DISCUSSION PAPER SERIES

Migration Gravity, Networks, and Unemployment

Arnab K. Basu
Nancy H. Chau
Gary C. Lin

DISCUSSION PAPER SERIES

Migration Gravity, Networks, and Unemployment

Arnab K. Basu
Cornell University and IZA

Nancy H. Chau
Cornell University and IZA

Gary C. Lin
Johns Hopkins University

ABSTRACT

Migration Gravity, Networks, and Unemployment*

We develop and estimate a theory-consistent gravity model 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 novel concepts (e.g. home bias). As a model of unemployment, a migration gravity approach uncovers hitherto under-appreciated interregional 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.

Corresponding author:
Nancy H. Chau
Charles H. Dyson School of Applied Economics and Management
Cornell University
Ithaca, NY 14853
USA

E-mail: hyc3@cornell.edu

1 Introduction

Can the search friction origins of unemployment shed light on observed patterns of labor mobility? What can barriers to migration teach us about the causes of unemployment? While there is strong evidence that aggregate labor mobility respond to unemployment-driven spatial disparities across regions (e.g., DaVanzo, 1978; Greenwood, 1985; Blanchard and Katz, 1992; Dorn and Zweimüller, 2021) and vice versa (e.g. Card, 1990; Dustmann et al., 2005), our understanding is notably scant regarding the co-determination of networks of labor mobility patterns and local unemployment incidence. To these ends, the workhorse gravity model of migration does not as yet allow for endogenous unemployment (e.g., Anderson, 2011; Grogger and Hanson, 2011; Bertoli and Fernández-Huertas Moraga, 2013; Tombe and Zhu, 2019), and theories on the inter-regional roots of joblessness have just begun investigating the role of commuting costs and distance (e.g., Manning and Petrongolo, 2017; Monte et al., 2018; Caliendo et al., 2019; Notowidigdo, 2020). In this paper, we incorporate lessons from a vibrant literature of networks as a predictor of job search success (Chau, 1997; Munshi, 2003; McKenzie and Rapoport, 2007; Bertoli, 2010; Beine et al., 2011) to formally connect the gravity of migration with local unemployment incidence. The resulting setup of multilateral mobility and unemployment sheds new light on longstanding puzzles (e.g. low mobility despite better communication), and provides a setting in which to accommodate more novel concepts (e.g. home bias in migration decisions).

Specifically, we develop and estimate a theory-consistent gravity model for inter-regional migration flows with co-determined 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. In this setting, 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. 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.

In conceptualizing individual locational preference, we allow for the possibility of home bias – the spatial expected utility premium a native resident attaches to her own origin relative to a

---

[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.

new immigrant (Faini and Venturini, 2001). This may be due to status quo bias (Samuelson and Zeckhauser, 1988), social capital (Borjas, 1992), information asymmetries between residents and new migrants (Bryan et al., 2014), sunk investments (Albert and Monras, 2017; Bilal and Rossi-Hansberg, 2018), identity and preference (Djajic and Milbourne, 1988; Heise and Porzio, 2019) or relationships with ethnic enclaves (Albert and Monras, 2017). In so doing, our model offers a theory-based empirical strategy to estimate home bias – and an opportunity to better understand its possible correlates – that may contribute to impeding or encouraging mobility even after controlling for migration dyad level factors such as migration cost, or network connection.

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) origin- and destination-specific expected utilities of employment as push and pull migration forces, and (iii) an analogue of the familiar multilateral migration resistance term (Bertoli and Fernández-Huertas Moraga, 2013). 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 as pull-push factors in the standard migration gravity model is an expected utility term adjusted to account for home bias. Specifically, since employment is not assured in our model where unemployment is co-determined with migration, the expected utility term allows for spatial disparities in the likelihoods of finding a job across all possible destinations, where the effective size of the job seeker pool per vacancy in turn depends on the network of links connecting any given destination with all other sending locations. For example, a high-wage high-amenity location may nonetheless be less desirable if it is strongly connected with other sending locations resulting in a intensely competitive job search due effectively to a large job seeker pool.

Finally, by incorporating endogenous search friction induced unemployment in our migration gravity setting, we find that the (outward) multilateral migration resistance term in our model turns out to be a sufficient statistic for the local unemployment rate. This finding is novel yet 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 other key insights from the standard gravity model that are closely related to these three blocks. In particular, we write down a structural migration gravity equation following by now standard practice to retrieve our analogue of the iconic population product term (Anderson, 2011). We find that the origin-destination population product term indeed requires revision, to respectively capture (i) sending location employment inclusive of emigrants, and (ii) the destination employment inclusive of all immigrants. The former is directly proportional to 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 U.S. 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 comes 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

---

[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
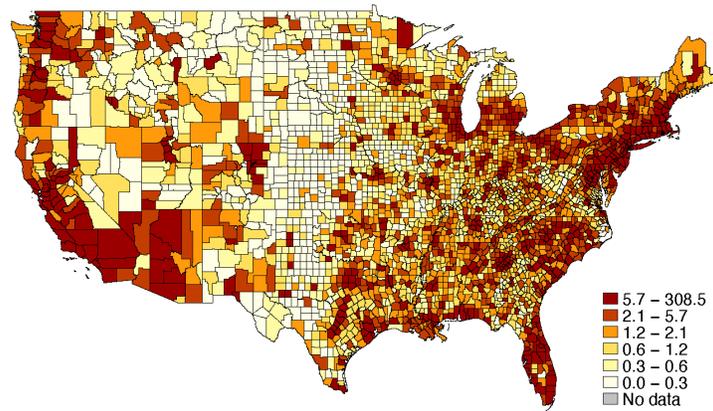
friendship links between counties in the contiguous 48 U.S. states. To our knowledge, with its 239 million users, the Facebook dataset is the only dataset that provides a comprehensive coverage of friendship networks at the national level in the United States. Figure 1 plots counties' average social 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 historic population composition to subsequent migration movements are commonly employed in empirical work (e.g. 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 implications of these empirical findings in three applications. First, we distinguish between inflow migration gravity and outflow migration gravity.[3] In empirical estimation on

---

[3]Respectively, these refer to the number of bilateral immigrants as a share of the total number of destination

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. 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 home bias in location preferences as the difference between origin and destination fixed effects of the same location.[5] 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, Lousiana, 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 systematically correlated with features of a local economy that are often overlooked in migration gravity research, including congestion (e.g. commute time and population density), employment composition (e.g. the share of information industry jobs), demographic factors (e.g. the share of population that are living alone, share of population receiving retirement income), and climate factors (e.g. temperature in winter and summer months).[6] These findings are consistent with Boustan (2013) and Albouy and Stuart (2014), for example, which demonstrate the important

---

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 immigration are also universally and simultaneously positive from 615 to over 32,000 and from 345 to over 23,000 respectively.

[5]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.

[6]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.

role that public goods and amenities play 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 to capture the collective unemployment impact of a network of job search connections summarized by the outward multilateral resistance term. Guided by our model, we construct an interaction term (a covariance term between bilateral network connection and destination expected utilities) to assess the role of the correlation between network connections and destination expected utility on unemployment, after controlling for level effects (the average level of bilateral network connection, the average destination expected utility proxied by destination fixed effects from the migration gravity model), as well as state or region fixed effects. We find consistently a negative and robust relationship between the interaction term and local unemployment, consistent with our theory that being better connected to high expected utility locations, conditional on home bias, lowers unemployment. By contrast, we show that the average social connectedness of a location is a poor predictor of local unemployment.[7]

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, 2010; Beine et al., 2011; Bertoli and Fernández-Huertas Moraga, 2012; Beine and Parsons, 2015). The role of friendship connections on 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.[8] 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

---

[7]Also, consistent with the findings in Bailey et al. (2018), we find that once network effects and state boundaries are accounted for, any corresponding level and interaction effects of geographic distance on unemployment are either not detectable, or of the wrong sign.

