Migration Gravity, Networks, and Unemployment*

Arnab K. Basu† Nancy H. Chau‡ Gary C. Lin§

October 19, 2021

Abstract

We develop and estimate a theory-consistent gravity equation for interregional migration flows in the presence of unemployment. Micro-founded in a setting where search friction regulates labor market transitions, we derive a migration gravity equation for bilateral mobility that embodies a co-determined local unemployment term. As a theory of migration, our model connects directly with longstanding migration puzzles (e.g. declining internal mobility) as well as more recent developments (e.g. home bias). As a model of unemployment, a migration gravity approach uncovers novel inter-regional roots of local unemployment, and furnishes an unemployment sufficient statistic interpretation to the familiar multilateral migration resistance term. We empirically test the predictions of the model using U.S. county-level data on bilateral migration and unemployment rates, bilateral connectedness data such as Facebook friendship links, and instrumental variable identification based on a novel similarity index of counties’ historical ethnic-composition.

Keywords: Friendship networks, gravity equation, internal migration

JEL Codes: J61, J64, R23

*We thank Cynthia Bansak, John McLaren, Steve Matusz, Riley Wilson and seminar participants at the 2021 Annual Conference of the Society of Labor Economists and the 2021 Midwest International Trade Conference for insightful comments. All errors are our own.

†Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca NY 14853. Email: arnab.basu@cornell.edu. Tel: (607)255-6280.

‡Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca NY 14853. Email: hyc3@cornell.edu. Tel: (607)255-4463.

§Department of Economics, The Johns Hopkins University, Baltimore MD 21218.
1 Introduction

What can barriers to migration teach us about the causes of unemployment? Can an investigation into the search friction origins of unemployment shed light on labor mobility patterns? Research has long found evidence demonstrating labor outflows to be an equilibrating response to unemployment-driven spatial disparities at a regional level (e.g., DaVanzo, 1978; Greenwood, 1985; Blanchard and Katz, 1992; Dorn and Zweimüller, 2021), and in reverse direction the impact of regional labor inflows on destination unemployment has also received popular attention (e.g. Card, 1990; Dustmann et al., 2005). But on the link between bilateral labor mobility and unemployment, our understanding is notably scant. In particular, the workhorse model of migration gravity does not account for unemployment (e.g., Anderson, 2011; Grogger and Hanson, 2011; Bertoli and Fernández-Huertas Moraga, 2013). Canonical models of local unemployment, meanwhile, have until recently rarely considered the inter-regional roots of unemployment (Hall, 2003; Rogerson et al., 2005). Indeed, while studies have repeatedly pointed to social connections as a predictor of job search success and mobility (Chau, 1997; Munshi, 2003; Mckenzie and Rapoport, 2007; Bertoli, 2010; Beine et al., 2011), the difference that a network of social connections can collectively make on local unemployment and mobility remains an open question in a growing literature on commuting, mobility and job search in a spatial setting (e.g., Stoll and Raphael, 2000; Manning and Petrongolo, 2017; Monte et al., 2018; Tombe and Zhu, 2019; Caliendo et al., 2019; Notowidigdo, 2020).

In this paper, we develop and estimate a theory-consistent gravity equation for inter-regional migration flows adjusted for endogenous unemployment. We depart from the random utility model origins of the gravity equation, and adopt a multi-locational search-theoretic setup as the starting point with Poisson job arrivals and heterogeneous individual location preferences. Job arrivals from any given destination location depends on job vacancy, dyad-specific job search network links, as well as other third-party linkages which govern the overall general equilibrium sizes of the job seeker pools in each location.\(^1\) The desirability of a job offer at any given location, on the other hand, is the outcome of a utility draw from a destination-specific distribution function. We allow for the possibility of home bias in individual locational preference – the spatial expected utility premium a native resident attaches to her own origin relative to a new immigrant (Faini and Venturini, 2001). This may be due to status quo bias (Samuelson and Zeckhauser, 1988), social capital

\(^1\)This is in keeping with Manning and Petrongolo (2017), where workers are found to be discouraged from applying to jobs in areas with a competitive job seeker pool.
(Borjas, 1992), information asymmetries between residents and new migrants (Bryan et al., 2014), sunk investments (Albert and Monras, 2017), identity and preference (Djajic and Milbourne, 1988; Heise and Porzio, 2019) or relationships with ethnic enclaves (Albert and Monras, 2017). Each worker then maximizes utility by choosing the best option out of all job arrivals, if any. Workers who are not matched with any viable job offer remain in local residence as unemployed.

We solve for the equilibrium migration rates between any pair of locations in closed form. The revised migration gravity features (i) pair-specific bilateral search intensities to capture the strength of job and location information flow, (ii) location-specific expected utilities of employment, and (iii) an analogue of the familiar multilateral resistance term. Each of these features have parallels in standard migration gravity models. Pair-specific bilateral search intensity is the analogue of the standard bilateral migration cost / friction in our setting, meant to capture the ease with which workers are met with job offers across locations, mediated by family and friendship network links (e.g. Chau, 1997; Munshi, 2003), industry and occupational connections (e.g. Chen and Rosenthal, 2008), border effects (Wilson 2021) in addition to distance considerations (Manning and Petrongolo, 2017; De Weerdt et al., 2021). Our analogue of wage and amenities in the standard model is an expected utility term. In addition to home bias, we introduce adjustments allowing spatial disparities in the likelihood of finding a job in the expected utility calculus, where the effective size of the job seeker pool per vacancy in turn depends on the network of links connecting any given destination with sending locations.

Furthermore, we show that our revision of the familiar (outward) multilateral migration resistance term is a sufficient statistic for the local unemployment rate. This finding is intuitive, as multilateral resistance aggregates the strength of bilateral search intensities, weighted by the expected utility of each destination. As such, multilateral resistance encapsulates the inter-regional roots of unemployment in a single term that we will further explore in the paper.

As all three essential building blocks of migration gravity are preserved in our setting, we verify and otherwise modify key insights from the standard gravity model. In particular, we write down a structural migration gravity equation following by now standard practice to retrieve the iconic population product term. We find that the two parts of the unemployment-adjusted population product are (i) the sending location employment inclusive of emigrants, and (ii) the destination employment inclusive of all immigrants. The former is directly proportional one minus the sending location unemployment rate, while the latter implicitly incorporates the unemployment rates in all third party sending locations, as well as that of the sending-receiving location pair.
We then apply the model in two ways, both analytically to gain insight on a puzzle, and empirically to explore county-level migration and unemployment patterns in the US. A longstanding phenomenon in the U.S. labor market is that of a persistently low and declining internal mobility rate (Basso and Peri, 2020; Molloy et al., 2011, 2016; Dao et al., 2017) despite stark spatial disparities in unemployment incidence and wage differences. These trends are particularly puzzling in light of the ease and popularity of social media as means of interpersonal communication.\(^2\) By recognizing the role of search friction on bilateral search intensities and multilateral resistance, our model shows improvement in communication technology to be a double-edged sword – universal improvements in communication fosters connections both bilaterally, and multilaterally. Put simply, if there is universal better access to jobs, stronger bilateral access is then challenged by a more competitive job market as the effective pool of job seeker expands everywhere. The net impact on mobility can in fact be nil, or indeed negative if improvement in search efficiency is uneven, and biased in favor of less desirable locations.

Guided by our model, we estimate a gravity equation that relates bilateral county-level migration to relevant cross-county differences in search intensities and expected utilities. The strength of county-to-county friendship networks is our primary proxy for the degree of search intensity. Our measure of county friendship links come from the Social Connectedness Index developed by Bailey et al. (2018). The index is based on the number of Facebook friendship links between every county pair and between every U.S. county and every foreign country. Our analysis focuses on the friendship links between counties in the contiguous 48 U.S. states. To our knowledge, with its 239 million users, the Facebook dataset is the only dataset that provides a comprehensive coverage of friendship networks at the national level in the United States. Figure 1 plots counties’ average social connectedness, showing dense friendship networks on the U.S. coasts but also parts in the Midwest and the South.

To address the endogeneity of counties’ friendship networks, we construct a county-pair ethnic composition similarity index using the 1940 full-count Census records. Our identifying assumption is that historical ethnic networks are persistent over time and only affect current spatial linkages in population flows only through the social channel. To provide evidence for the validity of our instrument, we first demonstrate that our measure of ethnic distance has strong predictive power of both historical and current cross-county migration flows. Such path dependency going from

\(^2\)An emerging literature has begun to document the impact of social media communication on international migration (Dekker and Engbesen, 2014). Research on the impact of media networks on internal migration in developed country is as yet lacking.
Figure 1: Geographic Distribution of Social Connectedness

historic population composition to subsequent migration movements are commonly employed in empirical work (e.g. Hanson and Woodruff (2003), McKenzie and Hildebrandt (2005), Woodruff and Zenteno (2007)), the rationale being that prior ethnic networks facilitate future migration flows by mitigating information barrier and migration cost. In order to rule out the possibility that other historic economic activities – coincidental to historic ethnic composition similarity – are what drive migration, we construct bilateral indices capturing historic levels of industry composition and occupational composition similarity. We then demonstrate that these alternative historic linkages in fact do a poor job at predicting historic ethnic composition similarity.

We explore the empirical implications of the model in three applications. In the first, we distinguish between inflow migration gravity and outflow migration gravity.\(^3\) In empirical estimation on bilateral migration, one readily finds a variety of examples including outflow gravity (e.g. Eaton and Kortum, 2002), or geometric means of outflow and inflow gravity (e.g. Head and Ries, 2001). Guided by our model, we demonstrate theoretically and confirm in our empirical work that the effects of bilateral search intensity on outflow and inflow gravity are similar regardless of which model is chosen. To this, we also find that the location fixed effects of outflow gravity, being ratios of expected utilities, have a relative expected utility interpretation. Meanwhile, location fixed effects of inflow gravity are functions of multilateral resistance terms, have a relative employment interpretation.

