states that have lots of migration between them. 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).[24]

Another key question in this literature is how quickly the migration adjustment takes place (Liu et al., 2021; 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. Of course, equilibrium forces in housing markets (Glaeser and Gyourko, 2005) or labor markets (Howard, 2020) could mean that utility adjusts slowly, but in this model, there is no delay that comes inherently from migration.[25]

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). Within our model, equation (9) is central to answering such questions. Because the second order effects are determined by these population elasticities, it shows that

[24] In the long-run of the workhorse model of migration, population elasticities converge to something akin to a static logit, in which everywhere has the same long-run population elasticity to their own shocks, and the cross-elasticities are given only by population shares of the affected states. There is quantititatively very little role for the amount of migration, or the distance between the states in determining long-run elasticities. We discuss this more in Section 4.

[25] This is not the case in the dynamic logit model that many of these papers use, in which migration is one of the main forces that slow down adjustment. We discuss further in Section 4.

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.

A major point that comes from Proposition 3 is that 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.

We illustrate this point by considering the changes in utility across the United States between 1970 and 2018. Using the model, we can find v_{it} by recalibrating the model to population shares in both 1970 and 2018.. Alternatively, we can also approximate the utility changes, up to a constant, by

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

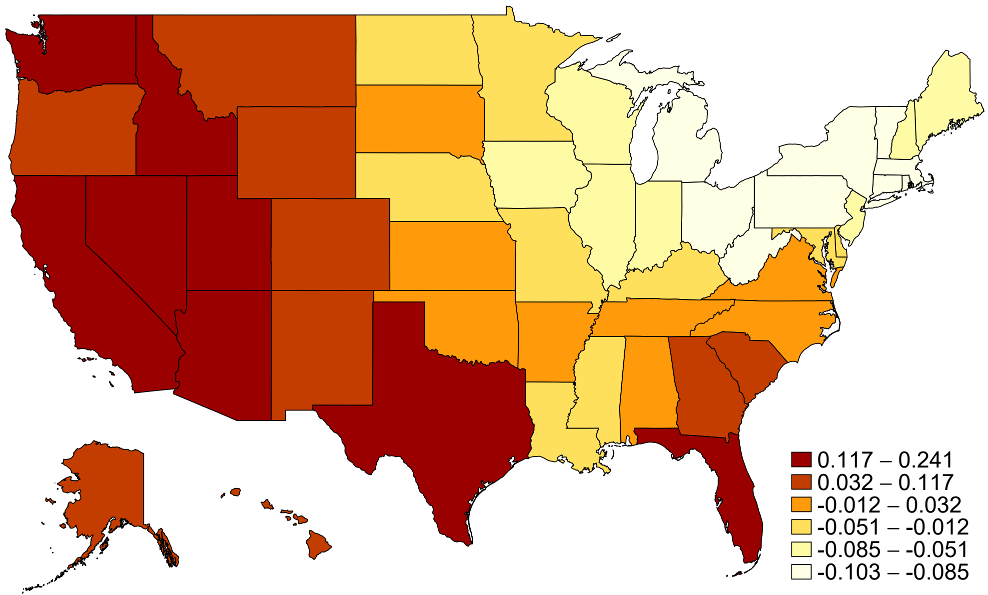
where dv^* is dv with one state removed to serve as a normalization, and the corresponding rows and columns removed from $\frac{\partial p}{\partial v}^*$. Similarly, dp^* is the change in population shares of all states but the normalized one. Figure 5a is a map of these utility changes which shows that utility increased by much more in the South and West of the country than it did in the Northeast.

Using the utilities implied by the model, we can focus on the question of how did the spatial nature of the utility changes affect migration insurance. From equation (9), we can calculate the second-order welfare term that comes from migration insurance, which is 0.021, or 16 percent of a standard deviation of the changes to v_i (weighted by 1970 populations).[26] To put this number in more easily interpretable context, the insurance that came from migration is about the same magnitude as the first-order effect of the relative utility declines in $dv_{\text{New York}}$ and $dv_{\text{Pennsylvania}}$, which were the largest relative utility declines in the nation.[27]

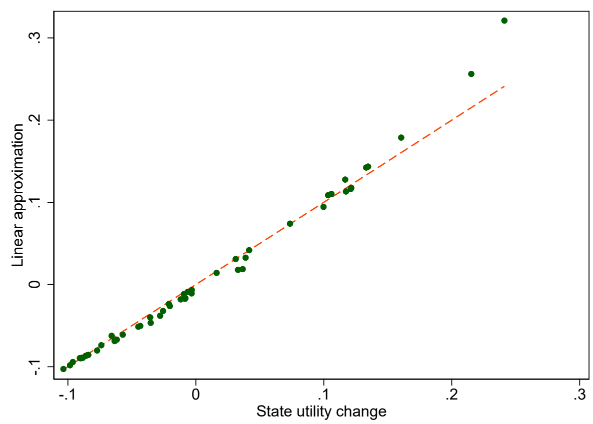
How important is the fact that the utility changes happened in the spatially concentrated way that they did? We consider two counterfactuals. In our first counterfactual, we shuffle the observed utility changes randomly between states, and calculate the second-order term for those states. In the second counterfactual, we draw each states' counterfactual utility from a random normal distribution, with the same variance as observed in the data (weighted by population). We do each of these 10,000 times. For the first counterfactual, the average second-order term is 0.041 and, for the second, the mean is 0.042. From either counterfactual, the average insurance is twice the insurance from the data. This tells us that the spatial correlation of utility changes has reduced the amount of insurance by 50 percent over the

[26] The first-order approximation is 0.0206 and the simulated effect is 0.0219, suggesting this approximation is fairly good in this instance.