[8]There is an extensive literature linking networks and migration, which network operates through providing job information (Borjas, 1992; Munshi, 2003), offering material support (Mayda, 2010; Munshi, 2014), and reducing migration cost (Chau, 1997; McKenzie and Rapoport, 2007).

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 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.

Mobility in our setup is driven by individual costs and benefits considerations conditional on origin- and destination-specific characteristics and individual preferences. Clearly alternative drivers abound, including mobility driven by a desire to be better positioned to access to jobs (Harris and Todaro, 1970), individual differences in locus of control (Caliendo et al., 2019), exodus to avoid environmental stress factors and congestion forces (Feng et al., 2012; Cattaneo and Peri, 2016), family reunification (Basu et al., 2022), return migration (Dustmann, 2003), retirement relocation (King et al., 2021), training and education Dustmann and Glitz (2011), and other long term dynamic migration motivations (Artuç et al., 2010; Caliendo et al., 2019) to name just a few. Our decision to maintain focus on static labor market considerations allows us to make component-by-component comparison with a long history of other models of migration gravity in the static context but without search friction induced unemployment or home-bias (e.g. Ahlfeldt et al., 2015; Morten and Oliveira, 2016; Amior and Manning, 2018; Monte et al., 2018; Tombe and Zhu, 2019). In addition, as we will demonstrate, some of these alternative drivers, such as climate, congestion, proximity to family, can be seen as embedded as correlates of our home bias estimates, after controlling for the push-pull forces of gravity.

The rest of this paper is organized as follows. In section 2, we spell out the multi-location search model, derive closed form solution for migration gravity, and make connection with structural migration gravity. We then provide in Section 3 the theoretical underpinnings for the three empirical applications that showcase the different ways in which our model contributes to understanding the role of network and search friction in migration decisions and unemployment incidence. Section 4 presents our empirical application, in which we (i) evaluate the 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 $n = 1, ..., M$. 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 $z_n = 0, 1, 2, ...$ offers is given by a Poisson distribution with parameter $\lambda_{mn} \geq 0$:

$$\Pr(z_n; \ \lambda_{mn}) = \frac{exp\left(-\lambda_{mn}\right)\left(\lambda_{mn}\right)^{z_n}}{z_n!}.$$

The job arrival rate λ_{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:[9]

$$\lambda_{mn} = \frac{a_{mn}v_n}{\sum_{k=1}^{M} a_{kn}N_k} \equiv \frac{a_{mn}v_n}{J_n}. \tag{1}$$

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.[10] All else equal an increase in a_{mn} raises job arrival λ_{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.[11] The probability distribution of this match-specific utility in location n, ω, 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 $F_{mn}(\omega) = F_n(\omega, 1 + b_n)$,

---

[9]The specification in (1) satisfies adding-up, namely total job arrivals in all M locations add up to the number of vacancies since:

$$\sum_{k=1}^{M} \lambda_{kn} N_k = \sum_{k=1}^{M} a_{kn} N_k \frac{v_n}{J_n} = v_n.$$

[10]We state λ_{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).

[11]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).

$m \neq n$, where we assume the following first order stochastic ordering:

$$F_n(\omega, 1 + b_n) \geq F_n(\omega, 1) \tag{2}$$

whenever $b_n \geq 0$. Put another way, positive (weakly negative) home bias in migration preference exists if and only if $b_n \geq (<)0$.[12]

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

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

otherwise

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

where $w_n \in [0, 1]$ and $w_n/(1 + b_n)$ are shift parameters, while ϵ is a shape parameter. The expected values of ω 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). b_n can also be equivalently interpreted as the expected utility discount that a non-native resident applies to moving to n. The sources of home biases are many, due for example to status quo bias (Samuelson and Zeckhauser, 1988), prior investment in social capital and connectedness to local friendship networks (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).[14]

At each destination n, the probability distribution of the maximal utility sampled by a worker

---

[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."

[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.

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

from m seeking a job in destination n is:

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

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

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:

$$\begin{aligned} p_{mn}(\omega) &= exp\left[-\lambda_{mn}(1-F_{mn}(\omega))\right] \\ &= \exp\left[-\lambda_{mn}w_n(1+b_n)^{-\mathbb{I}_{mn}}(1+\epsilon\omega)^{-1/\epsilon}\right] \end{aligned} \tag{6}$$

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 λ_{mn}, and $w_n(1+b_n)^{-\mathbb{I}_{mn}}$. Higher search intensity through better network connection, or a higher λ_{mn}, and a higher expected utility in n, through $w_n(1+b_n)^{-\mathbb{I}_{mn}}$, 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 μ_{mn} as the probability that a worker from m finds that the best utility draw in n, ω_{mn}, to be more appealing that any other one of the $M-1$ locations's best offers, ω_{mk}, $k \neq n$. Thus

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

Let α_{mn} denote the home-bias adjusted search intensity

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

---

[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 $\lambda_{mn}w_n(1+b_n)^{-\mathbb{I}_{mn}}$ 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

$$W_n = \frac{w_n v_n}{J_n} = \frac{w_n v_n}{\sum_k a_{kn} N_k}. \tag{8}$$

Now, by the law of large numbers, μ_{mn} represents the fraction of the workers in m who prefers location n to any of the other $M - 1$ locations.[16]

$$
\begin{aligned}
\mu_{mn} &= \int_0^\infty \prod_{k \neq n} p_{mk}(\omega) dp_{mn}(\omega) \\
&= \left(\frac{\alpha_{mn} W_n}{\sum_{i=1}^M \alpha_{mi} W_i} \right) \left(1 - exp \left[-\sum_{i=1}^M \alpha_{mi} W_i \right] \right).
\end{aligned} \tag{9}
$$

The expression

$$O_m \equiv \sum_{i=1}^N \alpha_{mi} W_i$$

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 α_{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,

$$\lambda_{mn} w_n = \frac{a_{mn} v_n}{\sum_k a_{kn} N_n} (1 - b_n \mathbb{I}_{mn}/(1 + b_n)) w_n = \alpha_{mn} W_n$$

by definition of λ_{mn} in (1), α_{mn} in (7), and W_n in (8). Thus

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

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 $t_{mn} = 1/\alpha_{mn}$. 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:

$$u_m = \exp(-O_m). \tag{10}$$

(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, μ_{mn}, depends on (i) bilateral home bias adjusted search intensities α_{mn}, (ii) destination expected utility W_n, and (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]

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

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

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

and symmetrically, outward multilateral can be expressed as

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

Thus, total migration between two locations depends on (i) bilateral home bias adjusted search intensity α_{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 relegated to Appendix A.

