I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14677

Migrant Networks and Destination Choice: Evidence from Moves across Turkish Provinces

Abdurrahman B. Aydemir
Erkan Duman

AUGUST 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14677

Migrant Networks and Destination Choice: Evidence from Moves across Turkish Provinces

Abdurrahman B. Aydemir
Sabancı University and IZA

Erkan Duman
Sabancı University

AUGUST 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Migrant Networks and Destination Choice: Evidence from Moves across Turkish Provinces

This paper estimates effects of birth place migration networks and other location attributes on destination choices of internal migrants conditional on migration. We also study heterogeneity in the role of these factors for migrant types who differ by skill group, age at migration, and reason of migration. We use data on male migrants from three rounds of Turkish censuses 1985, 1990 and 2000 who choose among 67 provinces. We find that migrants are drawn to provinces with larger networks, relatively better economic conditions, and distance is a significant deterrent for migration. There are, however, significant heterogeneities across migrant types. More educated move longer distances and rely less on networks for destination choice. Importance of labor market conditions increases and the effect of distance decreases with age. Among migrants with different reason of migration, labor market conditions play a significant role only for migrants moving for employment reasons and networks matter less for this group.

JEL Classification: J61, O15, R23, Z13

Keywords: migration, networks, destination choice, education, reason of migration, heterogeneous effects

Corresponding author:
Abdurrahman B. Aydemir
Faculty of Arts and Social Sciences
Sabancı University
Orhanli
Tuzla 34956
Istanbul
Turkey

E-mail: aaydemir@sabanciuniv.edu

1. Introduction

Destination choices of migrants shape migration flows with important implications for both receiving and sending regions. There is a growing literature on destination choices of international migrants that studies either flows between countries or location preferences of immigrants within a host country (e.g., Bertoli and Ruyssen, 2018; Bartel, 1989; Jaeger, 2000). There is relatively little work, however, on destination choices of internal migrants. Some of these studies are in developed country contexts (Davies et al., 2001 and Stuart and Taylor, 2021 on the US; Etzo, 2011 and Piras, 2020 on Italy) while only few consider less developed ones (Fafchamps and Shilpi, 2013 on Nepal; Swee, 2017 on Thailand; Davis et al., 2002 on Mexico). This paper contributes to the literature by studying location choice of internal migrants in the context of a middle-income country–Turkey. The main focus is on the role of networks while we consider several other location attributes. Our paper also studies the heterogeneity in the role of location attributes in shaping migrant flows across types of migrants who differ by skill group, age at migration, and reason of migration.

As in other developing countries Turkish economy has been undergoing a structural transformation where the economic activity shifted from agriculture to manufacturing to services. Large disparities in standards of living across regions —the largest regional gap in GDP per capita among OECD countries (OECD, 2018)— coupled with this structural transformation, has fueled large migration flows in the country that intensified over time. Annual interprovincial migration has increased from around 1.4% in 1980s to over 3.2% by 2017. In this high internal migration context, we study the choice of destination conditional on migration for a sample of male migrants who moved across province[1]. Within a conditional logit framework, we study the influence of location attributes that include network size, labor market characteristics, demographic structure, and distance.

While there is a voluminous literature on international migrants' location propensities that focuses on developed countries as hosts (see Bartel, 1989; Bauer et al., 2005, 2007; Davis et al., 2002; Dunlevy, 1991; Jaeger, 2000, 2007; Zavodny, 1999; Aslund, 2005; Damm, 2009), there are only few studies on internal migrants' location choices (see Davies et al., 2001; Davis et al., 2002; Fafchamps and Shilpi, 2013; Stuart and Taylor, 2021). This literature assesses the

[1] Migrants may differ from non-migrants in terms of both observable and unobservable characteristics. We do not study migration decision but choice of destination conditional on migration decision. Hence, similar to other studies on destination choice, our study abstracts from the issue of selection into migration.

role of location attributes including networks, economic conditions, and amenities in alternative destinations and distance to place of origin in the destination choices of migrants.

Studies on location choices of internal and international migrants reveal that ties of kinship, acquaintanceship, and birth place link former and latter migrants, and influence the destination choices of potential migrants by providing them information about labor and housing markets at destination, aiding adjustment, and reducing psychic costs.[2] While the definition of migrant ties and their measurement varies across studies, a finding that emerges from these studies is the significance of ties that are formed by sharing the same place of birth (e.g., same birth region/country/district/village). Turkish context provides a unique example for birth place ties that give rise to hometown associations which are called "hemşehri dernekleri". These associations are formed in destination regions with the aim of increasing solidarity between fellow countrymen and maintaining ties between members. Civil society associations are registered legal entities in Turkey and hometown associations constitute 13.7% of all civil society associations (STIGM, 2021). There is also a close relationship between interprovincial smigration and prevalence of hometown associations. The correlation between out-of-province birth share of population and province level share of hometown associations among civil society associations is 0.455 in Turkey.[3] The prevalence of these associations and their strong correlation with internal migration intensity indicates that birth place based ties are important part of the migrant networks in our context.[4] Thus, in this paper we proxy for networks based on province of birth.[5,6]

The first contribution of this paper to the literature is to test the role of birth place migration networks by modelling individuals' choices over destinations. There are only few

[2] The beneficial network externalities are discussed by numerous studies (e.g., Banerjee, 1983; Gottlieb, 1987; Grossman, 1989; Marks, 1989; Bartel, 1989; Dunlevy, 1991; Chiswick and Miller, 1996; Zavodny, 1999; Zahniser, 1999; Winters et al., 2001; Davis et al., 2002; Munshi, 2003; Bauer et al., 2005, 2007; Jaeger, 2000, 2007; Fafchamps and Shilpi, 2013).

[3] We calculate this correlation using two different data sources: ADNKS (Register based population counts) that reports out-of-province birth shares of province populations for 2020 and STIGM statistics on civil society associations by type as of June 2021 (STIGM statistics are not available for earlier years). Note that while some of the hometown associations are formed by out of province migrants others are formed by migrants who moved within a province.

[4] The practicality and prevalence of these associations in provinces in the Turkish context is discussed by Ergür (2006) and İnat (2006).

[5] In the public discourse on internal migration in Turkey source province is the single most important factor that identifies migrants. While forming new acquaintances, often one of the first questions posed is the province of birth as people try get to know each other.

[6] An alternative measure of network size would be prevalence of hometown associations at province level by source region of migrants; however, the information on source region of these associations is not available.

papers that similarly model individual choices over alternative destinations and they differ from our work in that they either consider other definitions of networks (e.g., Fafchamps and Shilpi, 2013) or study developed country contexts (e.g., Jaeger, 2000; Stuart and Taylor, 2021). In this line of work, our paper is the first one that tests the role of birth place networks –so called weak ties defined at province level– in a developing country context. It is not clear a priori whether weak ties matter in less developed country contexts –individuals may mainly rely on strong ties; i.e. friends and relatives. Our results provide new evidence on the role of birth place networks and complements findings in the existing literature.[7]

The effects of location attributes, such as network size or labor market characteristics, may differ across different types of migrants. For example, more educated individuals may put less weight on networks in their destination decisions if they value diversity more or benefit less from network-based job referrals.[8] More educated may also face lower costs of information acquisition about destinations, reducing both the deterring effect of distance and value of information networks provide. Similarly, the effects of location attributes on destination choice may differ by reason of migration. For example, migrants who move for job search or to start a job may be more likely to respond to labor market prospects relative to those moving as tied movers or for education purposes. There may also be differences by age if the preferences for certain location attributes change or individuals accumulate more migration capital as they get older. For example, as individuals get older they may put more weight on attributes such as climate or amenities. These types of heterogeneity have received limited attention in the literature. Our second contribution is to investigate such heterogeneities. We estimate our models by migrant type, distinguishing between education groups, age at migration, and reason of migration which provides new insights about the heterogenous effects of location attributes on destination choice.

[7] In the internal migration context Swee (2017) defines networks by contacts that originate from the same village and studies effects on job search process. Etzo (2011) defines networks by lagged out-flows from a region, Piras (2020) and Stuart and Taylor (2021) by common birth region. There are also other definitions of networks adopted within internal migration context, e.g., ethno-caste, language and religion proximity (Fafchamps and Shilpi, 2013) and kinship (Banerjee, 1983 and Davis et al., 2002). The role of birth place networks on destination choice is also studied in the international migration context (Bartel, 1989; Scott et al., 2005; Jaeger, 2000; Bauer et al., 2005).

[8] There are few papers that study the role of networks by skill group, mostly in international migration context: Bartel (1989), Carrington et al. (1996), Winters et al. (2001), McKenzie and Rapoport (2010), Beine et al. (2011) discuss that network effects decline with skill levels of international migrants. Hellerstein et al. (2011) report a similar result for residential labor market networks. The elasticity of migration with respect to distance is also shown to be lower for more educated (e.g., Schwartz, 1973 and Schultz, T.P., 1982).

We estimate our models by using data from three rounds of Turkish population censuses: 1985, 1990 and 2000. To the best of our knowledge, this is the first paper on destination choices within the Turkish context where internal migration has been an important factor in the past few decades shaping the economic and social landscape of the country. Estimating the effects of location attributes on destination choice is challenging due to both omitted variable and simultaneity bias. Omitted variable bias may arise, for example, if network size is correlated with other unobserved characteristics of regions that affect location decisions. Large inflows of migrants to a location may, on the other hand, affect province level characteristics such as unemployment rate leading to simultaneity issues. In our conditional logit specification, to account for potential biases, we include a set of fixed effects for destination provinces, define location attributes as differences between origin and destination, and use location characteristics of destinations from an earlier time period before migration takes place.