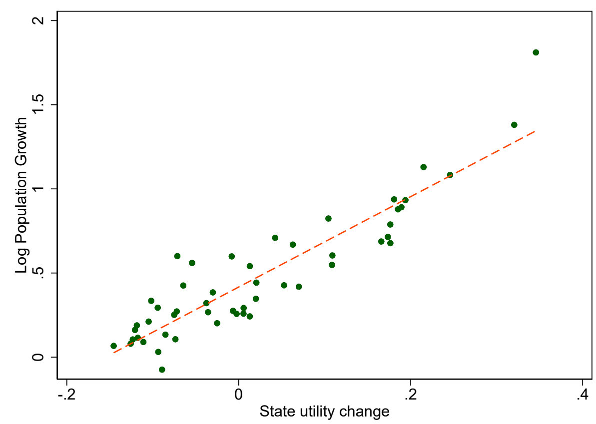
[27] Each of their utilities fell by a bit less than .15 relative to other states, and combined, they are a bit less than 15 percent of the population.



(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 5: 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 (11). The state with the largest utility increase is Nevada, followed by Arizona.

time period from 1980-2021.[28]

We also use this exercise to show that the changes in utility from the first-order approximation are close to the the changes in utility from simulations of the model that more closely match the true population changes. In Figure 5b, the approximation is very good except for the states that had the biggest increases in utility, where the linear approximation slightly overstates the utility changes.

We also stress that while the utility changes are related to population growth, there is not a one-to-one relationship. In Panel (c), we plot the utility changes against log population growth. While there is a significant correlation, a regression of population growth on utility changes would under-predict Nevada's population growth by 0.47 log points, and over-predict D.C.'s population growth by 0.25 log points. This reflects the fact that both of these states are high-gross-migration states, and so have a more elastic population than other states. However, the naive regression would also under-predict California's population growth by 0.21 log points, not because it is a high migration state, but because it is located near other states with even bigger increases, such as Nevada and Arizona.

2.4 Extensions of the model

Parameters are not geography-specific

In Appendix D, 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 in settings where the geography is not the state-level, as well. In fact, equation (7) is quite general. If one wants to ask what the population effect in any geography i is in response to a change in utility in geography j, one only need to know the migration from i to j and the distance, and one can use the parameters of this model to estimate this responsiveness.[29]

[28]We acknowledge that our counterfactual is only focusing on the migration block of the model. In a richer model, the correlations between states might be due to linkages between trade or other model features. In those models, assuming that the changes in v_i's are uncorrelated spatially may be unrealistic.

[29]The one caveat would be that this would not hold where the size of the region is so large that whatever

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, Smith and Wozniak, 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.[30]

Generalizing "Distance"

Another relatively simple 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 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 E. It leads to small improvements in matching migration patterns.

3 Tractability of the SPACE and workhorse models

While the primary purpose of this paper is not to disparage the typical dynamic logit model, it is worth comparing the model presented here to the by-far-the-most-common model in the literature. This section focuses on reasons why one might prefer our model to a dynamic logit model, focusing on the ease of modeling. Comparing the tractability of the models is important because one of the main reasons that the dynamic logit is popular in the literature is its tractability.[31]

value used for distance is no longer a good approximation for the distance to parts of that region.

[30] A model with this type of heterogeneity would have slightly different implications for 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.

[31] If a reader believes in Occam's Razor, that simpler arguments are more likely to be true, then this section could be taken to argue that the SPACE model is better than the workhorse model. However, given that neither model is going to capture all the features of migration anyway, we do not endorse this argument.

Later, in Section 4, we will show that the models also have different implications for counterfactual predictions and policy effects.

We consider the baseline dynamic logit model to be one in which, each period, agents make idiosynchratic draws from an extreme value distribution for each location, and pick the location which maximizes utility net of moving costs. For unfamiliar readers, we present what we consider to be the simplest version of the workhorse model in Appendix F.

3.1 Requires fewer state variables

A nice computational feature of this model, when the focus is on location-level outcomes, is that it has no state variables stemming from migration. In the dynamic logit, the state variables are the populations of each state at the end of the previous period, since the moving costs each person faces depend on their previous location. In contrast, the SPACE model has no state variables because the distribution of the person-location match-specific utility has a constant distribution. So while the specific people who live in a given location requires knowing their match-specific utility draws from the previous period, the spatial distribution of the population can be calculated without any state variables.

3.2 Dynamic moments

In Figure 4, we showed that SPACE model did a reasonably good job at hitting untargeted dynamic moments. The dynamic logit model does not. Looking back at the figure, in Panel (a), which plots the t year migration rate, the dynamic logit simulation has migration rates that increase proportional to t, much larger than the $\sqrt{t}$ of the data or the SPACE model simulation. In Panel (b), the dynamic logit model has many fewer people that never move over 14 years, and many more people that move exactly once. Almost no one moves more than four times, unlike the data or the SPACE model. Panel (c) does not draw the dynamic logit simulation, but that is because the probability of migration does not depend on previous migration in the baseline version of the dynamic logit. Hence, if we were to draw it, the conditional probability would always be approximately the same as the unconditional probability: about 3 percent, with most of that being onward migration and not return migration.[32]

Of course, many papers, going back to Kennan and Walker (2011), modify the baseline dynamic logit model in order to increase the amount of return migration. However, this

[32] The reason this statement is approximate is that people who recently moved are slightly more likely to live in high gross migration states, and since the state of residence does determine future migration probability, there is a very slight increase in the conditional probability of moving. In simulations, this concern is not visible when creating a figure similar to Figure 4c.

makes the previous point about the number of state variables even more salient. In order to match high rates of return migration, the dynamic logit model requires an additional state variable: how many people that live in location i and that used to live in location j. In order to match the facts in Figure 4c, the modeler would also need to keep track of how long each type of person had lived there.[33] Already this is $N^2 \times T$ number of state variables, where N is the number of locations and T is the length of time you wish to keep track of their presence. With 50 U.S. states plus D.C. and 10 years to match the moments in Figure 4c, that is 26,010 state variables. It may be that this reason is why many papers that include migration in quantitative spatial dynamic models do not account for return migration (Caliendo et al., 2019; Liu et al., 2021; Schubert, 2021).