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:

$$OI = \alpha$$

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

$$O_m = \sum_n \frac{\gamma \alpha_{mn} v_n w_n}{\sum_k \gamma \alpha_{kn} N_k} = \sum_n \frac{\alpha_{mn} v_n w_n}{\sum_k \alpha_{kn} N_k}$$

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

$$
\begin{aligned}
\mu_{mn} &= \int_0^\infty \prod_{k \neq n} p_{mk}(\omega) dp_{mn}(\omega) \\
&= \left(\frac{\alpha_{mn} v_n w_n / J_n}{\sum_{i=1}^M \alpha_{mi} v_i w_i / J_i} \right) (1 - u_m).
\end{aligned}
$$

is likewise invariant to equi-proportionate increases in α_{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 the ways in which changes in search intensity impact unemployment, through its influence on mobility. 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, α_{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 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 W_n/W_m, 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:

$$\frac{\mu_{mn}}{\mu_{nn}} = \left(\frac{a_{mn}}{a_{nn}}\right)\left(\frac{(1-u_m)\ln(1/u_m)}{(1-u_n)\ln(1/u_n)}\right)\left(\frac{1}{1+b_n}\right).$$

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 $[(1-u_m)\ln(1/u_m)]/[(1-u_n)\ln(1/u_n)]$, and (iii) home bias at destination n.

$$\ln \mu_{mn} - \ln \mu_{nn} = \ln a_{mn} - \ln a_{nn} + t_m - d_n, \tag{19}$$

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 μ_{mn}/μ_{mm} and inflow gravity (μ_{mn}/μ_{nn}) are *symmetrically* dependent on the relevant search intensities ratios, α_{mn}/α_{mm} and α_{mn}/α_{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),

$$T_i - D_i = \ln(1 + b_i) = d_i - t_i. \tag{20}$$

Importantly, therefore, the difference between the destination and origin fixed effects, when $m = n$, gives an estimate of the home bias 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 $\alpha_{ij} = a_{ij}(1 - \mathbb{I}_{ij}b_j/(1+b_j))$ to parameterize the overall barrier to migration between i and j. Quite intuitively, α_{ij}, and hence outward gravity from i to j is decreasing in b_j.[20] Our task here is to separately tease out b_i from α_{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 fine-tuned 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:

$$u_m = \exp\left(-\sum_i \alpha_{mi}W_n i\right) = \exp(-O_m).$$

Unemployment in location m is lower when workers there are better connected in high expected utility locations as O_m is a cross-product. Note that the expected utility of destination n is employment-adjusted, $W_n = w_n v_n / \sum_k a_{kn}N_k$, while search intensities are home-bias adjusted $\alpha_{mn} \equiv a_{mn}(1 - \frac{\mathbb{I}_{mn}b_n}{1+b_n})$.

To separately address the roles of home bias preference (b_n) and other search intensity controls (a_{mn}) on unemployment, 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

$$W_{mn} \equiv W_n\left(1 - \mathbb{I}_{mn}b_n/(1+b_n)\right),$$

where to recall, $\mathbb{I}_{mn} = 1$ 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,

---

[20]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 α_{ij} does.

distance, and other economic ties between m and n), consider a first order approximation of a_{mn}:

$$a_{mn} = \sum_i \beta_i x^i_{mn},$$

where x^i_{mn}, $i = 1, ..., I$ is a list of bilateral search intensity controls such as our friendship network indicator, and β_i the vector of corresponding marginal effects on search intensity. Now,

$$
\begin{aligned}
\ln(u_m) &= -\sum_n \sum_i \beta_i x^i_{mn} W_{mn} \\
&= -\sum_i \beta_i \left[Covar(x^i_{mn}, W_{mn}) + \bar{x^i_m} \bar{W}_m \right]
\end{aligned}
\tag{21}
$$

where $\bar{W}_m \equiv \sum_n W_{mn}$ is the overall expected utility accounting for home bias in locational preferences, $\bar{x^i_m} = \sum_n x^i_{mn}/M$ the average gains in search intensity attributable to each bilateral search intensity control x^i_{mn}, and $Covar(x^i_{mn}, W_{mn}) = \sum_n (x^i_{mn} - \bar{x}^i_m)(W_{mn} - \bar{W}_m)$ is the corresponding covariance term. The log of unemployment is determined by an interaction effect captured by the covariance term $Covar(x^i_{mn}, W_{mn})$, conditional on average search intensity and expected utility level effects $\bar{x}^i_m \bar{W}_m$ controls. Equation (21) provides a basis for assessing the inter-regional roots of unemployment, and brings together in one summary measure the network of home bias adjusted search intensity, as well as 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

$$SCI_{mn} = \frac{\text{Facebook Connections}_{mn}}{\text{Pop}_m \text{Pop}_n},
\tag{22}$$

where Facebook Connections is the number of Facebook friendship links and POP is county pop-

ulation.[21] Bailey et al. (2018) normalize the index such that the maximum value of the index is 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

$$d(m,n) = \sum_{k \in K} (s_{km} - s_{kn})^2, \tag{23}$$

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:

$$\ln\left(\frac{\mu_{mn}}{\mu_{mm}}\right) = \beta \ln\left(\frac{a_{mn}}{a_{mm}}\right) + D_n + T_m + \varepsilon_{mn} \tag{24}$$

where μ_{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}/a_{mm} as the search intensity ratio. The ratio a_{mn}/a_{mm} is a linear combination of our list of bilateral search intensity controls, including the Facebook Social Connectedness index ratio (SCI_{mn}/SCI_{mm}), 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, ε_{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

---

[21]Note: this variable is slightly different from the one used in Bailey et al. (2018), where friendship links are 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 (24) using ordinary least quares could result in a biased estimate of the true coefficient if some of the unobservables in ε_{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 upward bias. Furthermore, as friendship networks are often formed in the workplace, our estimates could also be confounded by reverse causality. In light of these potential biases, we also estimate the model using an instrumental variables approach.

The choice of instrument is informed by our theoretical model. The theory makes clear that identification 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 alleviate endogeneity concerns regarding using counties' contempraneous ethnic compositions, we approximate contemporaneous bilateral ethnic similarities between county pairs using bilateral historic ethnic distance. Similarly, to control for economic linkages' potential effect on cross-county migration, we include controls for historic bilateral industry and occupation linkages.

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 (2010), Blumenstock et al. (2019)). 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.[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.

---

[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
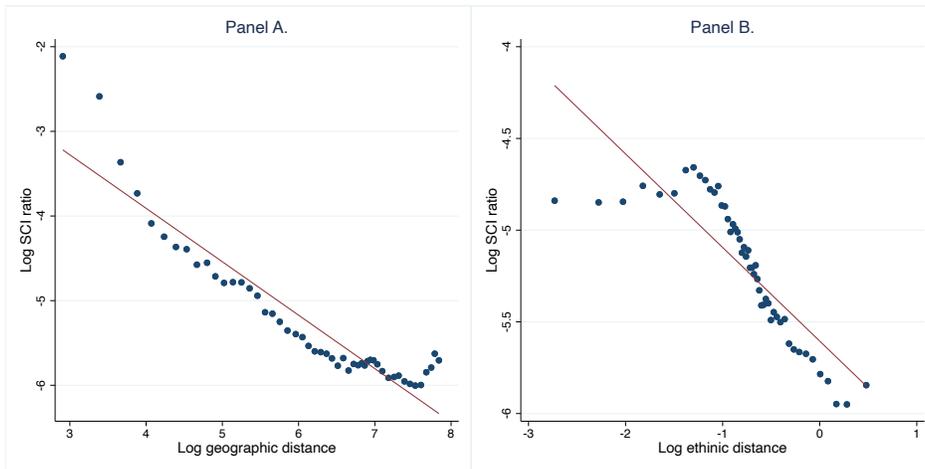
Note: This figure presents a binscatter plot of the relationship between SCI and geographic and ethnic distance between US counties. SCI refers to Facebook social connectedness index. Geographic distance is the distance between county pairs. Ethnic distance is constructed according to equation (23) using 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 regression 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 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, using a different migration data set.[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. One interpretation of the coefficients is that social linkages are more persistent through time than economic linkages.[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 due to factors 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 connnectedness 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.