In our context, the identifying variation in models estimated with single cross-sectional data comes from within region differences in covariates while the identification strategy in models estimated with pooled cross-sectional data exploits variation in location characteristics both within regions and within region variation over time. For example, consider the provinces of Şanlıurfa and Diyarbakır that border each other and are on par in level of development. These two provinces rank among the top five provinces in Turkey in terms of emigration intensity to other provinces between 1995 and 2000 (Kocaman, 2008). During this period the top three destination provinces with highest share of emigrants from the province of Diyarbakır are İstanbul with 17.7%, İzmir with 12.2% and Ankara with 7.3% while the corresponding destinations and fractions are Gaziantep with 13.3%, İstanbul with 9.9% and Aydın with 7.9% for Şanlıurfa (Kocaman, 2008).[9] This type of variation between provinces with similar source

[9] On top of being neighbors –relevant in holding fixed the migration cost in our estimation strategy, according to a socioeconomic development index calculated by Ministry of Development in 1996 (see Dinçer et al., 1996) Diyarbakır and Şanlıurfa were similar in terms of demographics, labor market, education and health characteristics and ranked 57th and 59th among 76 provinces, respectively. An observation consistent with positive migrant network externalities hypothesis is that the top three provinces that attract most migrants originated from Şanlıurfa and Diyarbakır between 1995 and 2000 also rank high among provinces with respect to migrant network measures of Şanlıurfa and Diyarbakır calculated using 1990 census: the foreign-born share in population is highest in İstanbul followed by İzmir and Ankara ranking 3rd and 4th respectively; Gaziantep is top ranking province with highest share of Şanlıurfa born emigrants in a province population, as well as housing highest frequency of Şanlıurfa borns among all provinces of Turkey while İstanbul ranked 3rd according to this latter measure. On the other hand, İstanbul houses Diyarbakır borns more than any other province while İzmir and Ankara are the 3rd and 6th provinces with highest frequency of Diyarbakır borns.

province characteristics and similar migration costs for a given destination helps us distinguish the effect of networks from other factors that determine location choices.

Our findings from various specifications and subsamples consistently show that migrants are drawn to provinces with larger networks, relatively better economic conditions, and that distance is a significant deterrent for migration. We find no evidence for adverse impacts of migrant enclaves that may stem from increased competition among compatriots for job opportunities. Important heterogeneities also emerge from our estimation. First, the results indicate that with increasing education, migrants move longer distances and the impact of networks on location choice decreases. Education groups respond to unemployment rate differences similarly with the exception of primary school graduates who seem to be more responsive. Secondly, the results by age group show that with increasing age migrants become more responsive to labor market conditions but less responsive to distance. Thirdly, the results by migration motive indicate that the influence of networks is lower while the importance of employment conditions is larger in location preferences of those migrating for employment reasons. We find that students are discouraged by greater distance and that the deterring effect of distance is larger among those migrating for education purposes than migrants with other reasons of migration.[10]

These results indicate that location attributes factor into destination choices differently depending on migrant characteristics. Estimates of network effects also imply that selection into networks occur on multiple dimensions including education, age, and reason of migration. This result is important for studies that aim to estimate the effect of networks on future outcomes –for example, our results show that those individuals who are less inclined to work are more likely to choose destinations with larger networks and this selection needs to be addressed in studies assessing the role of networks on employment and earnings. There are a number of important policy implications of our results. In addition to changing the intensity of migration (Aydemir et al. 2021), increasing education levels in the country may result in dispersal of migrants that differ from past episodes. Similarly, changes in the age structure of the population imply changes in destination choices. These results are important for forecasting future migration flows and efforts to accommodate inflows of migrants.

[10] The discouraging effect of distance for those migrating for education purposes is in line with the findings from college student migration (e.g., Alm and Winters, 2009).

The rest of the paper is organized as follows: Section 2 discusses the econometric model of location choice and the relevant locational attributes incorporated to the model. We explain the data and estimation samples in Section 3. Section 4 discusses identification and some estimation issues. Section 5 presents the summary statistics. In section 6 we present the results from the main specification. In section 7 we test the sensitivity of the results to changes in the main specification. Section 8 presents heterogenous effects of migrant networks. We conclude in the last section.

2. Empirical Model

We estimate the impact of location attributes on internal migrants' location propensities based on a discrete choice model, following the methodology applied in Fafchamps and Shilpi (2013) and Jaeger (2007). We model the utility that an internal migrant i from source province s gets from choosing province j as follows:

$$U_{ij} = \alpha L_{sj} + \gamma_j Z_j + \varepsilon_{ij} \quad , \quad j = 1, 2, \ldots, J \tag{1}$$

where the stochastic utility function is linear in parameters and there are J possible destinations in a migrant's choice set. Note that source province s need not be the province of birth since some individuals may have moved since birth. Z_j includes destination fixed effects that controls for unobserved differences across alternative destination regions. L_{sj} is a vector of destination characteristics that varies by the source province s of migrant measured, for a given characteristic j, as the difference between the destination and source province.[11] The coefficient vector α is fixed across choices and across individuals. The reason for having the subscript s in the righthand side term is that the attribute of a destination will be viewed differently by migrants from different source provinces. For example, if the characteristic refers to the unemployment rate, then the corresponding variable in the model captures labor market conditions in the destination relative to the source, i.e., unemployment rate difference between destination and source. Migrants compare location attributes of possible destinations with those characteristics in their origin provinces and decide on the location choice based on a utility maximization problem.[12] The only variable that is not expressed as a first difference is the variable that measures the distance between the source and the destination. This variable also has a similar interpretation. Migrants take distance into consideration while making their location choice as higher distance is associated with higher transportation and psychic costs.

[11] "Source province" and "origin province" terms are interchangeably used in the paper and refer to the migrant's province of residence before migrating.
[12] Decision maker i chooses destination j over destination k if and only if $U_{ij} > U_{ik}\ \forall\ k \neq j$.

Since distance to a destination differs across source provinces, the utility associated with a particular destination varies by the source province of migrants. If ε_{ij} follows type I extreme value distribution and is independent and identically distributed over alternatives j, then the model parameters can be estimated using McFadden's (1984) conditional logit model. Let y_{ij} be a random variable that takes value 1 if individual i chooses province j, and takes value 0 otherwise. Then the probability of migrant i from origin s choosing province j is given by:

$$P(y_{ij} = 1) = \frac{\exp(\alpha L_{sj} + \gamma_j Z_j)}{\sum_{k=1}^{J} \exp(\alpha L_{sk} + \gamma_k Z_k)} \quad (2)$$

Equation (2) gives the likelihood function of an individual i choosing province j as the destination. Using the likelihood function, we can estimate the model parameters α and γ by maximum likelihood. The marginal effect of a change in a location's characteristic l_j (i.e., an element of L_{sj}) on the probability of a migrant choosing that location over others is calculated by taking the derivative of equation (2) with respect to l_j:

$$\frac{\partial P(y_{ij} = 1)}{\partial l_j} = [P(y_{ij} = 1)(1 - P(y_{ij} = 1))]\alpha_l \quad (3)$$

where α_l corresponds to the coefficient of location characteristic l_j in equation (1). Equation (3) implies that the marginal effect of a covariate varies with location j unless each alternative is equiprobable in being selected. Since location probabilities are unlikely to be equal, the marginal effect of a change in a location attribute depends on the location which presents difficulties in interpreting the estimation results. Therefore, we follow Jaeger (2007) in defining average marginal effect of a change in covariate l_j on $P(y_{ij} = 1)$ as

$$\frac{\partial \hat{P}(y_{ij} = 1)}{\partial l_j} = [\frac{1}{J}(1 - \frac{1}{J})]\hat{\alpha}_l \quad (4)$$

$1/J$ is the average location propensity assuming that migrants are equally likely to live in any given province. Since we model the preferences of internal migrants over alternative destinations conditional on migrating, each migrant is observed in only one location and this location is different than the province of origin. The sample contains 67 provinces which covers the whole country. Thus, each migrant has a choice set with 66 destination alternatives. By multiplying the conditional logit estimates with $\left[\frac{1}{J}\left(1 - \frac{1}{J}\right)\right] = \frac{1}{66}\left(1 - \frac{1}{66}\right) \cong 0.0149$, we will be able to interpret the resulting product as the average marginal effect of a change in a province's attribute on the probability of a migrant deciding to live in that province.

We include three migrant network variables into our specification following Jaeger (2007). Let M_{ij} denote the number of individuals born in province *i* who are residing in province *j*; P_j denote the population of destination province *j*. For an individual residing in province *i* considering province *j* as a potential destination, the first migrant network variable is defined as the share of the migrant's birth province in the population of the destination province $\left(\frac{M_{ij}}{P_j}\right)$; the second variable is the share of the migrant's birth province population that lives in the destination province $\left(\frac{M_{ij}}{M_{ii}+\sum_{j\neq i} M_{ij}}\right)$; and the third one is the share of the destination province population who are born in provinces other than the destination $\left(\frac{\sum_{i\neq j} M_{ij}}{P_j}\right)$.

These three variables capture different aspects of networks that may be relevant in migrants' destination choices. The first stock variable accounts for the relative size of the network. The previous literature interprets this network measure as capturing the extent of ethnic goods available to the migrant in a destination location (Bauer et al., 2007; Jaeger, 2007). A higher relative size may also indicate that the network plays a more prominent role in destination region's economy. The second stock variable captures the impact of the extent of information available to the potential migrant about housing, labor markets and livelihoods at a destination relative to other destinations. Since the denominator is the same for all destination provinces, this variable accounts for the impact of absolute size of the migrant's birth province group in a destination. Migrants may also prefer to move to destinations that are welcoming to migrants of all origins. It may be the attitudes of locals against migrants or the differential services offered to the migrants and their families that attract migrants to these provinces. The last stock variable captures these effects. In a sense it controls for the herd behavior where the herd consists of migrants of all origins (i.e., regardless of their birth provinces). Since our models include location characteristics as differences between destination and source provinces, the network measures in our empirical specification reflect strength of networks in the alternative destinations relative to the source province.

In addition to the network measures, we include several other controls that may influence destination choices. The utility levels that migrants would get in different locations are not observable. Thus, we need to think of the factors that determine a migrant's utility from moving to a location. Previous research (e.g., Bartel, 1989; Davies et al., 2001; Jaeger, 2000, 2007; Bauer et al., 2007) has shown that migrants respond to variations in economic conditions of locations. Since these differences are supposedly a key determinant that draws migrants to a location both historically and contemporarily, omitting labor market conditions of locations thereby differences in economic prosperity across alternatives, may bias estimates of migrant network effects. Thus, we include variables that capture differences in labor market prospects and amenities that different destinations offer as well as measures of migration costs. Turkish census data used in this study does not provide information on labor income which can be used to estimate average earnings levels in provinces. Therefore, to capture the level of economic activity, level of labor demand, and job opportunities in a location we use as proxies the province's population size and unemployment rate that are calculated from the census data. Controlling for unemployment rates also account for autocorrelated shocks to local labor markets, an important confounder of the relation between migrant networks and the location propensities of internal migrants (Bauer et. al., 2007).