These location fixed effects motivate a second application, in which we leverage the fact that each location in our data set is both an origin and a destination.\(^4\) We back out an estimate of

\(^3\)Respectively, these refer to the number of bilateral immigrants as a share of the total number of destination non-movers, and as a share of the total number of origin non-movers.

\(^4\)Our data set has observations on county-to-country migration rates between 452 unique counties. In these 452 counties, unemployment rates are strictly positive ranging from 2.2% to 23.6%, while total outmigration and total
home bias in location preferences as the difference between origin and destination fixed effects of the same location.\footnote{In a world without home bias, these two fixed effects should coincide since they both represent the expected utility proxy of the same location. With positive (negative) home bias, expected utility as seen by residents will be strictly greater (less) than that of new migrants.} We demonstrate that on average, individuals attach a 6\% premium to staying where they are after social connections proxied by our social connectedness indicator are accounted for. Failure to account for social connectedness, our estimates show, implies a more than ten-fold increase in the estimated home bias at 77\%. Home bias is also highly heterogeneous across U.S. counties, with highest average (positive) premia in New York, Virginia, Texas, California, Louisiana, and Minnesota. To flesh out the intuition behind these estimates of home bias even after accounting for social ties, we employ a Least Absolute Shrinkage and Selection Operator (LASSO) to identify significant correlates of our estimates from a wide array of county-level characteristics. We find that home bias is highly negatively associated with commute time and population density, and positively associated with employment factors such as the share of retail jobs with low entry barriers, and other demographic factors possibly related to retirement, including the share of population that are living alone, public assistance utilization, and summer temperature.\footnote{Heise and Porzio (2019) uses a similar approach to estimate home bias among workers born in East and West Germany respectively, by comparing destination fixed effects between workers born in, or foreign to the destination, to determine work home bias between West- and East-born Germans place higher premium to living in Western and Eastern Germany respectively. While the primary motivation in this paper is to ascertain the role of relationship networks in determining mobility and unemployment patterns, and in particular to see the difference in estimated home bias with and without accounting for relationship networks, Heise and Porzio (2019) does not account for network effects, or the origins of home bias such as demography that we are able to highlight here.} These findings are consistent \cite{boustan2013} and \cite{albouy2014}, for example, which demonstrate the key role played by public goods and amenities in mobility decisions. Our study adds to these findings the importance of accounting for the intensity and the configuration of the network of job search linkages as critical mobility determinants when estimating local preferences.

Finally, we perform unemployment regressions. Guided by our model, we capture the collective unemployment impact of a network of job search connections summarized by the outward multilateral resistance term by including as controls a level effect (the average level of bilateral network connection, the average destination expected utility proxied by destination fixed effects from the migration gravity model), and an interaction effect (a covariance term between bilateral network connection and destination expected utilities). We confirm the role of the level and interaction effects of search intensity on unemployment, and furthermore, we find that once network effects and state boundaries are accounted for, any corresponding level and interaction effects of distance on immigration are also universally and simultaneously positive from 615 to over 32,000 and from 345 to over 23,000 respectively.
unemployment are either no longer detectable, or of the wrong sign. This is consistent with the findings in Bailey et al. (2018) in which bilateral migration’s dependence on geographical distance is no longer detectable once bilateral friendship networks are included as controls.

This paper contributes to several areas of research. In addition to providing a micro-founded migration gravity equation that incorporates an endogenous level of unemployment, our analysis contributes to evidence that strong social network ties to destination communities facilitate migration (Mayda, 2010a; Beine et al., 2011; Bertoli and Fernández-Huertas Moraga, 2012; Beine and Parsons, 2015). The role of friendship connections in mitigating workers’ adverse labor market outcomes through migration is an understudied subject in the context of industrialized countries, such as the United States. Our investigation reaffirms the role of networks in reducing county-level barriers of migration.7 Our work is also related to Bailey et al. (2018), who document positive associations between counties where local residents higher shares of friends living within 100 miles and the higher incidence of worse of socioeconomic outcomes, including income, educational attainment, health, and social mobility. We build on their work, which is mainly descriptive, by documenting the causal effects of social networks on migration.

This paper also contributes to the local labor markets, commuting and spatial search literatures (e.g. Stoll and Raphael, 2000; Moretti, 2011; Manning and Petrongolo, 2017; Monte et al., 2018; Notowidigdo, 2020). From the lens of mobility, we closely examine the importance of geographical distance, state-boundaries, and social connections on bilateral search intensity simultaneously. In particular, we find that once friendship networks and state-boundaries are taken into account, the role of geographic distance alone on mobility is often no longer ironclad. On unemployment in a spatial setting, our model leverages the familiar multilateral resistance term to solve the problem of creating a summary index that meaningfully aggregates the unemployment impact of heterogeneous county-to-county connectedness (e.g. distance, networks, and economic similarity), with locations that differ in desirability across space.

The rest of this paper is organized as follows. In section 2, we spell out the multi-location search model, and derive closed form solution for migration gravity. We discuss its various interpretations, with particular emphasis on structural gravity, and the role of bilateral search intensities. We then provide in Section 3 the theoretical underpinnings for the three applications that we will empirically estimate. Section 4 presents our empirical application, in which we (i) evaluate the

---

7There is an extensive literature linking networks and migration, which network operates through providing job information (Borjas, 1992; Munshi, 2003), offering material support (Mayda, 2010a; Munshi, 2014), and reducing migration cost (Chau, 1997; McKenzie and Rapoport, 2007).
simultaneously role of social connections, geographic distance, and other economic connections on bilateral migration both through the lens of inflow and outflow gravity, (ii) present estimates of home bias with and without controlling for social connections, and (iii) discuss the results of our unemployment estimation, motivated by its equivalence with the multilateral resistance term.

2 Model

We consider the migration decisions of \( N_m \) number of job seekers in each of \( M \) locations with \( N = \sum_{m=1}^{M} N_m \). Let there be \( v_n > 0 \) number of employment vacancies in destination \( n = 1, \ldots, M \). Search friction prevents job seekers in origin \( m \) from sampling all \( v_n \) number of jobs in destination \( n \). The likelihood that a worker is met with \( z_n = 0, 1, 2, \ldots \) offers is given by a Poisson distribution with parameter \( \lambda_{mn} \geq 0 \):

\[
\Pr(z_n; \lambda_{mn}) = \frac{\exp(-\lambda_{mn}) (\lambda_{mn})^{z_n}}{z_n!}.
\]

The job arrival rate \( \lambda_{mn} \) depends on (i) the search intensity of workers from \( m \) in \( n \), \( a_{mn} \geq 0 \), (ii) the number of vacancies \( v_n \), and (iii) the search intensity adjusted number of job seekers \( a_{kn}N_k \) from all \( M \) locations in \( n \), \( J_n > 0 \), defined as follows:\footnote{The specification in (1) satisfies adding-up, namely total job arrivals in all \( M \) locations add up to the number of vacancies since:
\[
\sum_{k=1}^{M} \lambda_{kn}N_k = \sum_{k=1}^{M} a_{kn}N_k \frac{v_n}{J_n} = v_n.
\]}

\[
\lambda_{mn} = \frac{a_{mn}v_n}{\sum_{k=1}^{M} a_{kn}N_k} \equiv \frac{a_{mn}v_n}{J_n}.
\] \hspace{1cm} (1)

\( a_{mn} \) reflects the level of search intensity as practised by workers in \( m \) for jobs in \( n \), facilitated for example by social / career networks, geographic barriers such as distance, and other institutional barriers such as state boundaries.\footnote{We state \( \lambda_{mn} \) as a multi-destination analogue of the canonical job arrival rate in models of job search with one single location, \( v_m/N_m \) (e.g. Mortensen 2003).} All else equal an increase in \( a_{mn} \) raises job arrival \( \lambda_{mn} \). Naturally, an increase in search intensity in any other \( kn \) pairing, \( k \neq m \) will have the opposite effect, as it raises the intensity of job competition in \( n \) with other job seekers. This rise in competition is reflected in a matching increase in the effective number of job seekers in \( n \), \( J_n \equiv \sum_k a_{kn}N_k \) as \( a_{kn} \) rises.