---

[23]Bailey et al. (2018) uses the Statistics of Income (SOI) Tax Stats Migration Data provided by the IRS. Another 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 1: Relationship between Population Outflow Ratio and Log SCI Ratio.

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Log SCI Ratio	1.059***	1.094***	1.083***	1.048***	1.056***	1.007***
	(0.012)	(0.016)	(0.015)	(0.035)	(0.050)	(0.052)
Same State Dummy	0.219***	0.226***	0.235***	0.241***	0.258***	0.299***
	(0.036)	(0.036)	(0.036)	(0.079)	(0.054)	(0.056)
Geog. Distance		0.047**	0.043**		0.021	-0.007
		(0.020)	(0.019)		(0.039)	(0.039)
Hist. Ind. Distance			-0.123***			-0.148***
			(0.041)			(0.045)
Hist. Occ. Distance			-0.054***			-0.059***
			(0.018)			(0.019)
Observations	22281	22281	22281	22281	22281	22281
Mean dep. var.	-8.755	-8.755	-8.755	-8.755	-8.755	-8.755
R2	0.727	0.727	0.727	0.589	0.590	0.590
Kleibergen-Paap				276.454	271.529	236.014

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: "Log SCI Ratio", "Geog. Distance","His. Ind. Distance", "His. Occ. Distance" and "Same State Dummy" respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference, occupation composition difference defined in (23) and same-state status. Robust standard errors clustered at the origin and destination county levels in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

As shown in Table 2, 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

From (20), home bias at a given location can be inferred from the corresponding estimated origin and destination fixed effects. Here, we rely on the OLS estimates because they are more efficiently estimated. In Table 3, we report home bias estimates resulting from both outflow and inflow gravity estimations. The estimates are similar, ranging from 3% to 7%.[25] To highlight the importance of

---

[25]With reverse causality, it is possible that the estimated origin and destination fixed effects are biased. Specifically, suppose that

$$\ln\left(\frac{\mu_{mn}}{\mu_{mm}}\right) = \beta \ln\left(\frac{a_{mn}}{a_{mm}}\right) + D_n + T_m + \varepsilon_{mn}, \ \ \ln\left(\frac{a_{mn}}{a_{mm}}\right) = \ln\left(SCI_{mn}/SCI_{mm}\right) + b \ln\left(\frac{\mu_{mn}}{\mu_{mm}}\right).$$

Table 2: Relationship between Population Inflow Ratio and Log SCI Ratio.

	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Log SCI Ratio	1.063***	1.103***	1.090***	1.070***	1.088***	1.033***
	(0.013)	(0.017)	(0.016)	(0.035)	(0.049)	(0.051)
Same State Dummy	0.254***	0.261***	0.271***	0.240***	0.274***	0.319***
	(0.035)	(0.035)	(0.035)	(0.078)	(0.054)	(0.056)
Geog. Distance		0.053***	0.048**		0.043	0.010
		(0.020)	(0.020)		(0.038)	(0.038)
Hist. Ind. Distance			-0.109**			-0.127***
			(0.043)			(0.047)
Hist. Occ. Distance			-0.076***			-0.080***
			(0.018)			(0.019)
Observations	21701	21701	21701	21701	21701	21701
Mean dep. var.	-8.781	-8.781	-8.781	-8.781	-8.781	-8.781
R2	0.740	0.740	0.741	0.602	0.602	0.603
Kleibergen-Paap				268.975	266.114	232.575

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: "Log SCI Ratio", "Geog. Distance", "His. Ind. Distance", "His. Occ. Distance" and "Same State Dummy" respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference, occupation composition difference defined in (23) and same-state status. Robust standard errors clustered at the origin and destination county levels in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

incorporating credible measures of bilateral search intensity, Table 3 also reports the estimated home bias when SCI is not included in the regression. The resulting home bias estimates are evidently hugely divergent, ranging from 77% to over 81%.

To show the spread of these estimates, Figure 3 plots the distribution of home bias estimated with and without SCI using fixed effects estimates of outflow gravity, and the full list of explanatory variables as in Columns 3 of Table 3. The figure shows two right-skewed distributions, with home bias estimated without SCI having higher density around smaller values. In our model, home bias b_n in location n is a catch-all term, introduced to account for any factors that may lead to locational inertia that drives a wedge between the use-value of existing residents to deviate from new- or exchange-value of potential new resident. 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, and industrial and occupational structures.[26] 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 to minimum Bayesian information criterion. Table 4 reports the set of predictors selected by LASSO and their coefficients.

Using this approach, we confirm that our estimates of home bias is highly positively correlated with county-level congestion forces such as commute time and population density. This suggests that new migrants put a higher weight on the disutility of congestion and commute time than local residents in determining locational choice. This is consistent with asymmetric perceptions about the cost of living associated with congestion forces, where for example existing residents have had the time to find ways to cope with the disutility of congestion but new residents have yet to do so. Gender ratios features prominently also and may be telling a similar story related to congestion externalities associated with urban living, for U.S. counties with higher shares of males tend to be less urban.

The reduced form relationship is

$$\ln\left(\frac{\mu_{mn}}{\mu_{mm}}\right) = \frac{\beta}{1-\beta b}\ln\left(SCI_{mn}/SCI_{mm}\right) + \frac{1}{1-\beta b}D_n + \frac{1}{1-\beta b}T_m + \varepsilon_{mn}.$$

The estimated home bias $1 + \hat{b}_i = exp((T_i - D_i)/(1 - \beta b))$ thus exaggerates true home bias $exp(T_i - D_i)$. Using our OLS results (column 3) as an estimate of $\beta/(1 - \beta b)$ and our 2SLS result (column 6) in Table 1 and 2 as an estimate of β, the size of the bias in our estimate of home bias, $1/(1 - \beta b)$ is between 1.055 and 1.079. Thus, even accounting for these estimation biases, our estimates of home bias in 3 change very little from 0.055 to 0.051(= 0.055/1.079) for column (3), and for column (6) from 0.042 to 0.040(= 0.042/1.055).

[26]See Table A3 for a complete list of the variables.

Table 3: Home Bias Estimates from Outflow and Inflow Gravity with and without SCI

	(1) Outflow SCI	(2) Outflow SCI+Geo.	(3) Outflow SCI+Geo +Ind+Occ	(4) Inflow SCI	(5) Inflow SCI+Geo.	(6) Inflow SCI+Geo +Ind+Occ
Observations	22281	22281	22281	21701	21701	21701
Mean dep. var.	-8.755	-8.755	-8.755	-8.781	-8.781	-8.781
Mean Home Bias	0.071	0.051	0.057	0.053	0.029	0.042
Mean Home Bias (no SCI)	.	0.692	0.767	.	0.681	0.806

Note: This table displays the estimated home bias ratio with and without SCI based on OLS outflow and inflow gravity regressions in columns (1) - (3) of Tables 1 and 2 respectively for columns SCI, SCI+Geo, and SCI+Geo+Ind+Occ, with robust standard errors clustered at the origin and destination county levels.