Migration, either international or internal, incurs some substantial costs including travel costs, psychic costs of leaving beloved ones behind, information gathering costs and costs associated with accommodating to a new environment (Bartel, 1989; Jaeger, 2007). To proxy for migration costs, we take the straight-line distance (in kilometers) between the centers of destination province and province of origin. While the marginal cost of distance is positive, it may decline as the distance between the destination and the origin province increases (Davies et al., 2001; Jaeger, 2007). We thus include a squared distance term to allow the relation between location propensities and distance to be non-linear.

Amenities may also play a role in the location choice of migrants. There is mixed evidence in the literature, Cragg and Kahn (1997) and Fafchamps and Shilpi (2013) finding a significant role while Jaeger (2007) finds that the differences in amenities between alternative locations have almost zero effect on location propensities of migrants. While we do not have information on amenities that provinces provide, we account for the impact of amenities such as weather, total land area, being on the coast, the amount of resource endowments by including destination region fixed effects.[13]

[13] The destination region fixed effects, based on NUTS2 classification, groups provinces which are similar in terms of socioeconomic development and geography. Hence, region fixed effects capture NUTS2 level differences in amenities at alternative destinations.

3. Data and Estimation Sample

The Turkish Statistical Agency (TurkStat) makes publicly available the 5% random sample of the population censuses from years 1985, 1990 and 2000. The census questionnaire asks current place of residence, place of residence five years ago, place of birth, detailed questions about demographics such as age, gender, marital status, highest school diploma received, employment status and occupation. The 2000 Census adds to the set of questions in previous censuses the reason of migration.

We use 5% microdata from 1990 and 2000 to determine the internal migrants. Internal migration, in our context, is defined as the change in the province of residence during the past 5 years. Hence, if T denotes the census year then we call an individual as a migrant if the individual changed the place of residence between $t_0 \equiv T - 5$ and T. Note that, for an internal migrant, the place of origin at t_0 can be either the place of birth or another province if the individual has moved before t_0.

As a result of redefining geographic boundaries of administrative units, the number of provinces in Turkey increased over time from 67 in 1985 to 81 in 2000.[14] The new provinces were counties of the existing ones and became a province of their own. The formation of new provinces did not cause existing provinces to change names or cease entirely but resulted in a change in their borders. To have a consistent definition of geographic units across different data sets, we use the available county information in our data and adopt the border definitions of 67 provinces that existed in the first wave of our data. Hence, the observed actual migration to the latter founded 14 provinces are treated as directed to the provinces which initially contained them. An important reason for using province definitions that existed in 1985, as opposed to province definitions from 2000, is that place of birth variables in both 1985 and 1990 censuses only involve province of birth information but not county of birth. As a result, for individuals who were contemplating to move between 1985-1990 and between 1995-2000, we cannot create migrant network variables for the 14 new provinces if we were to adopt the 81-province definition.

[14] The number of provinces were 67, 73, and 81 in 1985, 1990 and 2000 censuses, respectively. Aksaray, Bayburt, Karaman, Kırıkkale became provinces in 1989 and were former counties of Niğde, Gümüşhane, Konya, and Ankara, respectively. Batman and Şırnak were separated from Siirt and became provinces in 1990. All these cities were created after the 1985 census and were first included as provinces in 1990 census. Bartın and Karabük were former counties of Zonguldak and became provinces in 1991 and 1995, respectively. Ardahan and Iğdır were separated from Kars in 1992. Yalova, Kilis, Osmaniye and Düzce were former counties of İstanbul, Gaziantep, Adana and Bolu, respectively. The latter set of new provinces were first introduced as provinces in 2000 Census.

Since large inflows of migrants to a location may affect province level characteristics such as unemployment rate and ease of access to amenities, location characteristics of destinations need to reflect the time period before migration takes place to avoid simultaneity problems. Therefore, for internal migrants in the 1990 Census who changed their province of residence within the last five years, characteristics of potential destinations prior to migration are constructed using the 1985 Census. Similarly, for internal migrants who moved between 1995 and 2000 destination characteristics are constructed using 1990 Census data.[15] Since the analysis focuses on determinants of location choice conditional on moving between provinces, we drop non-movers from the sample.

There are many motives for migration including marriage, finding a better job or getting education. The determinants of location choice may differ across migrants with different migration motives. For example, for those migrating as a result of marriage, which is common among women, it may be unrealistic to assume that all destinations are in the choice set. As these individuals join their spouses, destination may be pre-determined as the province where the spouse resides which leads to an empirical challenge in identifying the choice sets. In our main analysis we aim to focus on a more homogenous group for whom the migration motive is likely to be for employment purposes. We thus restrict our sample to 28-54-years-old male internal migrants who are less likely to be moving for education purposes or as tied movers.[16] The lower bound of the age restriction aims to restrict the sample to those who are likely to have completed their undergraduate level education.[17] To capture migration destination preferences of males with strong attachment to the labor market, we set the upper bound of the age restriction to 54 and drop those observations where respondent states the reason for not

---

[15] Since there is no Census conducted in 1995, we draw on 1990 Census for province characteristics.

[16] We exclude military personnel from the estimation sample since their locations are primarily determined by assignment mechanism of Turkish armed forces. Civil servants constitute another group who may not freely choose among the alternative destinations. Their location choices may be restricted to a subsample of the alternatives, and the personnel needs of the government department they work restricts their choices. Hence, estimation sample also excludes civil servants.

[17] In Turkey, schooling starts at age 6 and typically graduation from university requires 15- or 16-years of full-time education – high academic achievers graduating from university by age 23. For some individuals it may take longer to get Bachelor's degrees due to school repetition or repeated national exam taking that determines placement at universities. Others may continue their education beyond Bachelor's degree. Some students may move to other provinces for their undergraduate education. For these students choice sets are may also be limited due to their rank in national exam and the availability of programs across different provinces.

working in the last week as being retired.[18] In addition, the 2000 Census provides information on reason of migration that allows us to restrict the sample to those who indicate their reason of migration as "job search/start a job". Since the sample consists of males aged 28-54, provincial unemployment rates are calculated for 28-54-years-old males including both natives and migrants.

Our main analysis involves estimating a conditional logit model of location choice using pooled census data from 1990 and 2000, that controls for time invariant province characteristics (e.g., amenities) by incorporating destination region fixed effects. In this case, the model assumes that the parameters are constant over time. Some location characteristics may have gained importance over a decade, and some others may have lost influence on location decisions of migrants. For example, the unemployment rates increased substantially between 1990 and 2000 for both migrants and non-migrants. The national unemployment rate for 28-54-years-old male migrants was 6.8% and the corresponding figure for non-migrant counterparts was 4.9% in 1990. A decade later the unemployment rates increased to 13.8% and 12.4% for male migrants and non-migrants, respectively. The significant decline in the economic performance of the country during late 1990s may have influenced migration dynamics. Therefore, we also investigate the determinants of location choice separately for years 1990 and 2000 to examine whether the results are sensitive to the changes over time in relative economic and demographic conditions of locations. In the analysis where we use the 2000 Census data, we also expand our sample to migrants who state reasons of migration other than job search/start a job and study heterogeneity in the impact of networks on location choice by migration motive.

4. Identification and Estimation Issues

Source province characteristics matter in location choice of migrants. Migrants from a source with high-income level may view the income level of a destination differently than migrants from a source with low-income level. However, source province characteristics do not vary over alternative destinations; thus, cannot be included as separate regressors in the conditional logit specification. In order to account for differences in origin province characteristics of migrants, we adopt Fafchamps and Shilpi's (2013) method and create

[18] According to the labor force statistics from TurkStat, in year 2000 labor force participation (LFP) among males was above 90% for those 28 to 44, declining after this age to 83% among 45-49, 69 % among 50-54. Beyond age 54, due to early retirement provisions, LFPs are much lower, around 58% among 55-59 and 48% among 60–64-year-olds.

explanatory variables that correspond to the differences in location characteristics between destination and origin provinces. This requires symmetric responses to changes either in destination or origin characteristics and therefore is restrictive. To relax this restriction interaction of province dummies with source province characteristics could be included in the model; however, there are many alternatives and numerous origin characteristics which results in convergence issues and renders this approach infeasible.

Consistent estimation of model parameters in equation (1) relies on addressing omitted variable bias besides simultaneity issues. In our specification we include NUTS-2 level destination region fixed effects to control for unobserved economic and noneconomic differences between provinces that may affect migrants' location choices.[19] Using NUTS2 level fixed effects reduces the number of parameters to be estimated dramatically and helps to overcome the convergence issues encountered with province fixed effects models.[20] When we include NUTS2 fixed effects to the location choice model estimated with pooled cross-sectional data, the identification comes from both spatial and intertemporal variation in location characteristics within NUTS2 regions while in models estimated using a single cross-section the coefficients are identified by within NUTS2 region differences in covariates.

Almost half of the migrants in the estimation sample moved from an origin province that is different than their birth province which implies that for those migrants the choice sets contain their birth provinces. In fact, in pooled cross-sectional dataset around 29% of migrants are return migrants who choose to move to their birth provinces. These return migrants may do so because there may be unobserved birth province characteristics, such as need to take care of family enterprises, need to look after parents who suffer from a disease, negative experience with former migration moves, etc., which draw them back to their birth province. The choice of birth provinces for such reasons may confound our results. For example, if a significant

[19] The groupings of alternative provinces used in the study is provided by TurkStat and is at NUTS-2 level (a total of 26 regions). This grouping includes provinces that are similar in characteristics such as population, socioeconomic development level, geography, per capita GDP, per capita output in industry, agricultural output, and urbanization rate.

[20]As a better alternative, province fixed effects could be included in the specification which implies adding 65 alternative specific parameters to the model (i.e., there are 66 alternatives for each migrant and one province dummy is omitted from the regression equation to prevent collinearity). However, estimation with province fixed effects lead to convergence problems in maximum likelihood estimation. It may be the case that provinces with favorable observable characteristics also have favorable unobserved attributes, e.g., cities with greater perceived economic conditions may also provide better access to amenities, which may lead to collinearity between observable province attributes and province fixed effects—that account for omitted province characteristics—and result in convergence problems in estimation.

fraction of migrants chooses their birth provinces as their destination due to family related reasons and these provinces happen to be economically thriving regions we may falsely conclude that thriving regions attract internal migrants. Thus, in order to control for such unobserved birth-province pull effects, we include a birth province dummy which takes the value of one if the alternative is the birth province of the migrant; zero otherwise.