In contrast, the model in this paper can match the dynamics in Figure 4 with zero state variables.

3.3 Return Migration

In the data, people are not just more likely to return to the state in which they most recently lived, but also to return to a state that they lived in before that. In the GCCP, if we focus on people that have lived in at least three different states over the last ten years, they have a probability of moving to the third most-recent state (i.e. not the current state of residence nor the most recent place from which they moved, but the state prior to either of those two states) of about 2.3 percent.[34] This is much higher than the probability of moving to that state if we condition only on living in the same state currently and having the same most-recent state. That conditional probability is only about 0.18 percent. So they are more than thirteen times more likely to move to that third state if they had previously lived there.

The SPACE model also has this feature. The conditional probabilities are a little bit higher. People move to this third place about 4.9 percent of the time if they had previously lived there, and only about 0.37 percent of the time conditional only on the last two states of residence. So like in the data, they are about thirteen times more likely to move to that third place if they had previously lived there.

Of course, to match this fact in a standard dynamic logit, one would need to keep track of the third most-recent state as another state variable. So the number of state variables increases not just in T, but also in N^3, if a modeler tries to match this feature of the data.

[33] Figure A3 shows that including the state in the past two years is not sufficient to generate a square root pattern.

[34] For this exercise, we focus on the subset of the GCCP for which we have the location of the person in every year.

3.4 Short-run migration elasticities

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

$$\lim_{\rho \to 1} \sqrt{1-\rho^2} \frac{\partial \log m_{i \to j,t}}{\partial v_{j,t}} = \frac{\sqrt{\pi}}{\sqrt{1-\Sigma_{ij}}} \tag{12}$$

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.

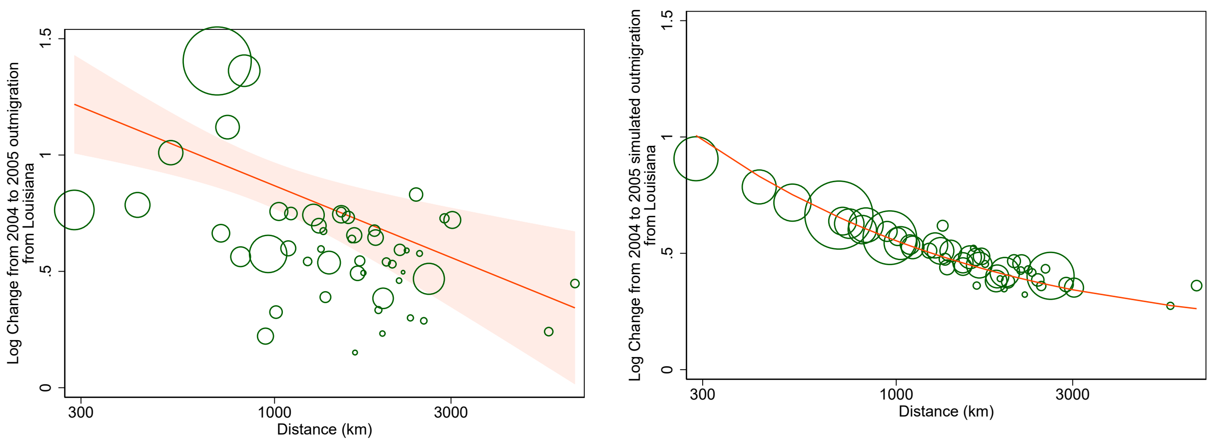
This contrasts to the baseline dynamic logit model, in which migration elasticities are constant.

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.[35]

We show the log-change in the amount of migration from Louisiana to other states after the hurricane in Figure 6. The left subfigure is the observed data, and the right 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

[35] 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 6: 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 $1/\sqrt{1-\Sigma_{ij}}$.

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.[36]

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.

Of course, one could modify the logit model to allow for differing elasticities across space, for example with a mixed logit (Train, 2009). The point of this exercise is not to falsify the whole class of dynamic logit models. Rather, the point of this exercise is to show that within the class of dynamic logit models, modeling migration requires additional features to match empirical facts that come naturally to the SPACE model.

[36] 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.

4 Implications of the SPACE and workhorse models

On several key questions in the literature, there are substantial differences between the predictions of the dynamic logit and the SPACE model. In this section, we explain a few of those differences. Testing any of these would be hard, which is why economists have typically turned to quantitative spatial models to answer these questions. So the goal is not to set up testable predictions with the aim of distinguishing the models. Rather, this section serves as a caution against accepting certain conclusions of a dynamic logit model, when another model that can match similar facts about internal migration would make other predictions.

4.1 Moving costs need not be large

Kennan and Walker (2011) estimate an average moving cost of $312,146 (in 2010 dollars).[37] This is more than six times the median household income in that year, which was $49,445 (Census, 2011).[38] Such a large cost is one of the main reasons that 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. In other words, the fact that most people do not move is not sufficient evidence to conclude that moving costs are large.

Another important difference regards changing the moving costs, a common counterfactual in the literature (Kennan and Walker, 2011; Schubert, 2021) and a policy used by some localities.[39] For example, Kennan and Walker (2011) finds that a moving subsidy could substantially increase the gross migration rate. In a dynamic logit 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 move, 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.[40]

[37]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.