Interestingly, home bias is also negatively correlated with communities with higher shares of retired individuals. The characteristics of such communities include a higher share population with retirement income or receiving social security payments. A preference for higher maximum January temperature may also be marker of older-aged communities. These observations tells a story often overlooked in migration studies, where retirement community amenities and climate factors can be drivers that attract new residents not captured in models where jobs are taken to be the primary reason behind mobility patterns. Beyond congestion forces and amenities for residents, housing also features prominently, where the share of relatively older housing (e.g. between 1940 and 1949) is positively correlated with home bias suggesting that an abundance of supply of older housing makes a friendlier environment for existing residents more so than newcomers. The opposite pattern can be observed for relatively new housing shares. 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 construction, wholesale, information, transportation and utility employment tend to exhibit higher and positive levels of home bias.

These lessons are a useful reminder that sources of spatial heterogeneity in living standards extend well beyond wages and job prospects. Once differences in social connectedness are accounted for, our estimated home bias can be seen as a reflection of spatial heterogeneity due to congestion forces, amenities and climate, neighborhood communities, housing and job differences. Naturally, different population subgroups may value these features differently and our home bias estimates highlight the differences depending on existing or new resident statuses. Of course, there may other predictors of preferences as well. The lack of social connectedness indicators specific to age group

27

Table 4: Predictors of Home Bias

Ranking	Predictor	Coefficient
1	Log avg. commute time	0.1718
2	% males	-0.1495
3	% household with retirement income	-0.0939
4	% poor	-0.0937
4	% wholesale ind.	0.0523
5	% construction ind.	0.0499
6	% housing built between 1990-1999	-0.0487
8	% grand parents caring for children	-0.0426
9	Log popultion density	0.0417
10	% household with food stamp	-0.0302
11	% carpool	-0.0237
12	% housing built between 1980-1989	-0.0232
13	% information ind.	0.0229
14	% younger than 20	0.0184
15	% transportation and utility occ.	0.0114
16	% housing built between 1940-1949	0.0082
17	% agriculture ind.	-0.0041
18	Maximum January temp.	-0.0032
19	% transportation ind.	0.0025
20	% household on social security	-0.0023

.

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 the Appendix

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 6 we rank counties according to the estimated level of home bias.[27] 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 (Figure 4). The list also contains smaller counties, suggesting that the estimated home biases 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.

5.3 Unemployment

Our model predicts that a county that is more socially integrated with high expected utility destination locations will experience lower unemployment in equilibrium. To test this prediction, we estimate the following equation using OLS:

$$\ln(u_m) = \alpha + \sum_i \beta_{ci} Covar(x^i_{mn}, W_{mn}) + \sum_i \beta_{li} \bar{x}^i_m \bar{W}_m + \phi_{r/s} + \varepsilon_m, \qquad (25)$$

where u_m is county unemployment rate. $\bar{x}^i_m = \sum_i x^i_{mn}/M_m$ is our proxy for the level effect of search intensity control i, and $\bar{W}_m = \sum_n W_n \equiv \sum_n W_n \left(1 - \mathbb{I}_{mn} b_n/(1 + b_n)\right)$ is the overall expected utility of the destinations of m, adjusted by our home bias estimates b_n. $Covar(x^i_{mn}, W_{mn}) = \sum_n (x^i_{mn} - \bar{x}^i_m)(W_{mn} - \bar{W}_m)$ is the covariance term following (21). We include as possible search intensity controls (x^i) social connectedness proxied by Facebook friendship data (SCI), geographical distance ($Geodist$), historic industry distance ($Inddist$) and historic occupational distance ($Occudist$). The $\phi_{r/s}$'s are either region or state fixed effects. The coefficients β_{ci} and β_{li} respectively capture the interaction and level effect of the corresponding search intensity control. Our theory informs us that β_{li} and β_{ci} are negative for search intensity proxies that mitigate against job search friction, such as the SCI. A priori, we do not want rule out the possibility that $Geodist$, $Inddist$ and $Occudist$ are determinants of a_{mn} and thus we have kept them in the list of search intensity controls as reference. Since these are distance rather than proximity indicators, our theory informs us that β_{ci} and β_{li} should be positive for these controls.

---

[27]Similar rankings of countries' trade costs are often used in the trade literature (e.g., Eaton and Kortum, 2002).

Figure 3: County-Level Home Bias (b_n) Dispersion.
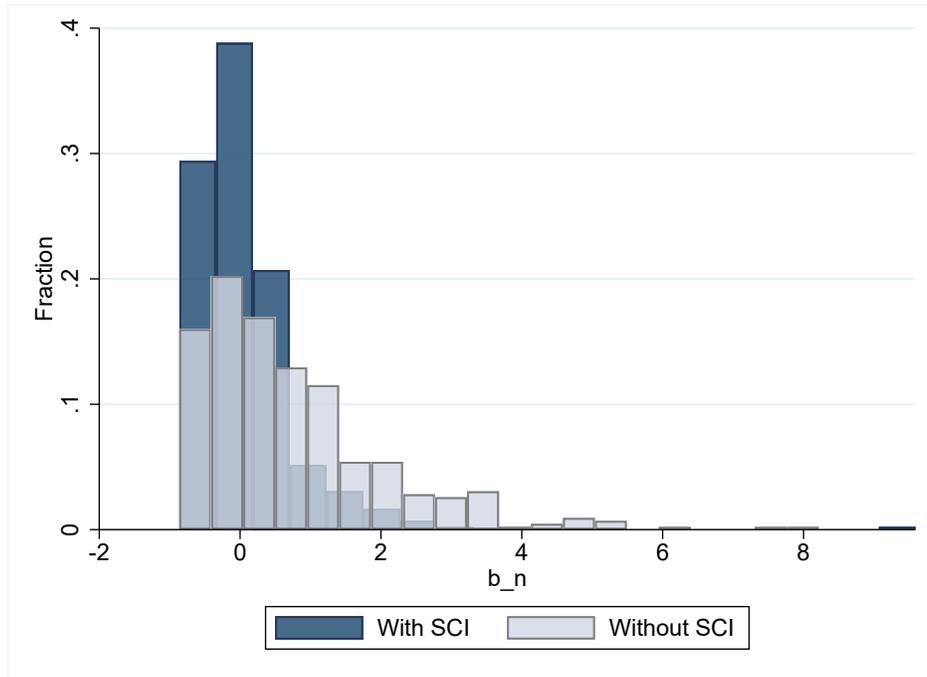
Note: This figure displays histograms of estimated home bias b_n with and without SCI based on OLS outflow gravity regressions in columns (3) of Table 3 with robust standard errors clustered at the origin and destination county levels.