A potential threat to model specification stems from a possible variation in values attached to the attributes of destination provinces across individuals. If tastes over attributes of alternatives vary by individual's observed characteristics, conditional logit model can capture taste variations by including interactions between province attributes and observable characteristics. However, if tastes vary by unobserved characteristics of migrants, then the assumption of errors being independent and identically distributed over alternatives cannot be satisfied which leads to model misspecification (Train, 2009). In our context, such taste variations can arise, for example, if migrants with low adaptability to a new region—a trait that is unobservable in our data—have high costs to migrate to provinces where their migrant enclaves are not densely populated; hence, they may prefer provinces with a higher concentration of compatriots. The consequence of not being able to include an interaction term between individual adaptability and migrant network is the violation of iid assumption for the error terms.[21] To control for differences across migrants in the values and importance they attach to location attributes, we include individual fixed effects to equation (1) hence the name

[21] For clarification, simplify the model in equation (1) and let the utilities depend only on migrant networks in a location and the distance of the location to the place of origin plus a stochastic error term that varies both over individuals and locations: $U_{ij} = \alpha_i Mig_{ij} + \beta dist_{ij} + \varepsilon_{ij}$. The subscript i in the coefficient of migrant network variable is included to allow the effect of migrant network to vary over migrants. Assume that the variation in migrant tastes over provinces is partly explained by the variation in migrants' adaptability. We can then decompose the network effect into two: an average effect and a migrant specific component—a deviation around the mean that differs across individuals. Let $\alpha_i = \alpha + \gamma adpt_i$ where the latter term controls for the migrant's adaptability. When we plug α_i into the above equation we reach: $U_{ij} = \alpha Mig_{ij} + \beta dist_{ij} + \gamma adpt_i Mig_{ij} + \varepsilon_{ij}$. Since adaptation capability is unobservable in the data, the interaction term ends up in the new error term $\tilde{\varepsilon}_{ij} = \gamma adpt_i Mig_{ij} + \varepsilon_{ij}$. The new error terms for different alternatives are not iid. To see that: $Cov(\tilde{\varepsilon}_{ij}, \tilde{\varepsilon}_{ik} | Mig_{ij}, Mig_{ik}) = Cov(\gamma adpt_i Mig_{ij} + \varepsilon_{ij}, \gamma adpt_i Mig_{ik} + \varepsilon_{ik} | Mig_{ij}, Mig_{ik}) =$ $Cov(\gamma adpt_i Mig_{ij}, \gamma adpt_i Mig_{ik} | Mig_{ij}, Mig_{ik}) + Cov(\gamma adpt_i Mig_{ij}, \varepsilon_{ik} | Mig_{ij}, Mig_{ik}) +$ $Cov(\varepsilon_{ij}, \gamma adpt_i Mig_{ik} | Mig_{ij}, Mig_{ik}) + Cov(\varepsilon_{ij}, \varepsilon_{ik} | Mig_{ij}, Mig_{ik}) = \gamma^2 Var(adpt_i | Mig_{ij}, Mig_{ik}) Mig_{ij} Mig_{ik}$. The original error terms ε_{ij} and ε_{ik} are iid and orthogonal to location and migrant characteristics by assumption. However, since adaptation capabilities vary over migrants, the resulting term is not equal to zero. Conditional variance of the new error terms is shown to be: $Var(\tilde{\varepsilon}_{ij} | Mig_{ij}) = Var(\gamma adpt_i Mig_{ij} + \varepsilon_{ij} | Mig_{ij}) =$ $Var(\gamma adpt_i Mig_{ij} | Mig_{ij}) + Var(\varepsilon_{ij} | Mig_{ij}) = \gamma^2 Mig_{ij}^2 Var(adpt_i | Mig_{ij}) + Var(\varepsilon_{ij} | Mig_{ij})$. This sum varies over alternatives as migrant networks vary over alternatives. As a conclusion, being unable to account for individual tastes over location attributes may violate iid errors over alternatives assumption of the conditional logit model.

fixed effects conditional logit.[22,23] Individual fixed effects also help control for differences in utilities —that a migrant gets from choosing an alternative— that is due to source province characteristics since source province characteristics can be considered as traits of migrants that do not vary across alternatives (Fafchamps and Shilpi, 2013). Another approach to deal with random taste variations is to implement mixed logit model that allows values attached to location attributes to vary across individuals (i.e., allows individual specific coefficients of location attributes). This flexible model also allows for a full relaxation of the independence from irrelevant alternatives (IIA) assumption of conditional logit model, however, the trade-off is a substantial loss in computing time (Davies et al., 2001; Christiadi and Cushing, 2007; McFadden and Train, 2000).[24] We test the sensitivity of conditional logit results to possible violations of IIA assumption (either through unproportionate substitution patterns or random taste variations) by running regressions using mixed logit model. The results are similar which suggests that IIA assumption holds in our context.[25]

We also consider the potential impact of the increase in the number of provinces on our estimation results. Job vacancies in the new local government positions and possible boom in economic activity might have induced high in-migration to the 14 new provinces that were formerly counties. Put it differently, the creation of new provinces may act as positive labor market shocks and increase the propensity of individuals to move to these localities which will be observed in our data as migration flows to the provinces that include the formerly 14 counties. The resulting pull impact of new job opportunities in newly created cities varies across alternatives and time in a way that NUTS2 region fixed effects cannot control.[26] A possible

[22] Estimation of alternative-constant determinants of a discrete choice model is infeasible under conditional logit framework. Therefore, handling of variability in alternative-constant individual characteristics that may bias the estimation results is achieved by implementing a panel fixed effects estimator in Stata®, namely `xtlogit` with `fe` option. Although we cannot estimate the impact of alternative-constant individual characteristics, controlling for them alleviates the concerns for having omitted variable bias.

[23] A crucial point is that for individual fixed effects to address omitted variable bias due to heterogeneity in unobserved individual characteristics that results in taste variation, the unobserved individual traits should be constant across alternative destinations.

[24] The assumptions of the mixed logit model allow disturbances to be correlated or heteroskedastic (or both) over alternatives which adds the model the capability to account for non-proportional substitution patterns and implies full relaxation of the independence from irrelevant alternatives assumption (Davies et al., 2001; Christiadi and Cushing, 2007). Mixed logit model is less restrictive compared to conditional logit and is flexible enough to approximate any random utility model (McFadden and Train, 2000); however, computationally it is much more intensive (Christiadi and Cushing, 2007).

[25] The mixed logit estimation results are available from authors upon request.

[26] If new city formation has the same influence on location choices of migrants in late 1980s and late 1990s, then destination province fixed effects would capture these impacts. Though, it is a strong assumption to believe in since the pull impact of city formation is most likely high in the short run when the city is newly founded and loses

correlation between the unobserved labor market shocks to (in terms of pull factors to 14 new provinces) and location characteristics of the provinces that yielded the newly created ones may bias the estimates; hence, needs to be accounted for in the analysis. In regressions using 1990 cross-section of migrants, we include a dummy variable which takes the value of one for the alternatives (potential destination provinces) that included counties which later became provinces between 1985 and 1990; zero otherwise. In regressions using 2000 cross-section of migrants, we similarly include a dummy variable which takes the value of one for the alternatives which included counties that became provinces between 1990 and 2000; and zero otherwise.[27] For the analysis using pooled data from 1990 and 2000 censuses, we allow the impact of potential job creation due to new province formation to vary over time by including interaction terms between census-year dummies (for 1990 and 2000) and the dummy variable for the group of alternatives that include counties which became provinces in between 1985 and 1990; and a separate interaction term between census-year dummy for 2000 and the dummy variable for the group of alternatives that include counties which became provinces after 1990.[28]

The estimation method of this study creates patterns of positive and negative correlations in error terms across alternatives for a given migrant (see Fafchamps and Shilpi, 2013). Correcting for the interdependence across observations for a given migrant requires clustering standard errors at individual level (Cameron and Miller, 2015). Nevertheless, if the destination choices of large outflows of migrants from the same origin province are correlated, then clustering standard errors at individual level would not be enough to achieve correct test sizes. Similarities in the characteristics of migrants originating from the same province that are unknown to the researcher may result in a correlation among migrants' location choices. To correct for this possibility, we cluster standard errors at origin province level.[29] This results in

---

its strength as time passes (as available job positions are getting occupied). Almost half of the cities were created before 1990; hence, it is possible to observe a declining impact of city formation process on location propensities over time. Furthermore, we are unable to include destination province fixed effects due to convergence issues.

[27] There is an implicit assumption in this method which implies that there is no impact of new job opportunities that were available during the city formation process for cities founded before 1990 on location choices of migrants who had moved between 1995 and 2000. We test with adding another dummy to the model specification for location choices of 2000 cross-section of migrants which takes value one for alternatives that included counties that became provinces in between 1985-1990 and takes value zero otherwise. This estimation yields similar results.

[28] Consistent with our argument omission of border change controls results in attenuation of estimates of demographic and labor market measures toward zero while the migrant network effects stay unchanged. Most of the new provinces were counties of provinces with relatively small population sizes and poorly performing labor markets. The sensitivity of the labor market estimates to the border change controls suggests that in the absence of new city creations the chances that individuals choosing the provinces embracing the latter founded provinces as migration destinations would have been lower.

[29] Clustering standard errors at origin province level encompasses clustering standard errors at individual level and also accounts for serial correlation in within-origin-province errors.

67 clusters of observations and abstracts from concerns about achieving incorrect test sizes due to few clusters (Angrist and Pischke, 2008).

5. Descriptive Statistics

Our estimation sample includes male internal migrants, i.e., individuals between the ages of 28 and 54 who changed their province of residence in the last five years, and excludes students, retirees, civil servants, and members of Turkish armed forces. The main estimation sample further restricts the sample to those who indicate their reason of migration as "job search/start a job" in the 2000 Census data. There are 25,325 and 16,655 individuals meeting these sample restrictions in the 1990 and 2000 Censuses, respectively.[30]

Table 1 presents the descriptive statistics for the variables from the 1990 and 2000 samples. We report average (over migrants) of the differences in a variable between destination and origin provinces; that is, the variables in Table 1 are of the form $\Delta_{ij}^{x} = x_j - x_i$ where x is the choice attribute, i is the province of origin and j corresponds to one of the remaining 66 provinces. For example, let x measure unemployment rate in a province; for actual destination of a migrant, Δ_{ij}^{x} is the difference in unemployment rates between the chosen province and origin province of the migrant. We present the average of Δ_{ij}^{x} over all migrants for the actual destination in columns (1) and (4). Columns (2) and (5) on the other hand present the average of Δ_{ij}^{x} over all migrants for the remaining 65 alternative destinations – i.e., those provinces that were in the choice set but were not chosen. The statistics for the distance variable, which corresponds to the straight-line distance (in kilometers) between the centers of destination and origin provinces, are computed similarly for the actual and alternative destinations. Columns (3) and (6) present t statistics for the tests of equality of means presented by the preceding two columns.