[38]Most papers focusing on the United States estimate moving costs of a similar magnitude (Monras, 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.

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

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

4.2 Aggregate dynamics

One major difference between the two models is the persistence of migration flows. In the dynamic logit model, if utility in region j increases permanently, the rate of migration from i to j increases permanently. This is because the idiosynchratic draws of utility for j are going to be new each period, and so a constant fraction are going to get a good enough draw to move.

In contrast, in the SPACE model, the rate of migration will increase only in the first period. In the period in which utility changes, people will move in response because there is a mass of people that now prefer the location whose utility increased. But afterwards, the amount of gross migration should be approximately the same.

In the data, migration is quite persistent. For example, the Rust Belt has had low inmigration for decades, and the Sun Belt has had high inmigration for decades. The two models interpret this persistence quite differently. In the dynamic logit model, a significant amount of this persistence is due to the fact that migration is inherently persistent (Liu et al., 2021), 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, and there are also ample possible mechanisms for persistence in utility in the literature. 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.[41] It could be that initial shocks are small, but then as people move in, they become amplified with a delay (Howard, 2020). The spatially-correlated persistent preferences model would emphasize these various forces as reasons for persistence in migration, whereas the dynamic logit model would attribute the persistence to migration itself, and find less explanatory power for these forces.

4.3 Implied Utility Changes

A third way in which the models differ is in how to calculate utility changes. In the SPACE model population responds according to the equation: $dp = \frac{\partial p}{\partial v} dv$. So relative utility changes can be solved by

$$dv^* = \left(\frac{\partial p}{\partial v}^*\right)^{-1} dp^* \tag{13}$$

[41] Liu et al. (2021) discuss how having location specific durable capital can keep wages high after a negative productivity shock.

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 dynamic logit model, populations are not the right statistic to look at to calculate utility. But interstate migration is given by:

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

where μ is the migration elasticity. So roughly, 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

$$d \log m_{i \to j} = du_j + \alpha_i + \epsilon_{ij} \tag{14}$$

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

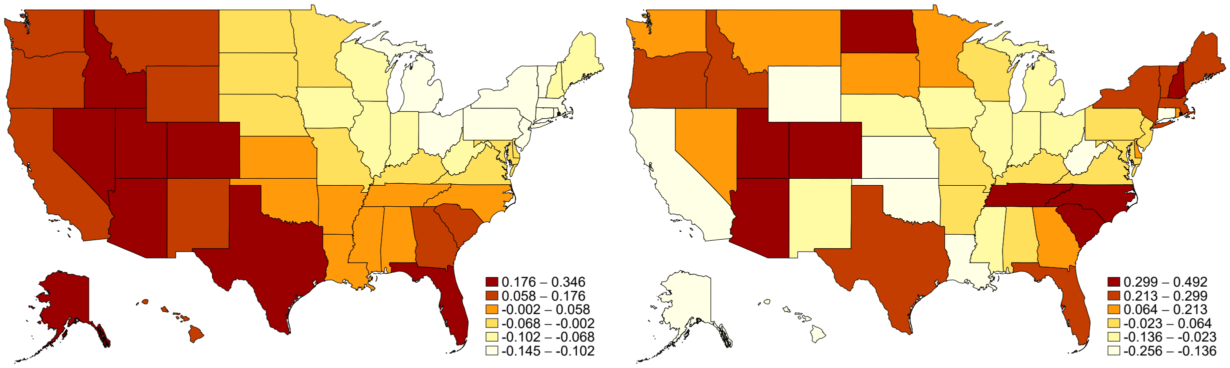
Fundamentally, there is a difference in what data identifies the utility of a location. In the SPACE model, it is the populations, and in the dynamic logit it is the migration rates. In the SPACE model, a place has gotten better to live if people have moved there. In the dynamic logit model, a place 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 dynamic logit model from 1990-2018.[42] 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 dynamic logit 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 a 0.16 correlation between the utility changes implied by the two models, but it is not statistically significant.

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

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 Figure 5c, we

[42] 1990 is the first year in which we have full state-to-state migration flows in the IRS data.



(a) SPACE model

(b) Dynamic logit model

Figure 7: Change in Utilities 1990-2018, implied by the SPACE model and the dynamic logit.

showed 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.

4.4 Long-run population elasticities

A fourth way in which the models differ is in their predictions for the long run population elasticities. In the short-run, the models both predict that a shock to a specific location will have larger population effects on places that have high migration rates with the shocked location. However, in the long-run the dynamic logit predicts that these cross-elasticities do not depend very much on migration, whereas the cross-elasticities in the SPACE model do.

The reason for this is that the dynamic logit's steady state is very similar to a static logit model. The population elasticities in a static logit are proportional to

$$\frac{\partial \log p_i}{\partial v_j} = \begin{cases} -p_j & \text{if } i \neq j \\ 1 - p_j & \text{if } i = j \end{cases}$$

where p_i is the population share of location i. The long-run elasticities for a dynamic logit are quite similar, which we show in Appendix F.

In the long-run, the dynamic logit 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 dynamic logit 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.

In contrast, both the short- and long-run elasticities for the model in this paper are given by Proposition 3. It predicts that states with higher migration shares, like D.C., will be more responsive to a shock, even in the long-run. And states that have more migration between them will have larger cross-elasticities, i.e. a shock to Maine will have larger long-run impacts on New Hampshire than it will on California. Neither of these implications would be true in the dynamic logit.[43]

As we argued in Section 2.3, it is important to get these cross-elasticities right for many of the questions in the literature. Unfortunately, this is a hard question to test empirically, but we can confidently say that conclusions related to the long-run cross-elasticities of population in a dynamic logit setting are likely not robust to using the SPACE model.