We assume that the utility of location \( n \) for a worker from \( m \), accounting for wages and non-wage benefits such as amenities, is random and specific to each vacancy-worker match.\footnote{Random destination utilities is a common assumption in the mobility literature. See for example, Bertoli and Fernández-Huertas Moraga (2013), Dix-Carneiro (2014), Monte (2015), Redding (2016).}
distribution of this match-specific utility in location $n$, $\omega$, is characterized by a cumulative distribution function $F_{nn}(\omega) = F_n(\omega, 1)$ for workers native to $n$. We allow migrant workers to have different utility perceptions relative to natives, with associated distribution function $F_{mn}(\omega) = F_n(\omega, 1 + b_n)$, $m \neq n$, where we assume the following first order stochastic ordering:

$$F_n(\omega, 1 + b_n) \geq F_n(\omega, 1) \quad (2)$$

whenever $b_n \geq 0$.\textsuperscript{11} Put another way, positive (weakly negative) home bias in migration preference exists if and only if $b_n \geq (\leq) 0$.\textsuperscript{12}

To gain further insights, let $F_n(\cdot)$ assume a generalized Pareto distribution with parameter $\epsilon \in (0, 1)$ and $w_n \in (0, 1]$,\textsuperscript{13}

$$F_n(\omega, 1) = 1 - w_n (1 + \epsilon \omega)^{-1/\epsilon}, \text{ if } m = n \quad (3)$$

otherwise

$$F_n(\omega, 1 + b_n) = 1 - \frac{w_n}{1 + b_n} (1 + \epsilon \omega)^{-1/\epsilon}, \text{ if } m \neq n. \quad (4)$$

where $w_n \in [0, 1]$ and $w_n/(1 + b_n)$ are shift parameters, while $\epsilon$ is a shape parameter. The expected values of $\omega$ associated with $F_n(\omega, 1)$ and $F_n(\omega, 1 + b_n)$ are simply $w_n(1 - \epsilon)^{-1} > 0$ and $w_n[(1 - \epsilon)(1 + b_n)]^{-1} > 0$ respectively. Thus, our home bias term $b_n$, if positive, gives the spatial expected utility premium that a local resident attaches to her origin relative to that of a new resident (Faini and Venturini, 2001). The sources of home biases are many, due for example to status quo bias (Samuelson and Zeckhauser, 1988), prior investment in social capital (Borjas, 1992), information asymmetries between residents and new migrants (Bryan et al., 2014), sunk investments (e.g. home, schooling) (Albert and Monras, 2017), identity and preference (Djajic and Milbourne, 1988; Heise and Porzio, 2019) or relationships with ethnic enclaves (Albert and Monras, 2017).\textsuperscript{14}

\textsuperscript{11}This setup assumes implicitly that pair-specific network connections do not impact the expected utility of migration, conditional on finding a job. Naturally, network connections may do more than offering job opportunities to encourage mobility. We show in the appendix that such extensions can be readily accounted for, by accordingly adjusting our interpretation of the term $a_{mn}$. For our purpose here, $F_{nn}(\omega)$ and $F_{mn}(\omega)$ are the foundational characteristics that determine expected utility respectively for native workers, and for workers who originate elsewhere.

\textsuperscript{12}Samuelson and Zeckhauser (1988) define status quo bias as a “tendency to adhere to status quo choices more frequently than would be predicted by the canonical model.”

\textsuperscript{13}The generalized Pareto as a distribution class is commonly used in extreme value theory (Balkema and de Haan, 1974; Coles et al., 2001). The familiar exponential distribution, and the Pareto distribution are examples of special cases.

\textsuperscript{14}There are alternative definitions in the trade literature that are notable here. For example, the constructed home bias indicator of Anderson and Yotov (2010) is based on trade costs differences across locations. In our estimation results, we will further demonstrate that the strength of our estimated home bias is systematically correlated with a
At each destination \( n \), the probability distribution of the maximal utility sampled by a worker from \( m \) seeking a job in destination \( n \) is:

\[
p_{mn}(\omega) \equiv \sum_{z_n=0}^{\infty} \frac{\exp(-\lambda_{mn}) (\lambda_{mn})^{z_n} F_n(\omega)^{z_n}}{z_n!} = \exp[-\lambda_{mn}(1 - F_n(\omega))]. \tag{5}
\]

\( p_{mn}(\omega) \) is the probability that the highest utility job a worker finds is not better than \( \omega \).

Each worker then maximizes utility by choosing the best option out of all job arrivals, if any. Workers who are not matched with any viable job offer remain in local residence as unemployed. Substituting into \( F_{nn} \) and \( F_{mn} \), the distribution of the highest offer for a worker from \( m \) in destination \( n \) is:

\[
p_{mn}(\omega) = \exp[-\lambda_{mn}(1 - F_n(\omega))]
\]

\[
= \exp[-\lambda_{mn} w_n (1 + b_n)^{-1} (1 + \epsilon \omega)^{-1/\epsilon}] \tag{6}
\]

where \( \mathbb{I}_{mn} \) is an indicator variable which takes on the value of 1 if \( m \neq n \), and zero otherwise. (6) shows that the probability distribution \( p_{mn}(\omega) \) of the best offer for a worker from \( m \) to \( n \) assumes the functional form of a generalized extreme value distribution function,\(^{15}\) with parameters \( \lambda_{mn} \), and \( w_n (1 + b_n)^{-1} \). Higher search intensity through better network connection, or a higher \( \lambda_{mn} \), and a higher expected utility in \( n \), through \( w_n (1 + b_n)^{-1} \), both give rise to a first order stochastically dominating change in the distribution of the best offer from \( n \), all else equal.

### 2.1 The Decision to Migrate

Denote \( \mu_{mn} \) as the probability that a worker from \( m \) finds that the best utility draw in \( n \), \( \omega_{mn} \), to be more appealing that any other one of the \( M-1 \) locations’s best offers, \( \omega_{mk}, k \neq n \). Thus

\[
\mu_{mn} = \int_0^{\infty} Pr \left[ \omega \geq \max_{k \neq n} \omega_{mk} \right] dp_{mn}(\omega).
\]

Let \( \alpha_{mn} \) denote the home-bias adjusted search intensity

\[
\alpha_{mn} \equiv a_{mn} \left( 1 - \frac{\mathbb{I}_{mn} b_n}{1 + b_n} \right) \tag{7}
\]

list of factors that are not related to migration cost, such as county-level population density, commuting costs, and share of retail jobs, for example.

\(^{15}\)Many commonly used extreme value distributions such as Fréchet, Gumbell and Weibull distributions are special cases of the generalized extreme value distribution. For example, the Fréchet distribution obtains by setting \( \lambda_{mn} w_n (1 + b_n)^{-1} \) to unity, and a change of variables \( y = (1 + \epsilon \omega) \), and \( \beta = 1/\epsilon \), so that \( F(y) = \exp(-y^{-\beta}) \).
where home-bias’ contribution to migration friction when \( m \neq n \) is on display explicitly. Also let \( W_n \) denote the employment-adjusted expected utility of location \( n \) where

\[
W_n = \frac{w_nv_n}{J_n} = \frac{w_nv_n}{\sum_k a_{kn}N_k}.
\]

Now, by the law of large numbers, \( \mu_{mn} \) represents the fraction of the workers in \( m \) who prefers location \( n \) to any of the other \( M - 1 \) locations.\(^{16}\)

\[
\mu_{mn} = \int_0^\infty \prod_{k \neq n} p_{mk}(\omega) dp_{mn}(\omega)
= \left( \frac{\alpha_{mn}W_n}{\sum_{i=1}^M \alpha_{mi}W_i} \right) \left( 1 - \exp \left[ -\sum_{i=1}^M \alpha_{mi}W_i \right] \right).
\]

The expression

\[
O_m \equiv \sum_{i=1}^N \alpha_{mi}W_m
\]

is the direct parallel of the outward multilateral resistance term, capturing outward mobility friction in the standard migration gravity equation and trade gravity equation (Anderson, 2011; Bertoli and Fernández-Huertas Moraga, 2013).\(^{17}\) In the current setting, \( O^m \) normalizes bilateral search intensity \( \alpha_{mn} \) to account for the influences of all other locations on the relative desirability of \( n \) for workers in \( m \).

Importantly, our analogue of the outward multilateral resistance term in migration gravity with search friction is a sufficient statistic for the equilibrium unemployment rate. To see this, note that the total number of employed location \( m \) workers is

\[
\sum_{n=1}^M \mu_{mn}N_m = \left( 1 - \exp \left[ -\sum_{i=1}^M \alpha_{mi}W_i \right] \right) N_m = [1 - \exp(-O_m)]N_m.
\]

\(^{16}\)This follows since,

\[
\lambda_{mn}w_n = \frac{\alpha_{mn}v_n}{\sum_k a_{kn}N_n} (1 - b_nI_{mn}/(1 + b_n))w_n = \alpha_{mn}W_n
\]

by definition of \( \lambda_{mn} \) in (1), \( \alpha_{mn} \) in (7), and \( W_n \) in (8). Thus

\[
\int_0^\infty \prod_{k \neq n} p_{mk}(\omega) dp_{mn}(\omega) = \int_0^\infty \alpha_{mn}W_n (1 + \epsilon\omega)^{-1}/\epsilon - 1 \exp \left[ -\sum_{k=1}^M \alpha_{mk}W_k (1 + \epsilon\omega)^{-1}/\epsilon \right] d\omega
= \left( \frac{\alpha_{mn}W_n}{\sum_{i=1}^M \alpha_{mi}W_i} \right) \left( 1 - \exp \left[ -\sum_{i=1}^M \alpha_{mi}W_i \right] \right)
\]

where the last equality follows by definition of \( p_{mn}(\omega) \).

\(^{17}\)To see this, denote the inverse of our search intensity as migration friction, say \( t_{mn} = \alpha_{mn} \). The term \( O^m \equiv \sum_{i=1}^N W_{mi}/t_{mi} \) is what Anderson (2011) refers to as outward migration friction in a search friction free world.
The unemployment rate in location $m$, $u_m = 1 - \sum_{n=1}^{M} \mu_{mn}$ is thus uniquely captured by outward multilateral resistance:

$$u_m = \exp(-O_m). \quad (10)$$

(10) spells out the inter-regional roots of local unemployment. The stronger the total outward multilateral migration resistance, the higher will be the total unemployment rate.