Figure 4: The Geography of Average State-Level Home Bias (b_n).
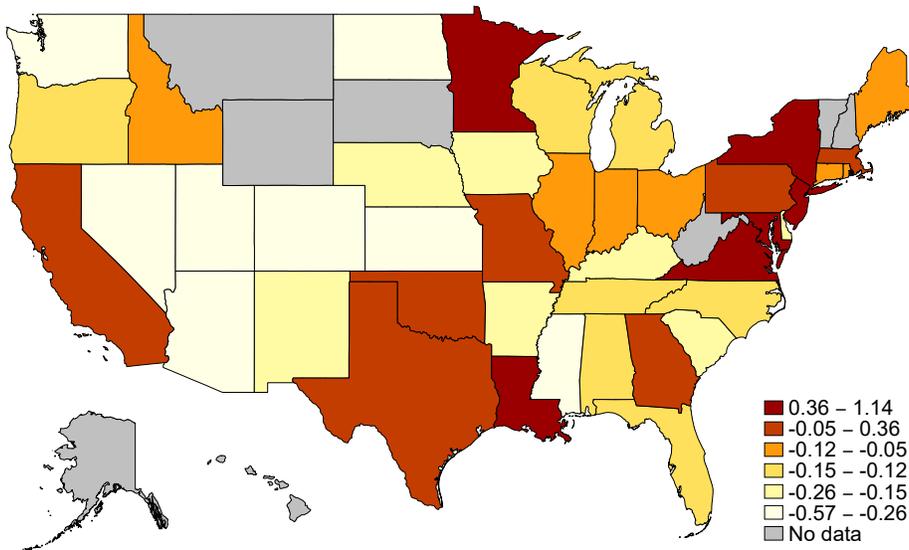
Note: This figure displays state-level average home bias (b_n) estimated based on OLS outflow gravity regressions in column (3) of Table 1 with robust standard errors clustered at the origin and destination county levels.

Table 5 presents the results. We use our own home bias estimates to construct the interaction and level effects associated with the SCI and destination expected utilities.[28] We will rely on home bias estimates from our outflow gravity regression in Column 3 of Table (1), since fixed effects from outflow regression correspond directly to expected utilities as required here from (17).

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. According to the estimates in Column 1, a one-standard-deviation increase in the covariance between search intensity and destination's expected utility lowers unemployment rate by 6.8 percent. 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.

We provide two additional sets of results. In Table A2, we present a set of naive regressions, in which we simply regress log unemployment rate against the simple unweighted average SCI of the county in question, controlling for state and /or regional fixed effects. We find that average SCI alone is not a good predictor of unemployment. Next, in Tables 5, we present unemployment regressions accounting for the level and interaction effects of the full list of search intensity controls, including geographic distance, industry as well as occupation difference. Interestingly, we find that once we control for social connectedness, the level and / or interaction effect of the other three distance variables take on the wrong sign, while the corresponding social connectedness terms remain more or less the same. In other words, once SCI is controlled for, we find that the other distance variables do not behave in ways a search friction control is expected to.

Finally, with our limited number of counties, we want to guard against the possibility that these results are driven by outliers as county-level unemployment rates range widely, from 2% to 23% in our data set. Accordingly, we ran a series of quantile regressions, with the full list of search intensity level and interaction effects as column 4 in Table (5) with state fixed effects and clustered standard errors at the state level. We plot the quantile regression coefficients and confidence intervals associated with the standardized SCI interaction effect in Figure (5) for the full range of quantiles. We also display the OLS coefficient and confidence interval for comparison. As shown the estimated coefficients are for the most part stable, and statistically significant almost always and strictly negative throughout. We take this as evidence that the negative relationship between the local unemployment rate and the SCI interaction effect is not driven by outliers at

---

[28]The outcomes are log unemployment rate multiplied by 100 and thus the coefficients can be interpreted as percent changes.

either extremes of the unemployment distribution.

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 interaction effect of friendship networks links with high expected utility locations. Thus, social conections do matter as a determinant of unemployment provided that it is of the right kind, connecting individuals with those residing in desirable (high expected utility) locations. Notably, locational desirability itself in our model is an inter-regional concept, depending on average utility w_n conditional on finding a job, 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.

We unpack the determinants of home bias, and found that factors related to congestion effects, population demographics, housing, and the composition of jobs to be significant correlates. These offer rich contexts for future research on labor mobility and unemployment, to complement a growing literature that have incorporated a diversity of non-wage determinants of the gravity of migration (e.g., Chau, 1997; McKenzie and Rapoport, 2007; Belot and Ederveen, 2012; Manning and Petrongolo, 2017).

Table 5: Unemployment and Outward Multilateral Resistance Using Home Bias from Outflow Gravity

	(1) ln_unempl_rate	(2) ln_unempl_rate	(3) ln_unempl_rate	(4) ln_unempl_rate
Std. Interaction Effect (SCI)	-6.805***	-5.875***	-8.358***	-6.876***
	(1.774)	(1.843)	(2.717)	(2.395)
Std. Interaction Effect (Geodist)			-15.454***	-11.765**
			(3.398)	(4.922)
Std. Interaction Effect (Inddist)			5.445	7.238**
			(3.521)	(3.456)
Std. Interaction Effect (Occudist)			-7.607***	-8.249***
			(2.742)	(2.824)
Std. Level Effect (SCI)	-5.223***	-5.650***	-6.150*	-4.130
	(1.574)	(1.977)	(3.371)	(2.678)
Std. Level Effect (Geodist)			-18.756***	-15.446***
			(4.213)	(4.028)
Std. Level Effect (Inddist)			9.014*	9.060*
			(5.001)	(5.226)
Std. Level Effect (Occudist)			-10.316*	-13.481***
			(5.438)	(4.671)
Observations	425	421	425	421
Fixed effect	Region	State	Region	State
R2	0.171	0.376	0.296	0.502

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), $x = SCI, Geodist, Inddist, Occudist$ – 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 1) and M; (ii) an interaction effect (x), $x = SCI, Geodist, Inddist, Occudist$ – the product of the covariance between each bilateral search intensity control and the expected utility of the corresponding destination $Cov(x_{mn}^i, W_n)$ and M, and (iii) a state level / regional level baseline. "SCI", "Geodist", "Indist" and "Occudist" respectively refer to Facebook social connectedness index, geographic distance, industrial composition difference and occupation composition difference defined in (23). Standardized coefficients are presented. Robust standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: County Rankings: Top 10 and Bottom 10 Degree of Home Bias

Ranking	Highest Migration Home Bias	Ranking	Lowest Home Bias
1	Ascension Parish, LA	414	Tippecanoe County, IN
2	Clayton County, GA	415	Santa Rosa County, FL
3	New York County, NY	416	Calhoun County, AL
4	Nassau County, NY	417	Ashtabula County, OH
5	Los Angeles County, CA	418	Coconino County, AZ
6	Richmond County, NY	419	Imperial County, CA
7	Prince George's County, MD	420	Wichita County, TX
8	Gwinnett County, GA	421	Harrison County, MS
9	Queens County, NY	422	Craven County, NC
10	Schenectady County, NY	423	Richland County, OH
11	Bronx County, NY	424	Kings County, CA
12	San Francisco County, CA	425	Yum County, AZ

Note: This table lists the top 12 and bottom 12 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.