5 Conclusion

We propose a new model of migration that can generate a gravity relationship and match realistic migration dynamics. We show how to use moments of the data to calculate population cross-elasticities, which is the key statistic for counterfactuals and welfare. We show that the model has several advantages over the workhorse model, and that it has very different implications when it comes to moving costs, the speed of adjustment, how to infer utility changes from the data, and long-run population elasticities.

[43] In the dynamic logit, the migration rates govern the speed of adjustment (Liu et al., 2021), but not the long-run effects.

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.

Bergstrand, Jeffrey H, "The gravity equation in international trade: some microeconomic foundations and empirical evidence," *The Review of Economics and Statistics*, 1985, pp. 474–481.

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: Journal of the Econometric Society*, 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.

Chaney, Thomas, "The gravity equation in international trade: An explanation," *Journal of Political Economy*, 2018, *126* (1), 150–177.

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, 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," Technical Report, Working paper 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.

Heckman, James J, "Heterogeneity and state dependence," *Studies in labor markets*, 1981, *31*, 91–140.

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

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating trade flows: Trading partners and trading volumes," *The Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

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.

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.

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.

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/.

Lind, Nelson and Natalia Ramondo, "Trade with correlation," Technical Report, National Bureau of Economic Research 2018.

Liu, Ernest, Benny Klieman, and Stephen Redding, "Dynamic spatial general equilibrium," 2021.

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.

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

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," Technical Report, National Bureau of Economic Research 2021.

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.

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.

Zerecero, M, "The Birthplace Premium," Technical Report, Mimeo. 2021.

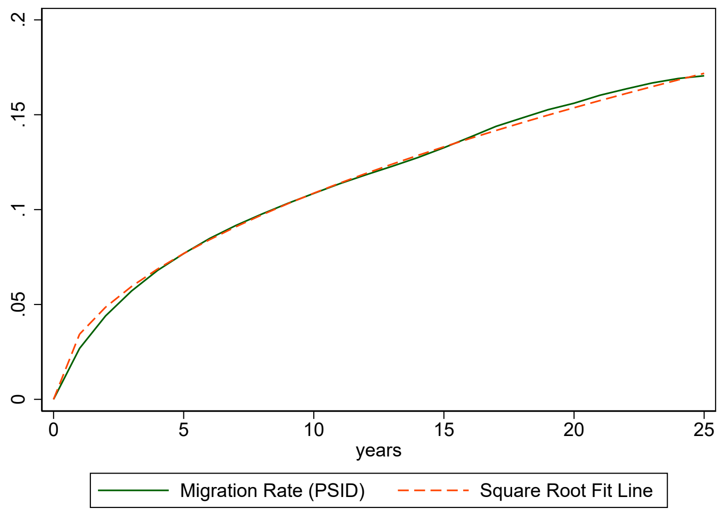


Figure A1: Square Root Fact in the PSID

A Robustness of the Square Root Fact

In this appendix, we show that the square root fact is robust to using other datasets and to considering subsets of the data. We also show that it cannot easily be explained by Markov models, even if they include age and recent location choices as state variables.

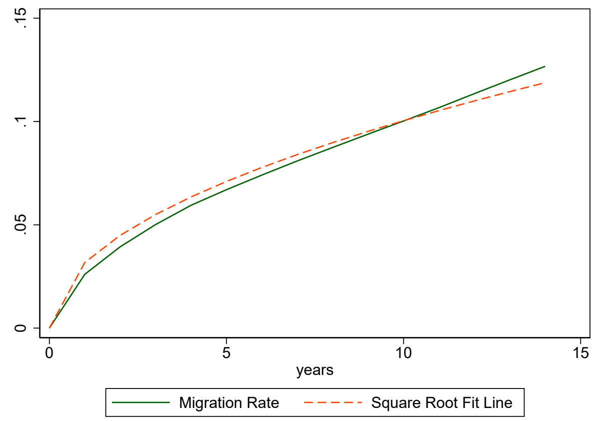
A reader may be worried that the GCCP is not a good dataset to study migration. For example, one worry would be that people do not always update their address information with their bank, especially for short-term moves, and therefore the amount of migration might be under-counted. For this reason, we show this fact also holds in the Panel Survey of Income Dynamics (1969-1997) (PSID). We use families that are surveyed from 1969 to 1997, and similarly calculate the t-year migration rate as the percent of people living in a different state than they did t years ago.[44] Our results are shown in Figure A1. We extend our analysis to 25 years because the dataset allows us to go beyond 14 years, and the square root line is a good fit throughout.[45]

A second worry that a reader might have is that migration is not measured well for young people. For example, college students may not be treated uniformly. 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.

[44] We focus on the years in which the PSID was adminstered annually. Since 1999, the survey has been done every other year.

[45] Beyond 25 years, the data is based on relatively few observations and jumps around a bit.

(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

Finally, we address the question of whether the square root fact is truly new by considering whether a Markov process could easily explain it. In Figure A3, we simulate Markov processes and compare the resulting t-year migration rates to the square root line.

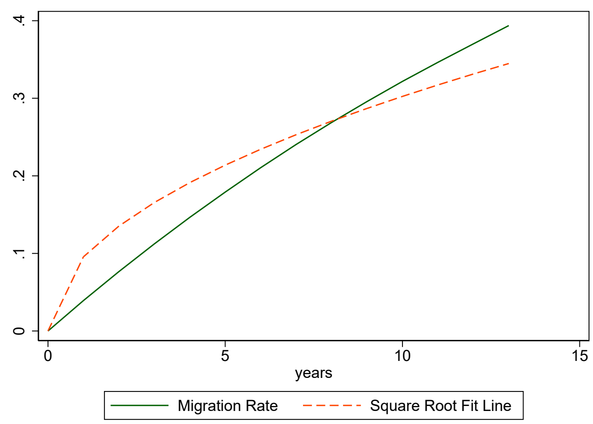
In our simplest Markov process, we assume the location of individual n at time t depends only on their location in $t-1$. We create a 51×51 Markov transition matrix and use it to simulate migration for several years. The result, shown in Figure A3a, is a t-year migration rate that is very close to linear, and does not have the curvature of the square root.