**Proposition 1.** Bilateral mobility rates from $m$ to $n$, $\mu_{mn}$, depends on (i) bilateral home bias adjusted search intensities $\alpha_{mn}$, (ii) destination expected utility $W_n$, (iii) an outward multilateral resistance term $O^m$:

$$\mu_{mn} = \alpha_{mn} W_n (1 - \exp(-O_m))/O_m.$$

Furthermore, outward multilateral resistance is a sufficient statistic for the unemployment rate:

$$u_m = \exp(-O_m).$$

### 2.2 Structural Gravity

It is straightforward to express (9) as a structural migration gravity equation. Doing so can reveal migration and unemployment as co-moving outcomes of population stocks and employment aggregates. Thus, let $M_{mn} = \mu_{mn} N_m$ denote total migration, and $L_n = \sum_m M_{mn}$ as total employment in $n$, we have:\(^{18}\)

$$M_{mn} = \frac{\alpha_{mn}}{O_m I_n} \frac{L_n \times [N_m(1 - u_m)]}{\sum_i N_i(1 - u_i)}.$$  \(11\)

where $I_n$ is denotes multilateral resistance capturing inward migration friction impacting mobility for destination $n$, with

$$I_n = \sum_m \frac{\alpha_{mn}}{O_m} \frac{N_m(1 - u_m)}{\sum_i N_i(1 - u_i)} \quad (12)$$

and symmetrically, outward multilateral can be expressed as

$$O_m = \sum_n \alpha_{mn} W_n = \sum_n \frac{\alpha_{mn}}{I_n} \frac{L_n}{\sum_i N_i(1 - u_i)}.$$  \(13\)

Thus, total migration between two locations depends on (i) bilateral home bias adjusted search intensity $\alpha_{mn}$ normalized by both outward and inward multilateral resistance $O_m$ and $I_n$, (ii) a population product, involving the total number of employed workers native to $m$, $N_m(1 - u_m)$,

\(^{18}\)The steps are exactly analogous to the structural trade gravity equation in Anderson (2011), and the proof is relegated to the Appendix.
and the total number of employed workers (inclusive of migrants) in \( n \), \( L_n = \sum_m M_{mn} \). The employment product is normalized by the overall employment level \( \sum_i N_i (1 - u_i) \).

Several observations are in order. First, (11) prescribes the product of a particular pair of population / workforce indicators as the determinant in our structural migration gravity equation, \( N_m (1 - u_m) \) and \( L_n \). Of course unemployment is featured in both expressions. In \( N_m (1 - u_m) \), the number of employed sending location workers \( N_m (1 - u_m) \) (inclusive of outward migrants) applies, whereas in \( L_n \), total labor supply in \( n \) (inclusive of inward migrants) \( L_n = \sum_k \mu_{kn} N_k = \sum_k \alpha_{kn} W_n (1 - u_k) N_k / O_k \) applies.

Second, consider the special case where the search intensity across all locations are symmetric \( \alpha_{mn} = \alpha > 0 \). In this case, bilateral mobility simplifies to\(^{19}\)

\[
M_{mn} = \frac{N_m L_n}{N}. \tag{15}
\]

Put simply, with universal symmetry in search intensity even after adjusting for home bias, the fraction of workers from \( m \) in all destinations will be equal to its share of workers in total population \( (M_{mn} / L_n = N_m / N) \). Furthermore, mobility defined as the share of migrants from \( m \) to \( n \) in \( m \)’s total population is equal to the share of employed workers in \( n \) in total population

\[
\frac{M_{mn}}{N_m} = \frac{L_n}{N}. \tag{16}
\]

These are directly analogous to the migration friction and search friction free counterparts, even though search friction remains and unemployment prevails. The reason for these observations is that with symmetric search intensity, unemployment rates are the same everywhere for outward

\(^{19}\)To see this, note that outward multilateral migration resistens simplifies to

\[
O_m = \alpha \sum_n W_n = \frac{\sum_n v_n w_n}{N}
\]

for all \( m \) and thus both \( O_m \) and unemployment rates will be equalized across all origins, with \( u_m = u \). Furthermore, and once again under symmetry \( \alpha_{mn} = \alpha \), the inward multilateral resistance:

\[
I_n = \sum_m \left( \frac{1}{\sum_i W_i} \right) \left( \frac{N_m}{\sum_i N_i} \right) = \sum_m \left( \frac{1}{\sum_i W_i} \right) \frac{N_m}{N} = I. \tag{14}
\]

and thus \( I_n \) will also be equalized across all destinations. Moreover, the product of the inward and outward multilateral resistance is can be simply expresed:

\[
OI \equiv \alpha
\]

from (12) and (13). (15) obtains upon substituting these expressions in (11).
multilateral migration resistance is:

\[ O_m = \alpha \sum_n W_n = \frac{\sum_n v_n w_n}{N} = O \]

for all \( m \). This reiterates the fact that when search intensities are identical, workers in any location have equal access to jobs anywhere. The symmetric mobility ratios in (15) thus naturally follow.

Third, consider a proportionate improvement in communication technology across all locations by a factor of \( \gamma > 1 \) everywhere. Note that unemployment is unaffected by this improvement, since

\[ O_m = \sum_n \frac{\gamma \alpha v_n w_n}{\sum_k \gamma \alpha N_k} = \sum_n \frac{\alpha v_n w_n}{\sum_k \alpha N_k} \]

since rising search capabilities is matched with a rise in job competition in every location through \( J_n \).

Consequently, improvements in communication technologies do not guarantee rising employment, nor does it guarantee rising mobility, since

\[ \mu_{mn} = \int_0^\infty \prod_{k \neq n} p_{mk}(\omega) dp_{mn}(\omega) \]

\[ = \left( \frac{\alpha v_n w_n / J_n}{\sum_{i=1}^M \alpha v_i w_i / J_i} \right) (1 - u_m) . \]

is likewise invariant to equi-proportionate increases in \( \alpha_{mn} \) for the same reason. We have thus:

**Proposition 2.** Symmetric proportionate improvements in search intensity has no impact on mobility as measured by bilateral migration as a share of destination \( n \) employment \((M_{mn}/L_n)\), or as a share of total employment of workers native to the origin \( m \) \((M_{mn}/[N_m(1 - u_m)])\). Also, symmetric proportionate improvements in search intensity does not affect the unemployment rate \( u_m \).

Proposition 2 speaks to the puzzling observation of a persistently low level of labor mobility in the US and elsewhere (Basso and Peri, 2020), despite advances in information and communication technology by leaps and bounds in recent decades, along with greater ease in long distance interpersonal communication assisted by electronic communication. Of course, in practice, improvement communication technology have had a skewed impact on different communities, with resulting implications on migration and unemployment that will change depending respectively on (9) and (10).

What remains to be fleshed out is in the ways in which changes in search intensity impact unemployment, through its influence of mobility \( a_{mn} \). In three applications below, we work step by step towards answering this question.
3 Three Applications

We now have an estimable model of migration gravity in which heterogeneous search intensity, $\alpha_{mn}$, and location-specific employment-adjusted expected utility $W_n$, and the outward multilateral resistance term $O_m$ are simultaneously featured. In three applications below, we (i) examine empirically the role of search intensity on migration, and the meaning of location fixed effects, (ii) leverage location fixed effects to back out the degree of home bias, and (iii) ascertain the impact of search intensities on unemployment through a level and an interaction effect, emphasizing not just better connection between locations, but better connections with more desirable locations accounting for home bias.

3.1 Outflow Gravity and Inflow Gravity

The migration gravity model in (9) can be estimated in a number of ways (Anderson and van Wincoop, 2004). Following Mayer and Head (2002) and Eaton and Kortum (2002) among others, we take ratios of (9) relative to a baseline pair of locations, and effectively replace the multilateral resistance terms with sending and destination location dummies. We look at two different baselines. From a sending location perspective, consider the outflow of migrants as a share of workers who are left behind, henceforth outflow gravity, as the product of:

$$\frac{\mu_{mn}}{\mu_{mm}} = \left( \frac{a_{mn}}{a_{mm}} \right) \left( \frac{W_n}{W_m} \right) \left( \frac{1}{1 + b_n} \right), \quad m \neq n. \tag{17}$$

Three sets of push and pull forces are featured in (17): (i) the relative search intensities $a_{mn}/a_{mm}$, (ii) the ratio of destination and sending location expected utilities, and (iii) home bias at destination $n$. Taking logs on both sides, we obtain a migration gravity model of worker outflows, henceforth outflow gravity, where for any $n \neq m$,

$$\ln \mu_{mn} - \ln \mu_{mm} = \ln a_{mn} - \ln a_{mm} - T_m + D_n, \tag{18}$$

where sending and receiving location fixed effects ($T_m = W_m$ and $D_n = W_n/(1+b_n)$) have expected utility interpretations as perceived by local residents at sending locations $W_m$, and by potential migrants at destination locations $W_n/(1+b_n)$.

Analogously, let inflow gravity denote the inflow of migrants as a share of employed destination non-movers:

$$\frac{\mu_{mn}}{\mu_{mm}} = \left( \frac{a_{mn}}{a_{mm}} \right) \left( 1 - u_m \right) \ln(1/u_m) \left( \frac{1}{1 + b_n} \right).$$
The push and pull factors associated with inflow gravity are (i) the relevant relative search intensities \(a_{mn}/a_{nn}\), and (ii) relative employment rates, and (iii) home bias at destination \(n\).

\[
\ln \mu_{mn} - \ln \mu_{mm} = \ln a_{mn} - \ln a_{nn} + t_m - d_n, \quad (19)
\]

where sending and destination fixed effects \(t_m = (1-u_m) \ln(1/u_m)\) and \(d_n = (1-u_n) \ln(1/u_n)(1+b_n)\) have employment interpretations.

There are two important takeaways. First, outflow gravity \(\mu_{mn}/\mu_{mm}\) and inflow gravity \((\mu_{mn}/\mu_{nn})\) are symmetrically dependent on the relevant search intensities ratios, \(\alpha_{mn}/\alpha_{mm}\) and \(\alpha_{mn}/\alpha_{nn}\). Thus, both outflow and inflow gravity are appropriate modeling choices in empirical investigations on the role of search intensities on migration rates, once destination and sending location fixed effects are incorporated. It should be noted that the expected utility and employment interpretation of outflow and inflow gravity equation, noted in (18) and (19) above, have analogous counterparts in the canonical structural migration gravity model without unemployment (e.g. Anderson, 2011). We single this property out here as a first step towards leveraging estimated location dummies to back out location-specific home bias.