[30] The information regarding migration within last five years is available for individuals who are at least 5 years old. There are a total of 2,864,207 observations in the 1990 Census and 3,444,456 observations in the 2000 Census data without any sample restrictions. In both Censuses birth province is known for any individual regardless of their ages – in 1990 data 598,067 individuals lived in a province which is different than their province of birth while in 2000 the corresponding number was 854,502. The number of individuals who changed their province of residence within the last five years is 204,949 and 239,727 in 1990 and 2000 Censuses, respectively. In our analysis we drop foreign-born individuals and individuals whose place of residence five years ago was another country from our estimation sample. Restricting the sample to males between the ages of 28 and 54 and excluding students, retirees, civil servants, and members of Turkish armed forces leads to sample sizes of 25,325 and 16,655 respectively for 1990 and 2000. The difference in the number of internal male work migrants between 1990 and 2000 censuses does not reflect a change in the migration intensity but is due to being able to better identify work migrants in 2000 Census thanks to the introduction of migration motives question.

Results in the first row of columns (1) and (4) show that both in 1990 and 2000 migrants move to provinces where migrant's birth province group constitutes a smaller fraction of population compared to the fraction in origin province. In both censuses for more than half of the migrants origin province was the birth province; thus, this result is expected. Columns (2) and (5) show that in the alternative destinations these differences are even larger, as birth province group constitutes, on average, much smaller share of the populations in alternative destinations. These results imply that migrants prefer to migrate to provinces that have higher percent of population that is from migrant's birth province (i.e., a higher relative size). The same pattern applies for the migrant network variable that measures the percent of migrant's birth province population that is living in a destination: migrants move to destinations that contain more people who share the same birth province with them. In both cross-sections, foreign-born (i.e., out-of-province born) individuals, on average, make up a larger share of actual destination populations relative to province of origin. Alternative destinations, on average, are less dense in terms of foreign-born settlers relative to origin province. Columns (3) and (6) show that in both cross-sections the differences in all three network variables between actual and hypothetical destinations are large in magnitude and statistically significant at 1% level.[31, 32]

Unemployment rate (UR), the first labor market attribute in Table 1, shows that, on average URs are higher in actual destinations than in the origin province for 1990 cross-section but this pattern reverses in 2000. In the Turkish context typically UR in agricultural sector is lower than in other sectors and most internal migrants move to urban centers where economic activity centers around non-agricultural sectors. Due to historically high prevalence of agricultural sector in Turkish economy—the share of agricultural sector in total employment is 34% in 1985 and 29% in 1990—the mean UR of a province may not be a good indicator of economic conditions in urban areas. Therefore, we also report statistics for nonagricultural employment rate (ER) and nonagricultural UR. These alternative measures of labor market performance in Table 1 reveal that, relative to both origin and alternative provinces, migrants prefer provinces with a higher nonagricultural employment rate (ER) and a lower

[31] Indeed, the differences in all explanatory variables between actual and alternative destination provinces are statistically significant at 1% level.
[32] The mean (sample standard deviation) of the three network variables for the places of residence of individuals observed in the pooled Census data are 17.8 (34.1), 19.4 (30.8), 30.9 (20.6) respectively for the first, second and third network variables.

nonagricultural UR in both years. This result is in line with migrants preferring locations that offer better labor market opportunities outside the agricultural sector.

Actual destination choices of migrants, on average, are more populous and denser than both origin and other alternative (nonchosen) provinces in both cross-sections. Destination population sizes are 45-to-55% larger than origin province and alternative destination populations are 43-to-44% smaller in 1985 and 1990, respectively.[33] Actual destinations are closer to migrants' origin provinces than alternative destinations and this difference is large– around one third of a standard deviation.

To sum up, migrants seem to prefer to move to provinces with a higher concentration of same birth province individuals relative to alternative provinces; a higher nonagricultural ER and a lower nonagricultural UR relative to origin and alternative destinations; a larger population and higher population density relative to origin and alternative destinations; and a shorter distance to the origin province relative to alternative provinces.

6. Results

Table 2 presents the coefficient estimates from our main specification for 28-54-years-old male migrants using fixed effects conditional logit to estimate the determinants of migrants' location choices. The specification includes linear and quadratic terms for three network measures, controls for labor market attributes and migration costs proxied by a quadratic function of straight-line distance between destination and origin provinces. Individual fixed effects, NUTS-2 level destination fixed effects, a dummy variable that captures cases when destination province is the same as birth province are included to control for destination and individual specific unobserved factors discussed in Section 4. Lastly, we include dummies to control for the differential labor market shocks that provinces with border changes might have experienced. Standard errors are clustered at origin province level which encompass clustering at individual level as well. Clustering standard errors at origin province level allows us to correct for the interdependence across alternatives for a given migrant.

[33] The average percentage difference in population sizes between actual destination and origin province is calculated by the formula: $(e^b - 1) * 100$; where b corresponds to the mean difference in log population sizes between actual destination and source province of migrants. The calculation of average percentage difference in population sizes between alternative provinces and origin is analogous.

Column 1 of Table 2 presents results from the pooled 1990 and 2000 cross-sections. Columns 2 and 3 present results using 1990 and 2000 cross-sections separately, respectively. The pooled estimation sample consists of 41,980 28-54-years-old male migrants with a choice set consisting of 66 alternatives resulting in 2,770,680 observations. The results show that conditional on controls included in the model, migrants are more likely to choose locations in which their compatriots are more highly concentrated compared to alternative locations. For two of the three migrant network variables there are diminishing returns to the size of the networks: migrant's birth city share of province population and province's share in migrant's birth city population have inverted U shaped effects on the probability of choosing a particular province. The inverted U shape implies that benefits of increasing network size are outweighed by costs beyond a certain level. For example, with increasing network size migrants get more contacts in the job market that may facilitate their job search. Increasing network size may, however, also imply more competition for jobs. Beyond a certain network size marginal benefit of increasing network size may thus fall short of the marginal cost. Albeit, in our regressions we do not detect such adverse effects of migrant networks on location propensities.[34]

Estimated coefficients for the first network measure imply that marginal effect of a 10 percentage points increase, evaluated at the sample mean of destination-origin difference in migrant's birth place share in province population, results in a 6.5 percentage points increase in the probability of locating in that particular destination.[35] For the second network measure, the marginal effect of a 10 percentage points increase, evaluated at the sample mean of the difference in shares of migrant's birth place population in destination and origin, leads to a 0.5

[34] Considering the results from the pooled sample, an increase in a migrant's birth place share in destination population is associated with an increase in the probability of choosing that destination over alternatives up to the point where the difference in the first network variable between destination and origin reaches approximately 88%. For differences larger than 88% the relation between location propensity and the size of the migrant enclave is positive but the effect of the enclave is declining. Only around a difference of 176% the effect becomes negative. The probability to choose a destination increases in the destination's share in migrant's birth place population up to a difference in the second network variable between destination and origin of size 69%; beyond that point the network effect starts declining. When the difference exceeds 138% the effect of the share of birth place population in destination on choosing that destination becomes negative. Since there is a convex relationship between location propensity and the share of foreign-born in destination and the ranges in our sample for the first and second migrant network variables are respectively (-96.28%, 96.25%) and (-92.97%, 92.97%), we do not find evidence for the adverse effects of migrant clustering on location choice in our sample.

[35] The average marginal effect of migrant network variables and distance, evaluated at the sample mean of variables, is calculated by the formula: $\frac{\partial \hat{P}(y_{ij}=1)}{\partial l_j} = [\frac{1}{J}(1-\frac{1}{J})](\hat{\alpha}_l + 2\hat{\theta}_{l^2}\bar{l}/100)$ where $\hat{\alpha}_l$ and $\hat{\theta}_{l^2}$ are the coefficient estimates of the linear and quadratic terms, respectively. $\bar{l}$ is the sample mean of the independent variable. The sample mean of the difference in first network variable between destination and origin is -48.67 while average probability of a location ($\frac{1}{J}$) is 0.0149.

percentage points increase in the probability of a migrant choosing that destination.[36] Among these two migrant network measures, the marginal effects are larger for the first one. This suggests that relative size of the network in the destination plays a more important role than its absolute size. The coefficient estimates for the third network measure imply that increasing foreign-born share increases the chances of a migrant to settle in that destination at an increasing rate. All migrant network variables and their squared terms have strongly significant coefficient estimates. In a nutshell, these results imply that other things constant the migrant's birth city share, the percent of migrant's birth city population living in the actual destination, and percentage of foreign-born individuals are higher in destination chosen by migrants than in alternative destinations. Stronger networks attract migrants which may be operating through reduced costs of migration, better labor market prospects, and higher availability of source region specific goods and culture that networks offer.

Unemployment rate has the expected negative sign and is statistically significant at 1% level in column (1) which implies lower unemployment rate of a destination works as a pull factor. A one percentage point increase in the destination-origin difference in unemployment rates reduces, on average, the probability of migrating to that destination by 0.21 percentage points.[37] This result implies that lack of job opportunities is a significant deterrent for migration decision. Migrants also prefer to move to relatively more populous provinces. Provinces with large populations may provide both a larger variety of jobs and more job opportunities for a given job type. In larger cities the number of institutions that provide job search assistance, such as government and private job search agencies, may be higher and easier to access which reduces the search costs of potential migrants and increases the returns on job search activities. In addition, migrants may benefit from their enclaves more in larger cities. The help from migrant's compatriots is more likely to yield a desired job position for the potential migrant in cities with abundance of job opportunities. The results on unemployment rate and population size together suggest that migrants take into consideration the economic conditions of locations while deciding on the migration destination. One standard deviation increase in the destination-origin difference in log population sizes, on average, increases the likelihood of migrating to that destination by 1.4 percentage points.