Of course, it is well-known that a significant amount of migration is return migration. Many models, including Kennan and Walker (2011), assume that migration depends on both the location at time t and the location at time $t-1$. We show this result in Figure A3b. Adding this extra dimension to the state space of the Markov process gives the model a bit of a kink between the 1-year and the 2-year migration rate, but the t-year rate beyond that is still close to linear and without the curvature of the square root function. So while the square root fact is related to the fact that many people move one year after a previous move, that fact is not sufficient to explain the entirety of the square root fact.

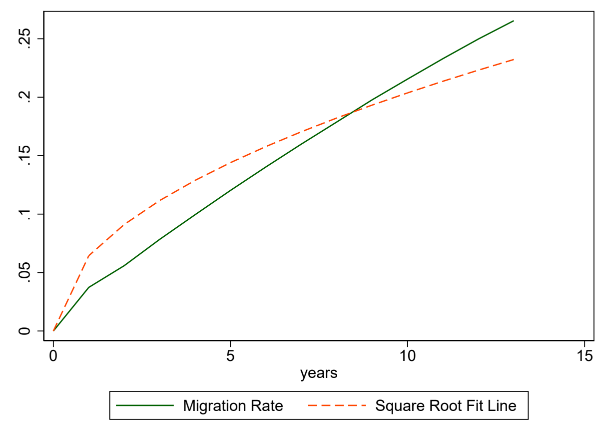
Similarly, including age as a state variable does not help move away from the linearity of the t-year migration rate (Panel c). So even though young people move a lot more, that is not sufficient to add noticeable curvature.[46]

Finally, in Panel d, we show that even including both the age and the state of residence

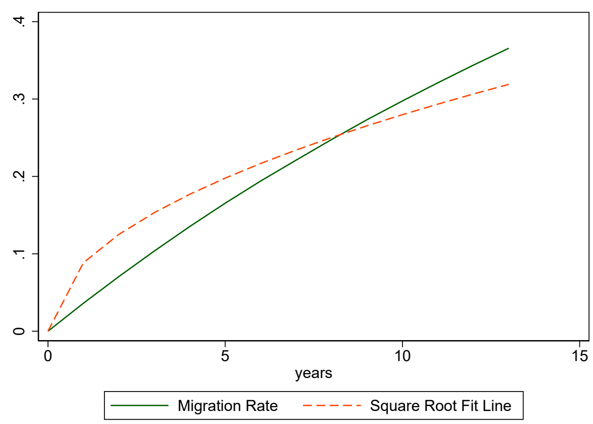
[46] One might wonder if other demographics would matter. For example, college educated workers are known to move more (Molloy et al., 2011). However, for immutable characteristics, the aggregate migration rate would just be the average of the different groups, so if the groups have linear t-year migration rates, then the aggregate would also be linear. We focus on age in this figure because it at least has the possibility that it would add some curvature to the t-year migration rate.



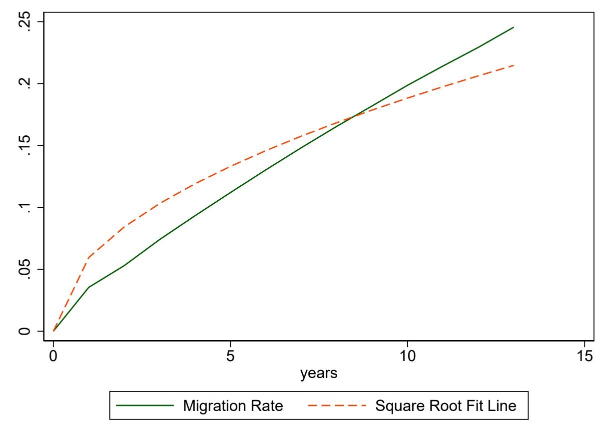
(a) State of residence in year t



(b) State of residence in year t and $t-1$



(c) State of residence in year t and age



(d) State of residence in year t and $t-1$ and age

Figure A3: The performance of simple Markov models to predict the square root fact

at $t-1$ does not help match the square root fact.

B Proofs

B.1 Proof of Proposition 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).[47] f is clearly non-constant, so it remains to show that $f(x)$ is completely monotone. Note that $\exp(-ax)$ is completely mono-

[47] 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.

tone. 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. □

B.2 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 $u_{in0} > u_{jn0}$ but $u_{jnt} > u_{int}$. By the recursive nature of the utility shocks,

$$u_{int} - u_{jnt} = v_{i0} - v_{j0} + \rho^t(\epsilon_{in0} - \epsilon_{jn0}) + \sum_{s=1}^{t} \rho^{t-s}\sqrt{1-\rho^2}(\eta_{ins} - \eta_{jns}) \tag{15}$$

The distribution of the shocks over t periods are given by

$$\sum_{s=1}^{t} \rho^{t-s}\sqrt{1-\rho^2}\eta_{ns} \sim N\left(0, (1-\rho^{2t})\Sigma\right) \tag{16}$$

So for a given utility difference $v_{in0} - v_{jn0}$, the probability that person moves to j in t is

$$1 - \Phi\left(\frac{v_{i0} - v_{j0} + \rho^t(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}\right) \tag{17}$$