### 3.2 Home Bias

To start, we note that each location \(i = 1, ..., M\) in a migration gravity model appears both as a destination as well as an origin. Thus, with a full set of sending location dummies and destination dummies, associated with each location are two estimated fixed effects, once as a sending location \((T_i\) and \(t_i\)), and once as a destination \((D_i\) and \(d_i\)). Using notations developed for outflow and inflow gravity where location dummies have expected utility interpretations, and relative employment interpretations respectively (18, 19),

\[
T_i - D_i = \ln(1 + b_i) = d_i - t_i. \quad (20)
\]

Importantly, therefore, the difference between the destination and origin fixed effects, when \(m = n\), gives an estimate of the migration cost of each location \(i = 1, ..., M\). This is possible using both the outflow gravity equation, and the inflow gravity equation.

By construction, \(b_i\) is the expected utility premium that individuals in \(i\) attach to staying put. A positive \(b_i\) naturally acts as a mobility barrier and discourages labor movement. The distinction between home bias as opposed to search cost as a mobility barrier is that \(b_i\) is origin-specific, whereas our search intensity characterization of mobility barriers, \(a_{ij}\), is location pair-specific. The
two can be combined to form a single parameter of home bias adjusted mobility barrier, as we have done in the definition of \( \alpha_{ij} = a_{ij}(1 - I_{ij}b_j/(1 + b_j)) \) to parameterize the overall barrier to migration between \( i \) and \( j \). Quite intuitively, \( \alpha_{ij} \), and hence outward gravity from \( i \) to \( j \) is decreasing in \( b_j \).

Our task here is to separately tease out \( b_i \) from \( \alpha_{ij} \), where home bias by definition is driven by preferences and arguably not readily changeable, but the pure search friction term \( a_{ij} \) may be finetuned using appropriate policies to facilitate broader access to jobs, for example. Having a gauge on home bias can provide key insights on how much mobility can feasibly be increased by removing the search-related migration barriers alone, when policy makers do not have the tools to change people’s preferences.

### 3.3 Unemployment

From (10), the outward multilateral resistance term has been shown to be a sufficient statistic for unemployment:

\[
  u_m = \exp \left( -\sum_i \alpha_{mn} W_n \right) = \exp(-O_m).
\]

Unemployment in location \( m \) is lower when workers there are better connected through higher home bias adjusted search intensities, particularly in high expected utility locations since \( O_m \) is a cross-product.

To separately address roles of mobility friction due to home bias preference \( (b_n) \), and other search intensity controls \( (a_{mn}) \), we can rewrite the unemployment rate as:

\[
  u_m = \exp \left( -\sum_i \alpha_{mn} W_n \right) = \exp \left( -\sum_i a_{mn} W_{mn} \right).
\]

where

\[
  W_{mn} \equiv W_n \left( 1 - I_{mn}b_n/(1 + b_n) \right),
\]

where to recall, \( I_{mn} = 1 \) if \( m \neq n \) and zero otherwise.

Since \( a_{mn} \) is a function of a list of potential determinants of search cost (e.g. social connections, distance, and other economic ties between \( m \) and \( n_i \)), consider a first order approximation of \( a_{mn} \):

\[
  a_{mn} = \sum_i \beta_i x_{mn}^i,
\]

In relation to the literature, Grogger and Hanson (2011) in their analysis of international migration, for example, found that the bilateral migration cost implied by observed difference in income per capita across countries is very large. The implied bilateral migration cost includes any effects associated with home bias, as \( \alpha_{ij} \) does.
where \( x_{mn}^i, i = 1, \ldots, I \) is a list of bilateral search intensity controls such as our friendship network indicator, and \( \beta_i \) the vector of corresponding marginal effects on search intensity. Now,

\[
\ln(u_m) = -\sum_n \sum_i \beta_i x_{mn}^i W_{mn}
\]

\[
= -\sum_i \beta_i [\bar{x}_m^i \bar{W}_m + Cor(x_{mn}^i, W_{mn})]
\]

where \( \bar{W}_m \equiv \sum_n W_{mn} \) is the overall expected utility accounting for home bias in locational preferences, \( \bar{x}_m^i = \sum_n x_{mn}^i / M \) the average gains in search intensity attributable to each bilateral search intensity control \( x_{mn}^i \), and \( Cor(x_{mn}^i, W_{mn}) = \sum_n (x_{mn}^i - \bar{x}_m^i)(W_{mn} - \bar{W}_m) \) is the corresponding correlation term. In our empirical application to follow, we will introduce additional flexibility to (21) by adding a baseline term \( \beta_o \), in order to capture state-level / regional differences in baseline search intensity.

The log of unemployment is thus given by two vectors. A vector of level effects of search intensity controls weighted by average expected destination utility through \( \bar{x}_m^i \) and \( \bar{W}_m \), and an interaction effect through the correlation term \( Cor(x_{mn}^i, W_{mn}) \). Together, the level and the interaction effects provide a basis for assessing the inter-regional roots of unemployment, which simultaneously account for home bias adjusted search intensity, and the desirability of the list of migration destinations.

4 Data and Methodology

We use three main data sets. We collect data on bilateral county population flows from the 2014-2018 American Community Surveys. The dataset contains yearly counts of individuals who have moved between counties. One disadvantage of the dataset is that migration is censored for small counties to avoid privacy concerns, and therefore we only have 425 unique counties.

Our measurement of bilateral friendship networks is based on the Social Connectedness Index (SCI) from Bailey et al. (2018). As mentioned earlier, the Social Connectedness Index is constructed using the total number of Facebook friendship links between individuals located in a pair of counties: for every county pair \( m \) and \( n \)

\[
SCI_{mn} = \frac{\text{Facebook Connections}_{mn}}{\text{Pop}_m \text{Pop}_n}, \tag{21}
\]

where Facebook Connections is the number of Facebook friendship links and POP is county population.\(^{21}\) Bailey et al. (2018) normalize the index such that the maximum value of the index is

\[^{21}\text{Note: this variable is slightly different from the one used in Bailey et al. (2018), where friendship links are...} \]
1,000,000 (Los Angeles, CA).

We also collect information on historical (1940) county ethnic origin, occupational, and industrial compositions from the public Census microdata. We use this dataset to construct historical social and economic ties. In particular, we define a “distance” measure between counties as

\[
d(m, n) = \sum_{k \in K} (s_{km} - s_{kn})^2,
\]

where \( K \) is the set of all available U.S. ethnic origin, occupation, or industry groups in 1940 and \( s_{kl} \) is the share of a specific group \( k \) in county \( l = m, n \). We use the 1940 sample because it is the most recent full-count census available, which provides detailed and complete coverage of county compositions across multiple dimensions. The ethnic origin distance measure is based on birthplace. Persons born in the 50 U.S. states are assigned their states of birth. To eliminate small cells, non-U.S. birthplaces are categorized into larger regions: U.S. territories, Canada, Mexico, other North America, South America, Central America, Western Europe, Central/Eastern Europe, Southern Europe, Northern Europe, Russian empire, East Asia, Southeast Asia, Southwest Asia, Middle East, and Oceania. Persons born at sea or with an unidentifiable birthplace are dropped from the analysis.

We estimate the following gravity equation:

\[
\ln \left( \frac{\mu_{mn}}{\mu_{mm}} \right) = \beta \ln \left( \frac{a_{mn}}{a_{mm}} \right) + D_n + T_m + \epsilon_{mn} \tag{23}
\]

where \( \mu_{mn} \) is the ratio of the number of migrants from \( m \) to \( n \) to the number of non-migrations from \( m \). We refer to \( a_{mn} \) as the search intensity ratio. We measure \( a_{mn} \) the arrival rate of jobs in destination \( n \) for residents in \( m \) using the Facebook social connectedness index between counties \( m \) and \( n \). The ratio \( a_{mn}/a_{mm} \) is a linear combination of our list of bilateral search intensity controls, including the SCI ratio, distance, state borders, and other economic ties. The variables \( D_n \) and \( T_m \) are source and destination county fixed effects and absorb county-specific unobserved “push” and “pull” factors of migration, including mean county worker expected utility draws. Finally, \( \epsilon_{mn} \) is a function of source-destination-specific shocks unrelated to social connections that affect migration, some of which may be unobservable.

There are two sets of econometric issues associated with estimating a gravity equation like (18 or 19). First, the average county in the data has 53 observable outward migration links, out adjusted by the number of Facebook users instead of by county population.
of a possible 425. This feature is not uncommon in migration data (e.g. Beine et al. (2011), Beine et al. (2016)). Two solutions have been adopted so far. These include a two-step Heckman estimation requiring an instrument for the extensive margin selection equation. Another possibility is a count regression model via a Poisson pseudo maximum likelihood regression (Santos Silva and Tenreyro (2021)). In our case, while unobserved migration links may indeed be due to the true absence of migration, treating all unobserved links as zero migration in a selection equation will be inappropriate, since in many cases, migration may simply have been censored due to privacy concerns, rather than actual zeros. Meanwhile, a count regression approach does not work in our case either, since our main estimation equation (18 and 19) present ratios of labor flows, rather than number of migrants.

The concern associated with ignoring unobserved flows (either because of the log of zeros with true zero migration flows, or missing /omitted observations) is that the influence of distance, networks, and other migration cost or search intensity related variables will be underestimated if the migration outcomes of the most remote / isolated locations are omitted. By the same token, in our context, ignoring unobserved flows can mean that estimated destination fixed effects will be inflated, while origin fixed effects may be underestimated. Consequently, home bias – being the difference between origin and destination fixed effects from (20) – will likewise be underestimated. In what follows, we proceed with our intensive margin estimation with the important caveat that our estimated search intensity variables as well as home bias are lower bounds. The same approach is adopted in Bailey et al. (2018) in the context of migration, and Eaton and Kortum (2002) in the context of international trade, among others.