Figure 5: Quantile Regression Coefficients: Standardized Interaction Effect (SCI)
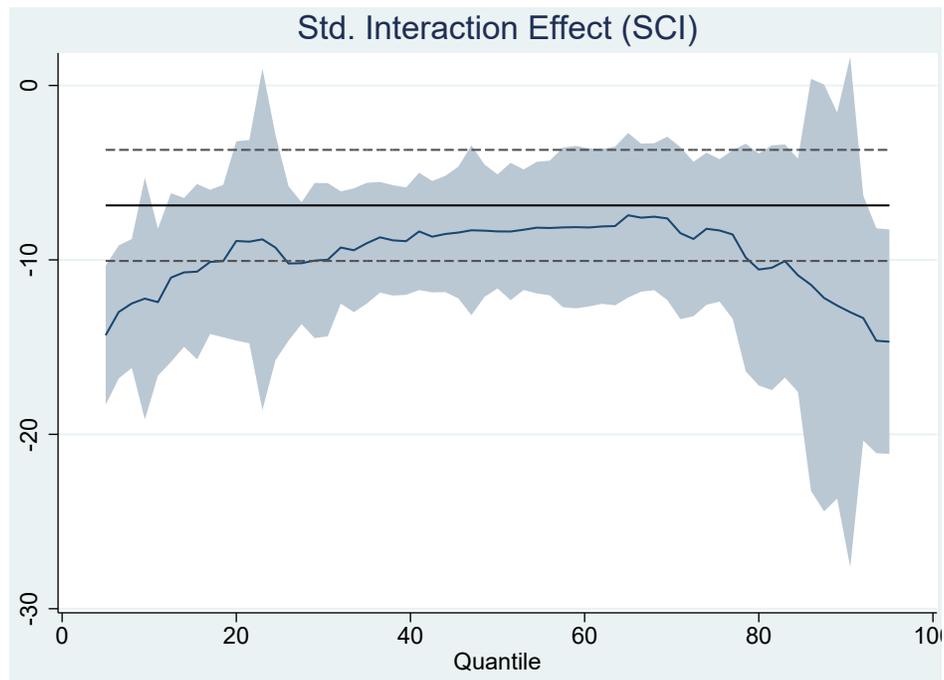
Note: This figure displays quantile regression coefficients of the standardized interaction effect (SCI) across quantiles. The unemployment regression includes the full list of covariance and level effects, in addition to state fixed effects and standard errors clustered at the state level as in Column 2 of Table 5. The shaded area displays the confidence intervals. The solid line is the OLS regression coefficient. The dashed lines are the confidence intervals associated with the OLS estimate.

References

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015, November). The Economics of Density: Evidence From the Berlin Wall. *Econometrica 83*(6), 2127–2189.

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.

Albouy, D. and B. Stuart (2014, February). Urban Population and Amenities: The Neoclassical Model of Location. Working Paper 19919, National Bureau of Economic Research.

Amior, M. and A. Manning (2018). The Persistence of Local Joblessness. *American Economic Review 108*(7), 1942–70.

Anderson, J. (2011). The Gravity Model. *Annual Review of Economics 3*(1), 133–160.

Anderson, J. E. and E. van Wincoop (2004, September). Trade Costs. *Journal of Economic Literature 42*(3), 691–751.

Anderson, J. E. and Y. V. Yotov (2010). The Changing Incidence of Geography. *American Economic Review 100*(5).

Artuç, E., S. Chaudhuri, and J. McLaren (2010, June). Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review 100*(3), 1008–1045.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018, August). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives 32*(3), 259–280.

Balkema, A. A. and L. de Haan (1974). Residual Life Time at Great Age. *The Annals of Probability 2*(5), 792 – 804.

Basso, G. and G. Peri (2020). Internal Mobility: The Greater Responsiveness of Foreign-Born to Economic Conditions. *Journal of Economic Perspectives 34*(3), 77–98.

Basu, A. K., N. H. Chau, and B. Park (2022). Rethinking Border Enforcement, Permanent and Circular Migration. *Economic Modelling 108*, 105733.

Beine, M., S. Bertoli, and J. Fernández-Huertas Moraga (2016). A practitioners' guide to gravity models of international migration. *The World Economy 39*(4), 496–512.

Beine, M., F. Docquier, and a. Özden (2011). Diasporas. *Journal of Development Economics 95*(1), 30–41.

Beine, M. and C. Parsons (2015). Climatic Factors as Determinants of International Migration. *Scandinavian Journal of Economics 117*(2), 723–767.

Belot, M. and S. Ederveen (2012). Cultural Barriers in Migration between OECD Countries. *Journal of Population Economics 25*(3), 1077–1105.

Bertoli, S. (2010). Networks, Sorting and Self-Selection of Ecuadorian Migrants. *Annals of Economics and Statistics* (97/98), 261–288.

Bertoli, S. and J. Fernández-Huertas Moraga (2012). Visa policies, networks and the cliff at the border. IZA Discussion Papers 7094, Institute of Labor Economics (IZA).

Bertoli, S. and J. Fernández-Huertas Moraga (2013, May). Multilateral resistance to migration. *Journal of Development Economics 102*, 79–100.

Bilal, A. and E. Rossi-Hansberg (2018, July). Location as an Asset. Working Paper 24867, National Bureau of Economic Research.

Blanchard, O. and L. Katz (1992). Regional evolutions. *Brookings Papers on Economic Activity 23*(1), 1–76.

Blumenstock, J. E., G. Chi, and X. Tan (2019). Migration and the Value of Social Networks. CEPR Discussion Paper DP 13611, CEPR.

Borjas, G. J. (1992). Ethnic Capital and Intergenerational Mobility. *The Quarterly Journal of Economics 107*(1), 123–150.

Boustan, L. P. (2013, July). Local Public Goods and the Demand for High-Income Municipalities. *Journal of Urban Economics 76*, 71–82.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica 82*(5), 1671–1748.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica 87*(3), 741–835.

Caliendo, M., D. A. Cobb-Clark, J. Hennecke, and A. Uhlendorff (2019). Locus of control and internal migration. *Regional Science and Urban Economics 79*, 103468.

Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *ILR Review 43*(2), 245–257.

Cattaneo, C. and G. Peri (2016, September). The Migration Response to Increasing Temperatures. *Journal of Development Economics 122*, 127–146.

Chau, N. H. (1997). The Pattern of Migration With Variable Migration Cost. *Journal of Regional Science 37*(1), 35–54.

Chen, Y. and S. Rosenthal (2008). Local amenities and life-cycle migration: Do people move for jobs or fun? *Journal of Urban Economics 64*(3), 519–537.

Coles, S., J. Bawa, L. Trenner, and P. Dorazio (2001). *An introduction to statistical modeling of extreme values*, Volume 208. Springer.

Dao, M., D. Furceri, and P. Loungani (2017). Regional labor market adjustment in the united states: Trend and cycle. *The Review of Economics and Statistics 99*(2), 243–257.

DaVanzo, J. (1978). Does Unemployment Affect Migration? Evidence from Micro Data. *The Review of Economics and Statistics 60*(4), 504–514.

De Weerdt, J., L. Christiansen, and R. Kanbur (2021). When Distance Drives Destination, Towns can Stimulate Development. Applied Economics and Policy Working Paper Series 310517, Cornell University, Department of Applied Economics and Management.

Dekker, R. and G. Engbesen (2014). How Social Media Transforms Migrant Networks and Facilitate Migration. *Global Networks 14*(4), 401–418.

Dix-Carneiro, R. (2014, May). Trade Liberalization and Labor Market Dynamics. *Econometrica 82*(3), 825–885.

Djajic, S. and R. Milbourne (1988). A general equilibrium model of guest-worker migration: The source-country perspective. *Journal of International Economics 25*(3-4), 335–351.

Dorn, D. and J. Zweimüller (2021). Migration and Labor Market Integration in Europe. Technical report.

Dustmann, C. (2003). Children and return migration. *Journal of Population Economics 16*(4), 815–830.

Dustmann, C., F. Fabbri, and I. Preston (2005). The Impact of Immigration on the British Labour Market. *Economic Journal 115*(507), F324–F341.