[36] The sample mean of the difference in second network variable between destination and origin is -43.73.
[37] The average marginal effect of unemployment rate is calculated according to equation (4).

Distance has the expected negative sign and the impact is precisely estimated. The coefficient on the squared distance has positive sign and is strongly significant which implies that distance has a deterrent effect on probability of migrating to a location but this effect is non-linear with a U shape. That is, as distance between destination and origin increases, the probability of choosing that destination decreases at a decreasing rate. The turning point is reached at a distance between destination and origin province of around 870 kilometers, and the impact becomes positive around 1750 kilometers. However, the sample does not include destination-origin pairs which are as distant as 1750 kilometers; hence, over the sample range distance has a negative effect on migrants' propensity to choose a location. The average marginal effect of a one standard deviation increase in the distance between destination and origin, evaluated at the sample mean of distance between alternative and origin provinces, is a 0.6 percentage points decrease in a migrant's likelihood to locate in that destination.

Columns 2 and 3 for 1990 and 2000 cross-sections imply coefficient estimates that are in line with the pooled sample. The inverted U-shaped relation between location choice propensities and the relative and absolute size of the migrant network remains, although the coefficients of the absolute size of the migrant network are imprecisely estimated when using 2000 cross-section of migrants. Unemployment rate differences between destination and origin appear to be more influential on 1990 cross-section migrants' than 2000 cross-section migrants' location choices.[38]

7. Robustness Checks

Table 3 presents the results of the sensitivity checks from alternative specifications using pooled sample. The main conclusion from this analysis is that migrant networks have a robust effect in determining migrants' destination choices.

[38] Starting with mid 1980s terrorist activities emerged in Southeast Turkey that intensified in 1990s. Although its magnitude cannot be quantified, this security situation may have induced migration away from terror inflicted regions. These individuals may have prioritized destination location characteristics other than the unemployment rate. We would like to note two points related to this issue. First, the 2000 sample that corresponds to column (3) is restricted to those who indicate their reason of migration as "job search/start a job" hence unlikely to include individuals migrating due to security concerns. We also estimate the model separately for those who indicate in the 2000 Census the reason of migration as "Safety issues" and present those results in the section on heterogenous effects. Moreover, although for the results presented in Table 2 we control for NUTS2 level fixed effects that may capture region specific factors we also check robustness of results to an alternative specification. In particular, the potential effects of terrorism call for individual specific coefficients of unemployment rate which is satisfied by a mixed logit specification. We estimated our model for the pooled sample with a mixed logit which results in coefficient estimates for unemployment rate that are very similar to the conditional logit specification. Due to convergence issues we could not estimate the same model specifically for the 2000 census.

In the first column we experiment with omitting destination region dummies. Except for a small attenuation of demographic and labor market effects, the results are unchanged. The slight change in the estimated effects of the demographic and labor market attributes suggest a confounding relation between unobserved destination characteristics and migrants' location choices which is tackled by the introduction of province-group dummies to the model.

In columns (2) to (5) we include alternative demographic and labor market measures. Replacing unemployment rate variable, column (2) controls for nonagricultural employment rate while column (3) controls for nonagricultural unemployment rate. Coefficient estimate for the nonagricultural employment, although precisely estimated, is close to zero. The coefficient estimate for nonagricultural unemployment rate is in line with the specification including aggregate unemployment rate suggesting that higher unemployment at destination deters migration. Column (4) replaces population size with population density while column (5) controls land area in addition to population size. The results imply that migrants prefer to move to less dense locations.[39]

Industrial cities in the country including Istanbul, Ankara, Izmir, Kocaeli and Bursa not only have large population shares of foreign-born but also account for large shares of migrants' birth place populations across Turkish provinces. In column (6) we exclude the third migrant network measure (Foreign-born share of province population) and estimate our model which yields migrant network effects that are similar to column (1) of Table 2 which mitigates concerns that collinearity may be affecting estimates of network effects.

In column (7) we only use the destination characteristics as regressors rather than destination-origin differences similar to Jaeger (2007). The migrant network effects remain qualitatively the same in this alternative specification. Lastly, we estimate our model by including non-migrants to the estimation sample. If non-migrants respond differently to location attributes in the case of migration than the migrants, then by examining only the migrants' responses we may compromise the external validity of the findings due to sample

[39] In the pooled sample, rural-out migration is less frequent to city centers with a rate of 41%. 53% of migrants from county centers prefer city centers as destination while this rate is 57% for migrants from urban locations. This is in line with urban-urban migration substituting rural-urban migration beginning in 1980s in Turkey (Terzi and Koçak, 2014).

selection bias. In this specification with non-migrant males, choice sets of individuals include their source provinces. Therefore, every individual has the same choice set that consists of all 67 provinces of Turkey. Davies et al. (2001) discuss that if an individual does not move and chooses current province among the set of alternatives, there may exist substantial differences between staying and moving due to unobserved costs associated with moving which cannot be controlled by region fixed effects. Following Davies et al (2001), we include a non-migration dummy which takes the value of one if the current province of residence is chosen and zero otherwise. The estimated coefficients in column (8) are similar to the main specification with the exception that the results for the extended sample imply a convex relationship for the first network measure. Coefficient estimate for the non-migrant dummy is large and significant implying that unobserved differences between staying and moving are important in location choice.

8. Heterogeneous Effects

The benefits that migrant enclaves offer may differ across migrant types. The literature on international migration, for example, provides evidence that immigrants' host country language ability plays a role in sorting across locations. Locations with large enclaves where there is less need to learn the local language attract immigrants with lower language ability while immigrants with high language proficiency choose locations with a small enclave (Lazear, 1999; Bauer et al., 2005). Jaeger (2007) finds that the magnetic effects of immigrant concentrations is more pronounced for the location choice of refugees than other types of immigrants. Similar to the international migration context, in the case of internal migration some demographic groups may benefit from networks more than others. Lower educated and younger individuals may lean on migrant networks in destination regions as support provided by former compatriots —in terms of finding a job and adapting to the environment— may be very beneficial. On the other hand, more educated and older cohorts may already possess skills, experience and contacts in job market that reduces potential benefits of migrant networks.

In Table 4 we examine the heterogeneity in migrant network impacts using subsamples defined by a given migrant characteristic. For example, considering age we divide our full sample (28-54 years-old) into distinct age groups and define five subsamples of 28-33, 34-38, 39-43, 44-49, and 50-54. We then run separate fixed effects conditional logit regressions of migrants' location choices on each subsample.[40] The analysis on education and age groups uses 1990 and 2000 pooled sample of migrants defined for our main analysis to preserve consistency with prior results. For the analysis of migration motives, we use the 2000 Census where information on motives is available and extend the sample to include individuals who report motives for migration other than "job search/start a job". We suppress the coefficient estimates of population size and of quadratic terms for migrant network and distance variables to save space. Nevertheless, Figures 1 and 2 depict the complete relation between location propensities and migrant networks for each subsample in Table 4 (holding other factors constant) by drawing the average marginal effect line of a 10 percentage points increase in each migrant network variable on location probabilities.[41]

The first panel of Table 4 presents results by education groups. The results show that the propensity of migrants to choose locations with larger migrant networks decreases with increasing education level. This negative relationship between network size and education is observed for the first two network measures while for the third network measure there is no clear pattern. This is depicted by the top panel of Figure 1 where the predicted marginal effect curves lie lower for higher education groups. As the relative size of the network at destination becomes smaller (more negative values on the x-axis), the magnetic effect of networks increases for each education group but much more so for the lower educated. As a result, the differential effect of networks across education groups becomes magnified in locations with small network sizes relative to the origin. These results indicate that networks benefit lower educated migrants more than higher educated migrants. This differential may stem from available job

[40] An alternative to estimating models separately for subsamples we could estimate a fully interacted fixed effects conditional logit model where each covariate in the model is interacted with a set of dummies capturing a given migrant characteristic (e.g., a set of dummies for each reason of migration). This leads to a substantial increase in the number of parameters to be estimated (e.g., for reason of migration, 50 additional parameters just for covariates related to location attributes and many more for the fixed effects) and convergence problems renders this estimation impossible. An alternative to a fully interacted model is a partially interacted model where only location attribute covariates are interacted with migrant characteristics but other covariates, such as region dummies, are not interacted. We prefer subsample analysis to a partially interacted model because the latter forces the effects of remaining covariates to be the same across different groups. This is an important limitation. For example, the effect of amenities on destination choice captured by destination fixed effects may be very different for different age groups or migrants moving for different reasons. Thus, we report the results from the subsample analysis and briefly refer to the results of partially interacted model in the remainder of this section.

[41] The average marginal effect line of a 10 percentage points increase in migrant network is drawn based on the following formula: $10 * \frac{\partial \hat{P}(y_{ij}=1)}{\partial l_j} = 10 * [\frac{1}{66}(1-\frac{1}{66})](\hat{\alpha}_l + 2\hat{\theta}_{l^2} l/100)$. $\hat{\alpha}_l$ and $\hat{\theta}_{l^2}$ denote the coefficient estimates of the linear and quadratic terms for migrant network variables and l, which represents the migrant enclave, takes values in the range of the estimation sample. Recall that the covariates in conditional logit regressions record differences in location characteristics between destination and origin; thus, the sample range for migrant enclaves include both positive and negative numbers.

opportunities transmitted by networks being predominantly lower skilled. Higher educated may also be receiving relatively more job offers from channels other than the birth place network. In addition, networks may offer shelter at initial stages of migration which may be especially valuable for lower educated for whom migration costs may be more binding while the need for housing support may be less among higher educated.

The first two figures in panel (a) reveal that among the three migrant network variables the most effective in location choice is the one which captures the relative size of the network that proxies for the prominence of the network in destination economy and availability of ethnic goods and culture. The marginal effects for the share of province population by birth place and province share of birth place population mostly lies above the zero line which implies that in the estimation sample the adverse impacts of migrant enclaves that may stem from increased competition among compatriots for job opportunities are not apparent.