A second potential issue is estimation bias due to omitted variables. Specifically, in our preferred specifications, we control for those factors using bilateral geographic distance, state border effects (whether counties are located in the same state), and historical economic linkages, such as the bilateral distance in 1940 occupation and industry compositions. Even with the inclusion of these controls, estimating (23) using Ordinary Least Squares could result in a biased estimate of $\beta$ if some of the unobservables in $\varepsilon_{mn}$ are still correlated with social connections. For example, omitting contemporaneous industrial or trade linkages, which are likely to be positively correlated with both friendship links and migration flows, could result in an upward bias of $\beta$. Furthermore, as friendship networks are often formed in the workplace, our estimates of $\beta$ could also be confounded by reverse causality. In light of these potential biases, we also estimate $\beta$ using an instrumental variables approach.
The choice of instrument is informed by our theoretical model. The theory makes clear that identification of $\beta$ comes from finding a valid instrument that is correlated with cross-county friendship links and uncorrelated with contemporaneous economic ties, conditional on the fixed effects. To this end, we construct a bilateral ethnic distance between county pairs based on residents’ birthplace.

Figure 2 plots both the relationships between log SCI ratio and log geographic distance (Panel A) and the relationship between log SCI ratio and log ethnic distance, adjusting for bilateral geographic distance (Panel B). There is a clear, strong, and negative relationship between geographic distance and log SCI ratio, which is consistent with the findings of prior work (Bailey et al., 2018). Additionally, counties that are historically more ethnically distant continue to exhibit less social integration today (more than 70 years later) even after conditional on geographic distance. These relationships suggest that a history of prior ethnic networks can facilitate future migration flows as tighter social integration also facilitates information flows and a lower cost of migration (e.g. Chau (1997), Munshi (2003), Mckenzie and Rapoport (2007), Mayda (2010b)). This leads to path dependency in migration patterns over time.

The implicit assumption of using this instrument is that bilateral social connections are more persistent than alternative economic linkages. On this, our study follows the work of an extensive list of studies that use historic migration rates as an instrument for current migration.\textsuperscript{22} Still, a potential challenge to the use of historic ethnic distance as an instrument for current social connection is that other historic economic relationships, such as industry or occupational specialization patterns across counties, may have been correlated with historic ethnic distance, and in turn it is a history of economic ties, rather than of social integration that drove current migration. For these reasons, in Appendix A we perform an OLS regression with clustered standard errors at the county-pair level which shows that historical industry- or occupational-differences do a poor job in explaining the cross-county pair variations in historic ethnic distance. Upon controlling for distance and same-state status, the fit of the regression substantially improves, but the both estimates on historic industry and occupation distance are insignificant. Furthermore, in the main gravity estimation, we will also demonstrate that our results do not change with or without the inclusion of historic industry, occupation and geographical distances.

\textsuperscript{22}See for example McKenzie and Hildebrandt (2005), Mckenzie and Rapoport (2007), Woodruff and Zenteno (2007), and López Córdova (2018)).
Figure 2: Relationships Between SCI and Geographic and Ethnic Distances

Note: This figure presents a binscatter plot of the relationship between SCI and geographic and ethnic distance between US counties. SCI refers to Facebook social connectedness index. Geographic distance is the distance between county pairs. Ethnic distance is constructed according to equation (22) using historical (1940) county ethnic origin compositions from the public Census microdata. The ethnic origin distance measure is based on birthplace. Persons born in the 50 U.S. states are assigned their states of birth. To eliminate small cells, non-U.S. birthplaces are categorized into larger regions: U.S. territories, Canada, Mexico, other North America, South America, Central America, Western Europe, Central/Eastern Europe, Southern Europe, Northern Europe, Russian empire, East Asia, Southeast Asia, Southwest Asia, Middle East, and Oceania. Persons born at sea or with an unidentifiable birthplace are dropped from the analysis.
5 Results

In this section, we provide empirical estimates of the determinants of county-level bilateral migration, home bias as well as unemployment, using empirical specifications guided by the three applications of our model.

5.1 Relationships between Bilateral Population Flows and Log SCI Ratio

Table 1 reports both the OLS and 2SLS estimates of the effect of log SCI ratio on log outflow ratios. The OLS estimates in the first three columns show that the correlation between log SCI ratio and log outflow ratio is very close to one. This positive relationship holds even after controlling for geographic distance (Column 2) and historical bilateral economic linkages and whether counties are located in the same state (Column 3). Bailey et al. (2018) also find that the elasticity of population flow to friendship links is close to one.\(^{23}\) The coefficients on industrial and occupational linkages are very small, suggesting that after controlling for bilateral social connections, historical economic linkages through industrial and occupation compositions have little explanatory power.\(^{24}\)

Furthermore, we introduce a binary variable indicating that the origin-destination pair are in the same state, to account for border effects in migration decisions that may arise due to within-state ease of movement beyond that of distance. Interestingly, we find that once friendship networks and border effects are accounted for, the distance variable has the wrong sign. We note that in Bailey et al. (2018) as well, a similar sign and significance effect on the distance term is observed once social connectedness indicator is included in the estimation.

Because of the endogeneity concerns raised earlier, we re-estimate the specifications in Columns 1 to 3 using 2SLS and report the results in Columns 4 to 6. With the exception of Column 6, the 2SLS estimates are slightly smaller than the OLS estimates, confirming our suspicion that the OLS estimates are likely biased upwards.

Table 2 reports the estimates using population inflows as the outcome. Our model predicts that we should expect to see similar coefficients on the log SCI ratio, and this is what we find.

5.2 Home Bias

As discussed in (20), the level of home bias of any location can be inferred from our estimated fixed effects respectively for origin and destination counties. Here, we rely on the OLS estimates because

---

\(^{23}\)One distinction is that they use aggregate bilateral population lows—both inflows and outflows—whereas our model is in terms of population outflows.

\(^{24}\)Further note that the coefficient on bilateral occupation distance, though small, is in the opposite direction.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td><strong>ln_sci_ratio</strong></td>
<td>1.110***</td>
<td>1.141***</td>
<td>1.083***</td>
<td>1.150***</td>
<td>1.251***</td>
<td>1.007***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.030)</td>
<td>(0.052)</td>
</tr>
<tr>
<td><strong>ln_dist</strong></td>
<td>0.040**</td>
<td>0.043**</td>
<td>0.134***</td>
<td>-0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.031)</td>
<td>(0.039)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ln_cbp_norm2_diff</strong></td>
<td>-0.123***</td>
<td>-0.148***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.045)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ln_occ_norm2_diff</strong></td>
<td>-0.054***</td>
<td>-0.059***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>same_state</strong></td>
<td>0.235***</td>
<td>0.299***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>22281</td>
<td>22281</td>
<td>22281</td>
<td>22281</td>
<td>22281</td>
<td>22281</td>
</tr>
<tr>
<td><strong>Kleibergen-Paap</strong></td>
<td>1409.161</td>
<td>571.570</td>
<td>236.014</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table displays the relationship between U.S. county level population outflow ratio (bilateral population outflow / total non-movers at source, 2014-2018 average, ACS) and search intensity controls. Five search intensity controls are included: “SCI”, “Geodist”, “Indist”, “Occudist” and “Same-state” respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference, occupation composition difference defined in (22) and same-state status. Robust standard errors clustered at the origin and destination county levels in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1.
Table 2: Relationship between Population Inflow Ratio and Log SCI Ratio.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>2SLS</td>
<td>2SLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>ln_sci_ratio</td>
<td>1.122***</td>
<td>1.157***</td>
<td>1.090***</td>
<td>1.172***</td>
<td>1.295***</td>
<td>1.033***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.028)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>ln_dist</td>
<td>0.045**</td>
<td>0.048**</td>
<td>0.162***</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.031)</td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln_cbp_norm2_diff</td>
<td>-0.109**</td>
<td></td>
<td>-0.127***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td></td>
<td>(0.047)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln_occ_norm2_diff</td>
<td>-0.076***</td>
<td></td>
<td>-0.080***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td></td>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>same_state</td>
<td>0.271***</td>
<td></td>
<td>0.319***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td></td>
<td>(0.056)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>21701</td>
<td>21701</td>
<td>21701</td>
<td>21701</td>
<td>21701</td>
<td>21701</td>
</tr>
<tr>
<td>Kleibergen-Paap</td>
<td>1373.489</td>
<td>562.212</td>
<td>232.575</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table displays the relationship between U.S. county level population inflow ratio (bilateral population inflow / total non-movers at destination, 2014-2018 average, ACS) and search intensity controls. Five search intensity controls are included: “SCI”, “Geodist”, “Indist”, “Occudist” and “Same-state” respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference, occupation composition difference defined in (22) and same-state status. Robust standard errors clustered at the origin and destination county levels in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1.
they are more efficiently estimated.

Figure 4 plots the distribution of home bias estimated with and without SCI. The figure shows two right-skewed distributions, with home bias estimated without SCI having higher density around smaller values. The means and standard deviations of the estimated home bias with SCI are .06 and .77 and without SCI are .77 and 1.34. In other words, once differences in bilateral social connections are accounted for, we find that on average, the expected utility premium that individuals attach to staying put is 6%. Without accounting for barriers to migration due to the lack of social connections, the resulting estimated level of home bias is substantially higher at 77%.