Dustmann, C. and A. Glitz (2011). Chapter 4 - migration and education. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of The Economics of Education*, Volume 4 of *Handbook of the Economics of Education*, pp. 327–439. Elsevier.

Eaton, J. and S. Kortum (2002, September). Technology, Geography, and Trade. *Econometrica 70*(5), 1741–1779.

Faini, R. and A. Venturini (2001). Home Bias and Migration: Why is Migration Playing a Marginal Role in the Globalization Process? Working Paper 21, Centre for Household, Income, Labour and Domestic Economics.

Feng, S., M. Oppenheimer, and W. Schlenker (2012, January). Climate Change, Crop Yields, and Internal Migration in the United States. Working Paper 17734, National Bureau of Economic Research.

Greenwood, M. J. (1985). Human Migration: Theory, Models, and Empirical Studies. *Journal of Regional Science 25*(4), 521–544.

Grogger, J. and G. Hanson (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics 95*(1), 42–57.

Harris, J. R. and M. P. Todaro (1970). Migration, Unemployment & Development: A Two-Sector Analysis. *American Economic Review 60*(1), 126–142.

Head, K. and J. Ries (2001). Increasing returns versus national product differentiation as an explanation for the pattern of u.s.-canada trade. *American Economic Review 91*(4), 858–876.

Heise, S. and T. Porzio (2019). Spatial wage gaps and frictional labor markets. Staff Reports 898, Federal Reserve Bank of New York.

King, R., E. Cela, and T. Fokkema (2021). New frontiers in international retirement migration. *Ageing and Society 41*(6).

López Córdova, J. E. (2018). Globalization, migration and development: The role of mexican migrant remittances. IDB Publications (Working Papers) 2652, Inter-American Development Bank.

Manning, A. and B. Petrongolo (2017). How Local Are Labor Markets? Evidence from a Spatial Job Search Model. *American Economic Review 107*(10), 2877–2907.

Mayda, A. M. (2010). International Migration: a Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics 23*(4), 1249–1274.

Mayer, T. and K. Head (2002). Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade. Working papers, CEPII research center.

McKenzie, D. and N. Hildebrandt (2005). The Effects of Migration on Child Health in Mexico. *EconomÃa Journal Volume 6 Number 1*(Fall 2005), 257–289.

McKenzie, D. and H. Rapoport (2007, September). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics 84*(1), 1–24.

Molloy, R., C. L. Smith, R. Trezzi, and A. Wozniak (2016, March). Understanding Declining Fluidity in the U.S. Labor Market. SSRN Scholarly Paper ID 2741559, Social Science Research Network, Rochester, NY.

Molloy, R., C. L. Smith, and A. Wozniak (2011, September). Internal Migration in the United States. *Journal of Economic Perspectives 25*(3), 173–196.

Monte, F. (2015, January). The Local Incidence of Trade Shocks. SSRN Scholarly Paper ID 2337270, Social Science Research Network, Rochester, NY.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018, December). Commuting, Migration, and Local Employment Elasticities. *American Economic Review 108*(12), 3855–3890.

Moretti, E. (2011). Chapter 14 - Local Labor Markets. In D. C. a. O. Ashenfelter (Ed.), *Handbook of Labor Economics*, Volume 4, Part B, pp. 1237–1313. Elsevier.

Morten, M. and J. Oliveira (2016, April). Paving the way to development: Costly migration and labor market integration. Working Paper 22158, National Bureau of Economic Research.

Munshi, K. (2003, May). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics 118*(2), 549–599.

Munshi, K. (2014, November). Community Networks and the Process of Development. *Journal of Economic Perspectives 28*(4), 49–76.

Notowidigdo, M. J. (2020). The incidence of local labor demand shocks. *Journal of Labor Economics 38*(3), 687 – 725.

Redding, S. J. (2016, July). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty 1*(1), 7–59.

Santos Silva, J. and S. Tenreyro (2021). The Log of Gravity at 15. School of Economics Discussion Papers 0121, School of Economics, University of Surrey.

Stoll, M. A. and S. Raphael (2000). Racial differences in spatial job search patterns: Exploring the causes and consequences. *Economic Geography 76*(3), 201–223.

Tombe, T. and X. Zhu (2019). Trade, Migration, and Productivity: A Quantitative Analysis of China. *American Economic Review 109*(5), 1843–72.

Woodruff, C. and R. Zenteno (2007). Migration networks and microenterprises in mexico. *Journal of Development Economics 82*(2), 509–528.

Appendix A

In this appendix, we demonstrate the structural gravity equation as displayed in Equation (11). To start, let $M_{mn} = \mu_{mn}N_m$ denote total migration, and $L_n = \sum_m M_{mn}$ as total employment in n. From (9),

$$L_n = W_n \sum_m \left(\frac{\alpha_{mn}N_m(1 - u_m)}{O_m} \right)$$

or equivalently

$$W_n = \frac{L_n}{I_n \sum_i N_i(1 - u_i)}$$

where

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

Substituting into (9), we obtain

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

as displayed in (11) and

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

as displayed in (13). As discussed, total migration between two locations depends on (i) bilateral home bias adjusted search intensity α_{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, $N_m(1 - u_m)$, 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 $\sum_i N_i(1 - u_i)$.

Appendix B

Table A1: (Non-)Drivers of Historic Ethnic Distance

	(1) OLS	(2) OLS
Ln His. Ind. Distance	-0.054 (0.048)	-0.032 (0.026)
Ln His. Occ. Distance	0.028 (0.022)	0.018 (0.013)
Ln Geog. Distance		0.018*** (0.006)
Same State Dummy		-1.532*** (0.037)
Observations	24459	24459
R^2	0.0007	0.7669
Mean dep. var.	-0.783	-0.783

Note: This table displays the relationship between historic ethnic distance and historic industry and occupation distance index defined in (23). Robust standard errors clustered at the origin and destination county levels in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Naive Unemployment and SCI Relationship

	(1) ln_unempl_rate	(2) ln_unempl_rate
Std. SCI	-1.140	-1.756
	(1.723)	(1.825)
Observations	425	421
Fixed effect	Region	State
R2	0.138	0.351

Note: This table displays the relationship between U.S. county-level log unemployment rates and un-weighted average county-level SCI controlling for state- or regional-level baseline. Standardized coefficients are presented. Robust standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: List of Variables in Lasso Regressions.

Variable Group	Variable List
Commute	log avg. commute time, % commuters, % drive alone, % carpool, % take public transport
Demographics	% 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
Environment	heat days, July precipitation, January maximum temperature, and July max temperature
Housing	% housing built 2010 or later, % built between 2000 and 2009, % built between 1990 and 1999, % built between 1980 and 1989, % built between 1970 and 1979, % built between 1960 and 1969, % built between 1950 and 1950, % built between 1940 and 1949, % built before 1940
Industry	% agriculture, % construction, % manufacturing, % whole sale, % retail, % transportation, % information, % finance, insurance, and real estate, % public education, health, and social services, % recreational and entertainment, % other industries, and % public admin
Marriage	% living alone, % with children, % divorced, % grand parents caring for children
Occupation	% management and professional, % construction, % farm and fish, % sales and office, % service, % transportation and utility
Public assistance	% 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
Religious	% Evangelical, % Catholic, and % Mainline Protestant
Social Capital	crime per capita, and republican vote share