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 and bounded. Then the total number of migrants over t years is

$$\int_{v_{j0}-v_{i0}}^{\infty} F_{ij}(\epsilon)\left(1 - \Phi\left(\frac{v_{i0} - v_{j0} + \rho^t\epsilon}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}\right)\right) d\epsilon \tag{18}$$

or with a simple u substitution,

$$\frac{1}{\rho^t}\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}} \int_{-\infty}^{\frac{(1-\rho^t)(v_{j0}-v_{i0})}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho^t}(w\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}} - (v_{j0} - v_{i0}))\right)\Phi(w)\,dw \tag{19}$$

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

$$\frac{1}{\rho^{t-1}}\sqrt{\frac{1-\rho^{2t}}{1-\rho^2}}\frac{\int_{-\infty}^{\frac{(1-\rho^t)(v_{j0}-v_{i0})}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho^t}(w\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}-(v_{j0}-v_{i0}))\right)\Phi(w)\,dw}{\int_{-\infty}^{\frac{(1-\rho)(v_{j0}-v_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(v_{j0}-v_{i0}))\right)\Phi(w)\,dw} \tag{20}$$

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

$$\begin{aligned}
\lim_{\rho\to1}\frac{1}{\rho^{t-1}} &= 1\\
\lim_{\rho\to1}\sqrt{\frac{1-\rho^{2t}}{1-\rho^2}} &= \sqrt{t}\\
\lim_{\rho\to1}\frac{(1-\rho^t)(v_{j0}-v_{i0})}{\sqrt{1-\rho^{2t}}\sqrt{2-2\Sigma_{ij}}} &= 0\\
\lim_{\rho\to1}\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(v_{j0}-v_{i0})) &= v_{j0}-v_{i0}
\end{aligned}$$

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

$$\sqrt{t}\frac{F_{ij}(v_{j0}-v_{i0})\int_{-\infty}^{0}\Phi(w)dw}{F_{ij}(v_{j0}-v_{i0})\int_{-\infty}^{0}\Phi(w)dw} = \sqrt{t} \tag{21}$$

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 $\sqrt{t}$. □

B.3 Proof of Proposition 3

From equation (19) in the proof of Proposition 2,

$$m_{ij} = \frac{1}{\rho}\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{\frac{(1-\rho)(v_{j0}-v_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(v_{j0}-v_{i0}))\right)\Phi(w)\,dw \tag{22}$$

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

$$\lim_{\rho\to1} m_{ij}\frac{1}{\sqrt{1-\rho^2}} = \rho\sqrt{2-2\Sigma_{ij}}\int_{-\infty}^{\frac{(1-\rho)(v_{j0}-v_{i0})}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}} F_{ij}\left(\frac{-1}{\rho}(w\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}-(v_{j0}-v_{i0}))\right)\Phi(w)\,dw$$

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

$$\lim_{\rho \to 1} m_{ij} \frac{1}{\sqrt{1-\rho^2}} = \sqrt{2 - 2\Sigma_{ij}} \int_{-\infty}^{0} F_{ij}\left(v_{i0} - v_{j0}\right) \Phi\left(w\right) dw$$

By the definition of F, the change in population is

$$\frac{\partial p_i}{\partial v_j} = -F_{ij}(v_{j0} - v_{i0}) \tag{23}$$

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,

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

$\square$

B.4 Proof of Proposition 4

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

$$1 - \Phi\left(\frac{v_{i0} - v_{j0} + \rho(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right) \tag{24}$$

where Φ is the standard normal cumulative density function.

If v_{j0} changes to v'_{j1}, then

$$1 - \Phi\left(\frac{v_{i0} - v'_{j1} + \rho(\epsilon_{in0} - \epsilon_{jn0})}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right) \tag{25}$$

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

$$\int_{v_{j0} - v_{i0}}^{\infty} F_{ij}(\epsilon) \left(1 - \Phi\left(\frac{v_{i0} - v'_{j1} + \rho\epsilon}{\sqrt{1-\rho^2}\sqrt{2 - 2\Sigma_{ij}}}\right)\right) d\epsilon \tag{26}$$

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

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

Evaluating it at $v'_{j1} = v_{j0}$,

$$\frac{\partial m_{ij}}{\partial v'_{j1}} = \frac{1}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}} \int_{v_{j0}-v_{i0}}^{\infty} F_{ij}(\epsilon)\phi\left(\frac{v_{i0}-v_{j0}+\rho\epsilon}{\sqrt{1-\rho^2}\sqrt{2-2\Sigma_{ij}}}\right) d\epsilon \tag{28}$$

Making a u-substitution,

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

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

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

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

$$\lim_{\rho\to 1}\sqrt{1-\rho^2}\frac{1}{m_{ij}}\frac{\partial m_{ij}}{\partial v'_{j1}} = \frac{1}{\sqrt{2-2\Sigma_{ij}}}\frac{F_{ij}\left(v_{i0}-v_{j0}\right)\int_{-\infty}^{0}\phi\left(w\right)dw}{F_{ij}\left(v_{i0}-v_{j0}\right))\int_{-\infty}^{0}\Phi\left(w\right)dw} \tag{31}$$

Simplifying,

$$\lim_{\rho\to 1}\sqrt{1-\rho^2}\frac{1}{m_{ij}}\frac{\partial m_{ij}}{\partial v'_{j1}} = \frac{\sqrt{\pi}}{\sqrt{1-\Sigma_{ij}}} \tag{32}$$

□

C 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 v_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 v_i that generate the populations in period 1.
3. Given that a and the v_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.