In our model, home bias is a catch-all term, introduced to account for any factors that may lead to locational inertia. Given our estimated level of home bias at the county level, we can now further unpack its determinants. We collect a wide array of county-level correlates, including crime rates, religiosity, demographics, family structure, housing attributes (e.g., age), and industrial and occupational structures.\footnote{See Table 8 for a complete list of the variables.} To unpack the interpretation of the estimated home bias, we employ a Least Absolute Shrinkage and Selection Operator (LASSO) to identify significant correlates of our estimates. We use a cross-validation method and select the shrinkage parameter according minimum Bayesian information criterion. Table 3 reports the set of predictors selected by LASSO and their coefficients.

Using this approach, we confirm that our estimates of migration costs continue to be highly negatively correlated with county-level congestion forces such as commute time and population density as displayed in Figure ???. Interestingly, home bias is positively correlated with population clusters such as retirement communities. In particular, the share of population living alone, the share of grandparents caring for grandchildren, the share of population with retirement income, the share of population with social security and maximum January temperature, are living arrangements, sources of income, and other preference markers of older-aged communities. Indeed, mobile communities are evidently more youthful, as home bias is negatively correlated with the share of working age population between 20 and 54. We also see some indication of credit constraints as a contributor to home bias, as the percent of individuals on food stamp is positively associated with home bias. Interestingly, housing figures prominently here as well, where the share of relatively newly built housing (between 1990 and 1999, between 2000 and 2009, and 2010 or later) is positively correlated with home bias. Finally, we also detect suggestive evidence that workers are attached to locations that specialize in particular lines of careers / work. For example, locations with plentiful
supply of flexible and low entry cost employment (e.g. retail trade) draw residents to stay, as do locations that have higher fractions of agricultural, as well as farming and fishing employment, for example.

Table 3: Predictors of Home Bias

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Predictor</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Log avg. commute time</td>
<td>-0.070</td>
</tr>
<tr>
<td>2</td>
<td>% living alone</td>
<td>0.068</td>
</tr>
<tr>
<td>3</td>
<td>% males</td>
<td>0.068</td>
</tr>
<tr>
<td>4</td>
<td>Log population density</td>
<td>-0.066</td>
</tr>
<tr>
<td>5</td>
<td>% on food stamp</td>
<td>0.060</td>
</tr>
<tr>
<td>6</td>
<td>Maximum January temperature</td>
<td>0.054</td>
</tr>
<tr>
<td>7</td>
<td>% working in retail trade</td>
<td>0.045</td>
</tr>
<tr>
<td>8</td>
<td>% housing built between 1990 and 1999</td>
<td>0.044</td>
</tr>
<tr>
<td>9</td>
<td>% divorced</td>
<td>0.037</td>
</tr>
<tr>
<td>10</td>
<td>% household with retirement income</td>
<td>0.033</td>
</tr>
<tr>
<td>11</td>
<td>% population aged between 20 and 54</td>
<td>-0.031</td>
</tr>
<tr>
<td>12</td>
<td>% working in agriculture</td>
<td>0.027</td>
</tr>
<tr>
<td>13</td>
<td>% working in wholesale trade</td>
<td>-0.024</td>
</tr>
<tr>
<td>14</td>
<td>% grandparents caring for grandchildren</td>
<td>0.023</td>
</tr>
<tr>
<td>15</td>
<td>% working in construction occupations</td>
<td>0.023</td>
</tr>
<tr>
<td>16</td>
<td>% working in education, health, and social services</td>
<td>-0.016</td>
</tr>
<tr>
<td>17</td>
<td>% household with social security</td>
<td>0.014</td>
</tr>
<tr>
<td>18</td>
<td>% population that are Hispanic</td>
<td>0.013</td>
</tr>
<tr>
<td>19</td>
<td>% housing built 2010 or later</td>
<td>0.011</td>
</tr>
<tr>
<td>20</td>
<td>% housing built between 2000 and 2009</td>
<td>0.007</td>
</tr>
<tr>
<td>21</td>
<td>% working in farm and fishing occupations</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: This table lists the top 20 contributors to county-level differences in estimated home bias and the corresponding coefficients. The analysis is based on a Least Absolute Shrinkage and Selection Operator (LASSO) estimator, and a cross-validation method that selects the shrinkage parameter according to minimum Bayesian information criterion. The full list of variables included in this exercise can be found in Appendix 7.

These lessons are a useful reminder that population demographic, and the skill requirement of jobs are spatially heterogeneous to begin with. Once differences in social connectedness are accounted for, our estimated home bias, at least in part, is a reflection of these inherent population and job differences across space. The fact that home bias tend to be related to demographic characteristics and the skill-specific supply of jobs would, perhaps ironically, testify to the fact workers are indeed choosing to be immobile rationally depending on her age and skill. This being the case, the lack of social connectedness indicators specific to age group and skills prevent us from
digging deeper into these issues, however, as we are not able to cater to each population subgroup and measure their respective levels of home bias, knowing now what a big difference ignoring social connectedness can make (6% as opposed to 77%).

With these important caveats in mind, in Table 4 we rank counties according to the estimated level of home bias. Among the counties with the highest home bias are several well-known dense urban centers, such as New York, San Francisco, Washington D.C. and its surrounding Virginia counties. The list also contains smaller counties, suggesting that the estimated migration costs are not simply picking up county size. By contrast, counties with the lowest level of home bias tend to be much smaller and more likely to be in the South and the West. One interpretation of the patterns exhibited by the two lists is that there is more variation in preferences for small counties’ amenities. In the Appendix, we also report county rankings based on sending county fixed effects, which capture average utility draws at the destinations. The list indicates that the highest expected utility counties are primarily located in Mid-Atlantic regions (e.g., VA and GA) and the lowest expected utility counties are located in the Midwest and the South.

Table 4: County Rankings: Top 10 and Bottom 10 Degree of Home Bias

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Highest Migration Home Bias</th>
<th>Ranking</th>
<th>Lowest Home Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ascension Parish, LA</td>
<td>416</td>
<td>Calhoun County, AL</td>
</tr>
<tr>
<td>2</td>
<td>Clayton County, GA</td>
<td>417</td>
<td>Santa Rosa County, FL</td>
</tr>
<tr>
<td>3</td>
<td>New York County, NY</td>
<td>418</td>
<td>Tippecanoe County, IN</td>
</tr>
<tr>
<td>4</td>
<td>Nassau County, NY</td>
<td>419</td>
<td>Nueces County, TX</td>
</tr>
<tr>
<td>5</td>
<td>Los Angeles County, CA</td>
<td>420</td>
<td>Douglas County, KS</td>
</tr>
<tr>
<td>6</td>
<td>Richmond County, NY</td>
<td>421</td>
<td>Lowndes County, GA</td>
</tr>
<tr>
<td>7</td>
<td>Prince George’s County, MD</td>
<td>422</td>
<td>Escambia County, FL</td>
</tr>
<tr>
<td>8</td>
<td>Gwinnett County, GA</td>
<td>423</td>
<td>Sussex County, DE</td>
</tr>
<tr>
<td>9</td>
<td>Queens County, NY</td>
<td>424</td>
<td>Kern County, CA</td>
</tr>
<tr>
<td>10</td>
<td>Schenectady County, NY</td>
<td>425</td>
<td>El Paso County, TX</td>
</tr>
</tbody>
</table>

Note: This table lists the top 10 and bottom 10 counties ranked based on estimated home bias. Home bias is calculated as a difference between the origin and destination fixed effects of the same location obtained from the outflow gravity regression displayed in 5.

5.3 Unemployment

Our model predicts that a county that is more socially integrated with high expected utility destination locations will have lower unemployment in equilibrium. To test this prediction, we estimate

\[ \text{Unemployment}_t = \beta_0 + \beta_1 \text{Home Bias}_t + \epsilon_t \]

26 Similar rankings of countries’ trade costs are often used in the trade literature (e.g., Eaton and Kortum, 2002).
the following equation using OLS:

$$\ln(u_m) = \alpha + \sum_i \beta_{li} \bar{x}_i^m \bar{W}_m + \sum_i \beta_{ci} \text{Covar}(x^i_{mn}, W_{mn}) + \phi_m + \varepsilon_m,$$

(24)

where $u_m$ is county unemployment rate. $\bar{x}_i^m = \sum_i x_{mn}/M_m$ is our proxy for the level effect of search intensity control $i$, and $\bar{W}_m = \sum_n W_n \equiv \sum_n W_n \left(1 - I_{mn}b_n/(1 + b_n)\right)$ is the overall expected utility of the destinations of $m$, adjusted by our home bias estimates $b_n$. We include as possible search intensity controls social connectedness proxied by Facebook friendship data ($SCI$), geographical distance ($Geodist$), historic industry distance ($Inddist$) and historic occupational distance ($Occudist$). $\text{Covar}(x, y)$ denotes the sample covariance between $x$ and $y$, $\phi_m$ are either region or state fixed effects. The coefficients $\beta_{li}$ and $\beta_{ci}$ respectively capture the level and interaction effect of the corresponding search intensity control. Our theory informs us that $\beta_{li}$ and $\beta_{ci}$ are negative. We note that while we include geographic, historic industry and occupational differences, our preferred specification includes simply the SCI, for in Table (1), geographic distance is either statistically insignificant or of the wrong sign, and the impact of historic industry and occupational distance are extremely small.

Table 5 presents the results.\textsuperscript{27} To make the coefficients easier to interpret, we report the effects in terms of a one standard deviation increase in the regressor of interest. Robust standard errors are clustered at the state level. Consistent with the model, we find strong and negative relationships between the unemployment rate and social connectedness. The level and interaction effects are both negative, consistent with the predictions of our model. Interestingly, we find that once we control for social connectedness, the level and interaction effect of distance take on the wrong sign, while the corresponding social connectedness terms remain more or less the same.