Estimates by education groups in Table 4 also show that higher educated migrants move longer distances. This may be because higher educated conduct job search over a larger geographic market. The benefits of search in more distant markets may result in returns that compensate larger migration costs for this group. Education groups respond to unemployment rate differences similarly with exception of primary school graduates who seem to be more responsive than the other groups.[42]

Results by age groups are presented in the second panel of Table 4. These results show that younger migrants choose locations that are closer to the origin. Put it differently, the impact of distance as a deterrent to migration becomes less with increasing age. Older individuals may have moved previously and therefore possess higher migration capital. This may reduce their cost of migration and enable them to choose locations that are more distant to the origin. The

[42] The discussion of differences across subsamples presented in Table 4 should be interpreted with caution since we cannot test whether these differences are statistically significant. Testing statistical significance of reported differences would require estimating a fully interacted fixed effects conditional logit model for which we encounter convergence problems as discussed in the first footnote of this section. As a partial solution, we estimate partially interacted fixed effects conditional logit model. For example, for the differences across education groups, we interact only the location attribute covariates with dummy variables that correspond to five education groups but do not interact remaining covariates such as region dummies with education dummies. Note that the coefficient estimates from this partially interacted model are not identical to the coefficient estimates reported in Table 4 – identical estimates would require a fully interacted model. Partially interacted model produces estimates that lead to the same qualitative conclusions discussed based on Table 4. The differences we discuss across education groups are found to be statistically significant with one exception: although all three network measures imply that more educated put less weight on networks, only significant difference emerges for the second network variable.

results by age group also show that migrants become more responsive to unemployment rate as age increases.[43] This may be because migration to longer distances enables older individuals to choose over a larger set of locations with a larger variety of employment conditions. Panel (b) of Figure 1 demonstrates differences in network characteristics of locations by age group. There is no significant heterogeneity in terms of first two migrant network measures and a modest difference in the third one suggesting that older individuals are more likely to choose locations with a larger share of foreign-born population.

Bottom panel of Table 4 estimates models by migration motive. The 2000 Census distinguishes between the following migration reasons: job search/start a job, transfer to a different province (including mainly public service employees such as teachers rotating between provinces), tied mover, education, earthquake, and safety issues. In August 1999 there was a major earthquake of 7.4 magnitude that hit northwestern Turkey with a death toll of around 18,000 people and many more homeless. Following the earthquake many people migrated from the severely affected regions. During mid 1990s there was also an escalation of terrorist activities in Southeastern Turkey which led to movement of people away from regions with security concerns. These reasons for migration are captured by the 2000 Census by earthquake and safety issue categories respectively. In our estimation sample those who move for employment reasons (job search/start a job) are 57.8% of movers, followed by transfers 11.4%, tied movers (migration dependent on a household member or due to marriage) 4.5%, earthquake 4.5%, education 2.9%, and safety reasons 0.7%.[44]

The results by migration motive show that economic prospects of destinations are a significant factor only for those migrating for employment reasons. Distance is a deterrent for all migration motives with a larger coefficient for those migration for education purposes. The negative effect of distance among those moving for education purposes, most likely for postsecondary education, is in line with the literature on college student migration that also reports a discouraging effect of distance (e.g., Alm and Winters, 2009). Our results further show that individuals who are moving for education purposes have a stronger preference relative to other migrants for locations closer to the origin where their families are likely to be residing.

[43] In a partially interacted fixed effects conditional logit model where only location attribute covariates are interacted with age group dummies, different effect of distance by age is found to be significant at 5% level but there is no significant difference in response to unemployment rate by age.
[44] Among migrants 16.5% state “other reasons” for migration while the unknown category is 1.4%.

Among the network measures the first and third network measures have statistically significant effects in location preferences across all migration motives while the second network measure is significant only for those relocating due to the earthquake. As Figure 2 shows the largest effects are observed for the first network measure and estimated effects for those migrating for employment reasons is smaller compared to tied movers and those moving due to safety reasons. The magnetic effect of the network is largest for those migrating due to safety reasons who are pulled towards provinces with a larger share of compatriots born in the same birth place. The effect is lowest for those moving due to earthquake for whom priority in location choice might have been locations where close relatives live.[45] The heterogenous effects of networks by demographic groups and reason of migration reveal that all migrants benefit from the positive externalities associated with migrant enclaves, even those who migrate for education purposes or due to transfers.[46]

9. Conclusions

This paper estimates the determinants of location choice conditional on migration within a fixed effects conditional logit framework. We focus on the impact of birth province migration networks on destination choices of internal migrants along with the role of other location attributes including labor market and population characteristics and distance from origin. The data used for estimation comes from three rounds of Turkish population censuses –1985, 1990 and 2000– and the sample is restricted to 28-54-years-old male migrants. We also estimate the models for subsamples by skill group, age at migration, and reason of migration in order to investigate potential heterogeneities in the role of these factors in shaping destination choices.

The results from our main specification show that internal migrants respond to differences in migrant networks, labor market and population attributes between locations while deciding on the migration destination. We use three measures of network variables that capture

[45] In a partially interacted fixed effects conditional logit model where only location attribute covariates are interacted with dummies for reason of migration, we find that economic prospects of destinations have a significant effect for employment migrants but not for other groups; deterring effect of distance is larger for those moving for employment reasons but the difference is not significant at 10% level; the difference in the effect of first network variable between employment migrants and those moving as tied migrants or due to safety reasons is significant at 1% level.

[46] The group that migrates to get education includes parents that follow their children in their pursuit of education. Therefore, we also estimated our model restricting the sample to at least high school graduates who are more likely to be the student rather than parents (who tend to be lower educated). The results are quantitatively similar. The results concerning migration due to transfers may be driven by the preference of public employees to return to their hometowns after completing their rotations in other provinces.

different aspects of network characteristics in alternative destinations. The results on migrant network variables together imply that migrants are drawn to cities in which their former compatriots and former migrants from all source provinces are highly concentrated. Differences in population size and unemployment rate have the expected impact on migrants' location probabilities and imply that migrants are more likely to move to cities in which economic conditions are relatively better. Distance between destination and source province is shown to be a significant deterrent of migrant's location choice. These estimates of the effects of location characteristics on migration destination choice may help local administrators and policy makers to predict future waves of migration and plan for services and infrastructure for upcoming migrant flows.

Estimation results in subsamples reveal significant heterogeneity in the effects of location attributes on destination choice. The results indicate that with increasing education, migrants move longer distances and the impact of networks on location choice decreases. Migrants with different education levels respond to unemployment rate differences similarly with the exception of primary school graduates who seem to be more responsive. With increasing age migrants are found to be more responsive to labor market conditions but less responsive to distance. The results by migration motive, on the other hand, show that the influence of networks is lower while the importance of employment conditions is larger in location preferences of those migrating for employment reasons. We also find that migrants who move as a result of push factors such as earthquakes or terror attacks are particularly attracted to locations with a higher concentration of former compatriots. For this group of migrants, economic prospects of destination do not appear to be an important determinant of location choice. These results highlight the importance of distinguishing between motives of migration in understanding the role of networks.

The differences in the effects of networks on destination choices of migrant types imply that the role of beneficial network externalities, such as job referral, information provision, cultural support may also differ by migrant characteristics. This also calls for a better understanding of which support mechanisms matter most for different migrant types. With this understanding, policies that aim to help internal migrants may be tailored towards the needs of migrants with different characteristics. The results in this paper are also relevant for studies that are interested in the effect of migrant/ethnic enclaves on economic outcomes such as employment and wages. Our results imply that migrants self-select into enclaves by reason of migration, age, and education and these patterns matter for proper identification of enclave effects.

References

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of education review, 28*(6), 728-738. https://doi.org/10.1016/j.econedurev.2009.06.008

Angrist, J. D., & Pischke, J. S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Aslund, O. (2005). Now and forever? Initial and subsequent location choices of immigrants. *Regional science and urban economics, 35*(2), 141-165. https://doi.org/10.1016/j.regsciurbeco.2004.02.001

Aydemir, A., Kırdar, M. G., Torun, H. (2021). *The Effect of Education on Geographic Mobility: Incidence, Timing, and Type of Migration*. IZA Discussion Papers, No. 14013, Institute of Labor Economics (IZA), Bonn.

Banerjee, B. (1983). Social Networks in the Migration Process: Empirical Evidence on Chain Migration in India. *J Dev Areas, 17*(2), 185-196.

Bartel, A. P. (1989). Where Do the New U.S. Immigrants Live? *J Labor Econ, 7*(4), 371-391. https://doi.org/10.1086/298213

Bauer, T., Epstein, G., & Gang, I. (2007). The Influence of Stocks and Flows on Migrants' Location Choices. In *Aspects of Worker Well-Being* (Vol. 26, pp. 199-229). Emerald Publishing Ltd. https://EconPapers.repec.org/RePEc:eme:rleczz:s0147-9121(06)26006-0

Bauer, T., Epstein, G. S., & Gang, I. N. (2005). Enclaves, Language, and the Location Choice of Migrants. *Journal of Population Economics, 18*(4), 649-662. https://doi.org/10.1007/s00148-005-0009-z

Beine, M., Docquier, F., & Özden, Ç. (2011). Diasporas. *Journal of Development Economics, 95*(1), 30-41. https://doi.org/10.1016/j.jdeveco.2009.11.004

Bertoli, S., & Ruyssen, I. (2018). Networks and migrants' intended destination. *Journal of Economic Geography, 18*(4), 705-728. https://doi.org/10.1093/jeg/lby012

Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *The Journal of Human Resources, 50*(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Carrington, W. J., Detragiache, E., & Vishwanath, T. (1996). Migration with endogenous moving costs. *The American Economic Review, 86*(4), 909-930.

Chiswick, B. R., & Miller, P. W. (1996, 1996/02/01). Ethnic networks and language proficiency among immigrants. *Journal of Population Economics, 9*(1), 19-35. https://doi.org/10.1007/PL00013277

Christiadi, C., B. (2007). *Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration*. https://researchrepository.wvu.edu/rri_pubs/65

Cragg, M., & Kahn, M. (1997, 1997/09/01/). New Estimates of Climate Demand: Evidence from Location Choice. *Journal of Urban Economics, 42*(2), 261-284. https://doi.org/https://doi.org/10.1006/juec.1996.2027

Damm, A. P. (2009). Determinants of Recent Immigrants' Location Choices: Quasi-Experimental Evidence. *Journal of Population Economics, 22*(1), 145-174. https://doi.org/10.1007/s00148-007-0148-5

Davies, P. S., Greenwood, M. J., & Li, H. (2001). A Conditional Logit Approach to U.S. State-to-State Migration. *Journal of Regional Science, 41*(2), 337-360. https://doi.org/10.1111/0022-4146.00220

Davis, B., Stecklov, G. U. Y., & Winters, P. (2002). Domestic and international migration from rural Mexico: Disaggregating the effects of network structure and composition. *Population Studies, 56*(3), 291-309. https://doi.org/10.1080/00324720215936

Dinçer, B., Özaslan, M., Satılmış, T. (1996). *İllerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması*. DPT-BGYUGM: Ankara.