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

We start with an initial guess for the vector of v_i's and use any number close to 1 for ρ,[48] simulate two periods of the model with 10 million people,[49] add up the m_{ij}'s in the simulation, and then calculate the approximate cross-partials. 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 $v_1 = 0$. We update the other v_i's based on inverting an $I-1$ by $I-1$ matrix of the $\partial p_i / \partial v_i$ 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.

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.[50] 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

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

[49]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.

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

"too-high" or "too-low" guess with the previous guess. We repeat this loop until we match the gravity distance coefficient to four digits.

D Commuting Zones

In this appendix, we demonstrate that the parameters estimated in Section 2.2 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 v_i using commuting zones instead of states. We do not recalibrate either ρ or a.

	(1)	(2)
	Migrants (IRS data)	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.

E 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.[51]

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

[51] 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.

the education level another dimension:

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

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 $\sqrt{b} = 908.4$.[52] 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 3 that the baseline model can capture that non-linearity.

F Dynamic Logit Model

Throughout the paper, we compare simulations of the SPACE model to a standard dynamic logit model which we present here.

As in the main model, individuals are denoted by n and locations by i. An agent that lived in j at $t-1$ has utility:

$$u_{nt}(j) = \max_i u_{int} = \max_i v_{it} - \delta_{ij} + \epsilon_{int} \tag{35}$$

where v_{it} is the baseline utility of 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. In this case,

$$\frac{m_{i\to j,t}}{p_{it}} = \frac{e^{v_{jt}-\delta ij}}{\sum_k e^{v_{kt}-\delta_{ik}}} \tag{36}$$

where δ_{ii} is normalized to 1.

[52] We actually calibrate a^2/b^2 in the code for computational purposes, so that a has less of an effect on the importance of the education difference.

Note that the agents in this model are myopic. This avoids the feature of these models that location i becomes more desirable if it has lower moving costs to an actually desirable location j. Of course, this can be relaxed.

Long-run population elasticities

In the long-run of this dynamic logit model, the population elasticities are quantitatively similar to those of a static logit model. As discussed in Section 4.4, the population elasticities of a static logit are proportional to:

$$\frac{\partial \log p_i}{\partial v_j} = \begin{cases} -p_j & \text{if } i \neq j \\ 1 - p_j & \text{if } i = j \end{cases}$$

where p_i is the population share of location i.

Calculating these elasticities for the dynamic logit is not straightforward, but easy to do numerically. For each state, we change the utility by a small amount, calculate the new migration matrix from equation (36), and simulate 500 periods to see how much population in each state changes.

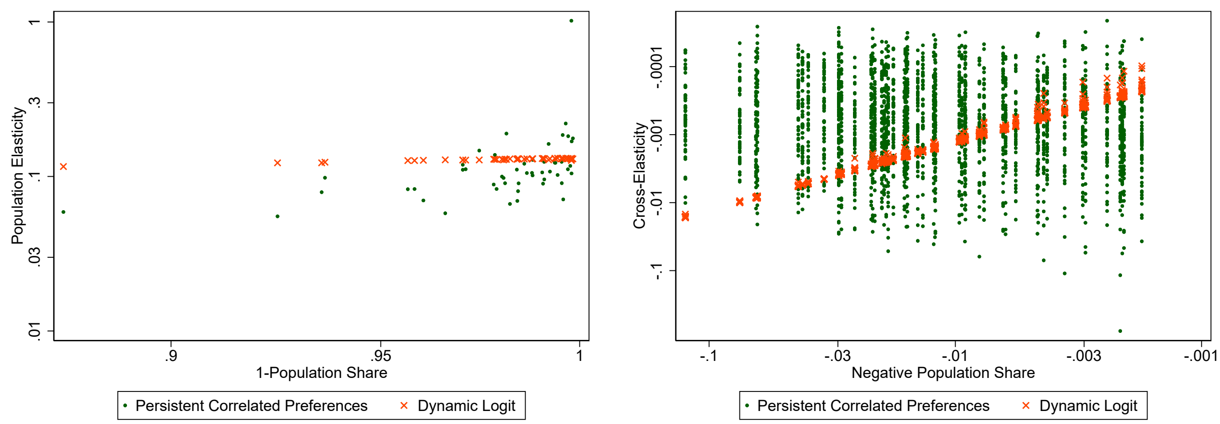
The overall correlation between the long-run elasticities of the dynamic logit and the static logit is 0.99996. Splitting it between same-state and cross-state elasticities, the correlation is 0.948 and 0.9994, respectively. Hence, it is a reasonable approximation to say that the dynamic logit approaches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

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.

In Figure A4, we show the comparison of the elasticities. In Panel (a), since all states are small compared to the whole U.S., the population elasticities are all about the same.[53] Each dot represents the population elasticity of a state with respect to a change in utility in that state. Although the elasticities are proportional to one minus the population share, that is hard to see on this graph since there is not much difference between those quantities. In contrast, the SPACE model has a wide range of elasticities. The highest elasticity is D.C. which has a high elasticity because it has a lot of migration to Maryland and Virginia and is quite close to them.

In Panel (b), we show the cross-state elasticities. For the dynamic logit, they follow a linear relationship with the negative of the population share of the state that is shocked. Each point is a state pair, so there are 50 green dots and 50 orange X's in each of the vertical

[53]Since the claims are about the relative size of population elasticities, we normalize so that the average same-state elasticity is the same for the dynamic logit and the SPACE model.



(a) Same-state. (b) Cross-state.

Figure A4: Population elasticities in the SPACE model and the dynamic logit vs. population shares. Log scaling.

stripes in the figure. The orange X's show a strong relationship between the cross-elasticity and population share. In contrast, the green dots do not follow the same relationship, and are hardly correlated to the population share at all. Rather, they are determined by the migration between the two states and the distance between them, following the formula in Proposition 3.