\textsuperscript{27}The outcomes are log unemployment rate multiplied by 100 and thus the coefficients can be interpreted as percent changes.
<table>
<thead>
<tr>
<th></th>
<th>(1) ln_unempl_rate</th>
<th>(2) ln_unempl_rate</th>
<th>(3) ln_unempl_rate</th>
<th>(4) ln_unempl_rate</th>
<th>(5) ln_unempl_rate</th>
<th>(6) ln_unempl_rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Level Effect (SCI)</td>
<td>-5.223***</td>
<td>-4.883</td>
<td>-6.150*</td>
<td>-5.650***</td>
<td>-2.583</td>
<td>-4.130</td>
</tr>
<tr>
<td></td>
<td>(1.574)</td>
<td>(3.561)</td>
<td>(3.371)</td>
<td>(1.977)</td>
<td>(2.760)</td>
<td>(2.678)</td>
</tr>
<tr>
<td></td>
<td>(1.774)</td>
<td>(2.372)</td>
<td>(2.717)</td>
<td>(1.843)</td>
<td>(2.482)</td>
<td>(2.395)</td>
</tr>
<tr>
<td>Std. Level Effect (Geodist)</td>
<td>-22.184***</td>
<td>-18.756***</td>
<td>-20.411***</td>
<td>-15.446***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.394)</td>
<td>(4.213)</td>
<td>(5.063)</td>
<td>(4.028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Interaction Effect (Geodist)</td>
<td>-16.916***</td>
<td>-15.454***</td>
<td>-12.356***</td>
<td>-11.765***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.816)</td>
<td>(3.398)</td>
<td>(4.056)</td>
<td>(4.922)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Level Effect (Inddist)</td>
<td>9.014*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Level Effect (Occudist)</td>
<td>-10.316*</td>
<td></td>
<td></td>
<td>-13.481***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.438)</td>
<td></td>
<td></td>
<td>(4.671)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Interaction Effect (Inddist)</td>
<td>5.445</td>
<td></td>
<td></td>
<td>7.238**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.521)</td>
<td></td>
<td></td>
<td>(3.456)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Interaction Effect (Occudist)</td>
<td>-7.607***</td>
<td></td>
<td></td>
<td>-8.249***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.742)</td>
<td></td>
<td></td>
<td>(2.824)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>425</td>
<td>425</td>
<td>425</td>
<td>421</td>
<td>421</td>
<td>421</td>
</tr>
<tr>
<td>Fixed effect</td>
<td>Region</td>
<td>Region</td>
<td>Region</td>
<td>State</td>
<td>State</td>
<td>State</td>
</tr>
</tbody>
</table>

Note: This table displays the relationship between U.S. county-level log unemployment rates (2014-2018 average, ACS) and outward multilateral migration resistance. Outward multilateral migration resistance, a function of destination expected utility weighted bilateral search intensity has three parts (21) (i) a level effect \( x = SCI, Geodist, Inddist, Occudist \) – the product of the average level of each search intensity control, the average expected utility of destination expected utility (the estimated destination fixed effects from Table ??) and \( M \); (ii) an interaction effect \( x = SCI, Geodist, Inddist, Occudist \) – the product of the covariance between each bilateral search intensity control and the expected utility of the corresponding destination \( Cov(x_{mn}, W_n) \) and \( M \), and (iii) a state level / regional level baseline. “SCI”, “Geodist”, “Inddist” and “Occudist” respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference and occupation composition difference defined in (22). Standardized coefficients are presented. Robust standard errors clustered at the state-level in parentheses; *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \).
The key takeaway here is that local unemployment rates do indeed appear have inter-regional roots. In particular, $u_m$ is significantly negatively correlated with the extent to which individuals in $m$ are connected with others elsewhere (level effect), where our measure of integration $\alpha_{mn}$ accounts for home bias. But perhaps equally importantly, local unemployment rates will be lower still if social linkages are of the right kind, namely, connecting individuals with those residing in desirable (high expected utility) locations. Importantly, expected utility itself in our model is an inter-regional concept, depending on average utility $w_n$, the availability of jobs, $v_n$, as well as the totality of social connections each destination has with other potential sending counties $\sum_k \alpha_{kn}N_k = J_n$.

6 Conclusion

In this paper, we develop a model of migration in the presence of friendship networks and job search frictions. The model delivers predictions about bilateral migration flows in a simple and tractable equation, in which bilateral migration and regional unemployment are simultaneously determined depending on the degree of integration between locations.

In addition to developing the micro-foundations of a migration gravity equation with unemployment, as a theory of migration, our model provides novel insights on the persistence in low mobility rates in the U.S. despite stark spatial contrasts in wages and unemployment rate, and advancement in communication technologies. Our model also offers clues on how to unpack migration friction into its two components: home bias and search friction, respectively capturing barriers to migration that are driven by preferences and path dependency, as well as the lack of access to markets. As a model of unemployment, we show how information on home bias and search friction jointly allows us to furnish an unemployment sufficient statistic interpretation to the familiar multilateral migration resistance term.

We verify the predictions of the model using cross-sectional data on bilateral county friendship links from Facebook. Because county friendship networks are endogenous, we develop an original instrumental variable approach based on a similarity index of counties’ ethnic-composition. We find that the predictions of our model in terms of migration flows, as well as unemployment are well supported by the data. Furthermore, we find that home bias and network connections go hand in hand, not just as individual contributors to migration friction. Indeed, we find that attempts to estimate home bias can be seriously flawed unless detailed dyad-level job search connections are appropriately accounted for.
References

Albert, C. and J. Monras (2017). Immigrants’ residential choices and their consequences. CReAM Discussion Paper Series 1707, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London. 2, 8


Appendix A.

Table 6: (Non-)Drivers of Historic Ethnic Distance

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td>ln(_{cbp}_norm2_diff)</td>
<td>-0.054</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>ln(_{occ}_norm2_diff)</td>
<td>0.028</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>ln(_{dist})</td>
<td>0.018**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td>same(_{state})</td>
<td>-1.532***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>24459</td>
<td>24459</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.0007</td>
<td>0.7669</td>
</tr>
<tr>
<td>Mean dep. var.</td>
<td>-0.783</td>
<td>-0.783</td>
</tr>
</tbody>
</table>

Note: This table displays the relationship between historic ethnic distance and historic industry and occupation distance index defined in (22). Robust standard errors clustered at the origin and destination county levels in parentheses; *** \(p < 0.01\), ** \(p < 0.05\), * \(p < 0.1\).

Table 7: County Rankings: Top 10 and Bottom 10 Origin County Fixed Effect.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Highest Origin FE</th>
<th>Ranking</th>
<th>Lowest Origin FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Webb County, TX</td>
<td>416</td>
<td>Arlington County, VA</td>
</tr>
<tr>
<td>2</td>
<td>Wichita County, TX</td>
<td>417</td>
<td>Portage County, OH</td>
</tr>
<tr>
<td>3</td>
<td>Calhoun County, AL</td>
<td>418</td>
<td>Fort Bend County, TX</td>
</tr>
<tr>
<td>4</td>
<td>Muskegon County, MI</td>
<td>419</td>
<td>Comal County, TX</td>
</tr>
<tr>
<td>5</td>
<td>Peoria County, IL</td>
<td>420</td>
<td>Paulding County, GA</td>
</tr>
<tr>
<td>6</td>
<td>Allen County, OH</td>
<td>421</td>
<td>Johnston County, NC</td>
</tr>
<tr>
<td>7</td>
<td>Shelby County, TN</td>
<td>422</td>
<td>Clackamas County, OR</td>
</tr>
<tr>
<td>8</td>
<td>Ouachita Parish, LA</td>
<td>423</td>
<td>Seminole County, FL</td>
</tr>
<tr>
<td>9</td>
<td>Harrison County, MS</td>
<td>424</td>
<td>Wright County, MN</td>
</tr>
<tr>
<td>10</td>
<td>Ashtabula County, OH</td>
<td>425</td>
<td>Cobb County, GA</td>
</tr>
</tbody>
</table>

Note: This table lists the top 10 and bottom 10 counties ranked based on estimated origin fixed effects. Origin fixed effects are obtained from outflow gravity regression displayed in 1.
<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Variable List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commute</td>
<td>log avg. commute time, % commuters, % drive alone, % carpool, % take public transport</td>
</tr>
<tr>
<td>Demographics</td>
<td>% males, % black, % Hispanic, % foreign-born, % with at least Bachelor’s, log population density, % younger than 20, % aged between 20 and 54, % aged older than 54</td>
</tr>
<tr>
<td>Environment</td>
<td>heat days, July precipitation, January maximum temperature, and July max temperature</td>
</tr>
<tr>
<td>Industry</td>
<td>% agriculture, % construction, % manufacturing, % wholesale, % retail, % transportation, % information, % finance, insurance, and real estate, % public education, health, and social services, % recreational and entertainment, % other industries, and % public admin</td>
</tr>
<tr>
<td>Marriage</td>
<td>% living alone, % with children, % divorced, % grand parents caring for children</td>
</tr>
<tr>
<td>Occupation</td>
<td>% management and professional, % construction, % farm and fish, % sales and office, % service, % transportation and utility</td>
</tr>
<tr>
<td>Public assistance</td>
<td>% household on social security, % household with retirement income, % household with supplemental security income, % household with public assistance, % household with food stamp, and % percent household below poverty line</td>
</tr>
<tr>
<td>Religious</td>
<td>% Evangelical, % Catholic, and % Mainline Protestant</td>
</tr>
<tr>
<td>Social Capital</td>
<td>crime per capita, and republican vote share</td>
</tr>
</tbody>
</table>
Figure 3: Distribution of Average State-Level Home Bias $b_n$. 

[Map showing the distribution of average state-level home bias with color coding for different intervals of values.]
Figure 4: Distribution of Average State-Level Expected Utility ($W_n$).