Ergür, E. D. (2006). *Sivil Toplum Kruluşları ve Kültürel Etkinlikleri* İstanbul University, Theology, Sociology of Religion.

Etzo, I. (2011). The Determinants of the Recent Interregional Migration Flows in Italy: a Panel Data Analysis. *Journal of Regional Science, 51*(5), 948-966. https://doi.org/10.1111/j.1467-9787.2011.00730.x

Fafchamps, M., & Shilpi, F. (2013). Determinants of the Choice of Migration Destination. *Oxford Bulletin of Economics and Statistics, 75*(3), 388-409. https://doi.org/10.1111/j.1468-0084.2012.00706.x

Gottlieb, P. (1987). *Making Their Own Way: Shorthorn Blacks' Migration to Pittsburgh, 1916–30*. University of Illinois Press.

Grossman, J. (1989). *Land and Hope: Chicago, Black Southerners, and the Great Migration*. University of Chicago Press.

Hellerstein, J. K., McInerney, M., & Neumark, D. (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics, 29*(4), 659-695. https://doi.org/10.1086/660776

İnat, Ş. (2006). *Fellow Citizenship Associatons as NGOs and the Process of Urbanization* Süleyman Demirel University, Institute of Social Sciences, Department of Sociology.

Jaeger, D. A. (2000). *Local Labor Markets, Admission Categories, and Migrant Location Choice*.

Jaeger, D. A. (2007). Green Cards and the Location Choices of Migrants in the United States, 1971-2000. *Research in Labor Economics, 27*, 131-183.

Kocaman, T. (2008). *Türkiye'de İçgöçler ve Göç Edenlerin Nitelikleri (1965-2000)*. DPT.

Lazear, E. P. (1999). Culture and Language. *Journal of Political Economy, 107*(S6), S95-S126. https://doi.org/10.1086/250105

Marks, C. (1989). *Farewell – We're Good and Gone: The Great Black Migration*. Indiana University Press.

McFadden, D. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics, 2*, 1395-1457.

McFadden, D., & Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics (Chichester, England), 15*(5), 447-470.

McKenzie, D., & Rapoport, H. (2010). Self-selection patterns in Mexico-US migration: The role of migration networks. *The Review of Economics and Statistics, 92*(4), 811-821. https://doi.org/10.1162/REST_a_00032

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics, 118*(2), 549-599. https://doi.org/10.1162/003355303321675455

OECD. (2018). *OECD Regions and Cities at a Glance 2018*. https://doi.org/doi:https://doi.org/10.1787/reg_cit_glance-2018-en

Piras, R. (2020). Internal Migration in Italy: The Role of Migration Networks. *Italian Economic Journal, 6*(1), 157-195. https://doi.org/10.1007/s40797-019-00106-y

Schultz, T. P. (1982). Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model. *Economic Development and Cultural Change, 30*(3), 559-593. https://doi.org/10.1086/452576

Schwartz, A. (1973). Interpreting the Effect of Distance on Migration. *The Journal of Political Economy, 81*(5), 1153-1169. https://doi.org/10.1086/260111

Scott, D. M., Coomes, P. A., & Izyumov, A. I. (2005). The Location Choice of Employment-based Immigrants among U.S. Metro Areas. *Journal of Regional Science, 45*(1), 113-145. https://doi.org/10.1111/j.0022-4146.2005.00366.x

STIGM. *Sivil Toplumla İlişkiler Genel Müdürlüğü, İçişleri Bakanlığı, Statistics (Directorate for Relations with Civil Society, Ministry of Interior)*. Retrieved June 27, 2021 from https://www.siviltoplum.gov.tr/sivil-toplumla-iliskiler-genel-mudurlugu

Stuart, B. A., & Taylor, E. J. (2021). Migration Networks and Location Decisions: Evidence from US Mass Migration. *American Economic Journal: Applied Economics, 13*(3), 134-175. https://doi.org/10.1257/app.20180294

Swee, E. L. (2017). Migrant Networks and Job Search: Evidence from Thailand. *Economic Development and Cultural Change, 66*(1), 113-146. https://doi.org/10.1086/693716

Terzi, E. a. K., Y. (2014). Hemşehri Dernekleri, Hemşehrilik Bilinci ve Kentlileşme İlişkisi Üzerine Bir Araştırma: İstanbul/Sultangazi'deki Karslı Hemşehri Dernekleri Örneği. *Selcuk University Journal of Institute of Social Sciences, 32*, 137-150.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2 ed.). Cambridge University Press.

Winters, P., de Janvry, A., & Sadoulet, E. (2001). Family and Community Networks in Mexico-U.S. Migration. *Journal of Human Resources, 36*(1), 159-184. https://doi.org/10.2307/3069674

Zahniser, S. (1999). *Mexican Migration to the United States: The Role of Migration Networks and Human Capital Accumulation*. Garland Publishing Inc.

Zavodny, M. (1999). Determinants of Recent Immigrants' Locational Choices. *The International Migration Review, 33*(4), 1014-1030. https://doi.org/10.2307/2547361

Table 1 Descriptive Statistics

Cell contents are relative to the province of origin	1990 cross-section of migrants: Mean in chosen destinations (1)	Mean in alternative destinations (2)	t-test of difference in means (3)	2000 cross-section of migrants: Mean in chosen destinations (4)	Mean in alternative destinations (5)	t-test of difference in means (6)
Migrant networks						
Birth place share of province population (%)	-23.412	-44.601	46.745	-44.332	-55.512	23.384
	(71.913)	(45.373)		(61.461)	(43.301)	
Percent of birth place population in province	-17.573	-40.084	57.415	-36.620	-50.010	32.014
	(62.208)	(38.401)		(53.782)	(36.788)	
Foreign-born share of province population (%)	9.613	-7.753	89.349	12.703	-6.533	83.685
	(30.820)	(21.149)		(29.549)	(21.122)	
Labor market attributes						
Unemployment rate (%)	0.077	-0.295	32.122	-0.524	-0.308	-13.612
	(1.830)	(2.167)		(2.013)	(2.668)	
Nonagricultural employment rate (%)	8.201	-8.238	88.182	9.212	-5.546	78.633
	(29.528)	(23.124)		(24.105)	(19.115)	
Nonagricultural unemployment rate (%)	-0.724	0.326	-50.999	-1.550	-0.079	-51.382
	(3.234)	(4.465)		(3.646)	(4.751)	
Population and distance						
Population size (log)	0.375	-0.563	98.040	0.436	-0.582	85.911
	(1.517)	(1.142)		(1.524)	(1.142)	
Population density (# of people/km^2)	115.001	-105.659	69.598	177.783	-111.203	60.336
	(503.161)	(301.060)		(616.577)	(351.680)	
Distance from origin province ('00 km)	4.963	5.714	-34.245	4.881	5.700	-30.877
	(3.468)	(3.161)		(3.400)	(3.122)	

Notes: t-test of difference in means column reports the t statistic for the test of equality of means. t-test assumes unknown and unequal population variances. All differences in means are statistically significant at 1% level. The number of individuals is 25,325 and 16,655 in the 1990 and 2000 cross-sections respectively. Standard deviations are in parenthesis.

Table 2 Determinants of location choice - using 28-54 years old male work migrants

Difference in destination and origin province characteristics	Pooled sample (1990 & 2000) (1)	1990 cross-section (2)	2000 cross-section (3)
Migrant networks			
Birth place share of province population (%)	0.285***	0.286***	0.286***
	(0.020)	(0.023)	(0.021)
Birth place share of province population sq. ÷ 100 (%)	-0.162***	-0.161***	-0.169***
	(0.017)	(0.019)	(0.016)
Percent of birth place population in province	0.021***	0.027***	0.011
	(0.007)	(0.007)	(0.007)
Percent of birth place population in province sq. ÷ 100	-.0152***	-0.023***	-0.006
	(0.004)	(0.004)	(0.004)
Foreign-born share of province population (%)	0.040***	0.052***	0.030***
	(0.005)	(0.005)	(0.005)
Foreign-born share of province population sq. ÷ 100 (%)	0.007***	0.007**	0.011***
	(0.003)	(0.003)	(0.004)
Labor market attribute			
Unemployment rate (%)	-0.144***	-0.191***	-0.098**
	(0.023)	(0.026)	(0.048)
Demographics and distance			
Population size (log)	0.965***	0.980***	0.939***
	(0.060)	(0.067)	(0.059)
Distance from origin province (‘00 km)	-0.377***	-0.380***	-0.381***
	(0.030)	(0.031)	(0.034)
Distance from origin province sq. ÷ 100 (‘00 km)	2.158***	2.262***	2.024***
	(0.225)	(0.241)	(0.241)
Pseudo-R^2	0.398	0.431	0.355
Number of individuals	41,980	25,325	16,655
Number of observations	2,770,680	1,671,450	1,099,230

Notes: The table presents coefficient estimates from fixed effects conditional logit. We control for within-origin-province error correlation by implementing cluster robust variance estimator and the resulting standard errors are in parenthesis. All models also include province-group dummies, province of birth fixed effects, and dummy/dummies that accounts for the impact of province border change on migrants' location propensities. * significant at 10%; ** significant at 5%; ***significant at 1%.

Table 3 Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant networks								
Birth place share of province population (%)	0.306***	0.285***	0.281***	0.266***	0.281***	0.310***	0.388***	0.029***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.034)	(0.007)
Birth place share of province population sq. ÷ 100 (%)	-0.179***	-0.162***	-0.159***	-0.146***	-0.159***	-0.182***	-0.236***	0.051***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.016)	(0.026)	(0.004)
Percent of birth place population in province	0.020***	0.020***	0.021***	0.030***	0.022***	0.016***	0.076***	0.061***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.009)	(0.010)
Percent of birth place population in province sq. ÷ 100	-0.010***	-0.015***	-0.016***	-0.016***	-0.015***	-0.014***	-0.082***	-0.009***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.003)
Foreign-born share of province population (%)	0.035***	0.042***	0.037***	0.073***	0.058***		0.081***	0.051***
	(0.002)	(0.005)	(0.005)	(0.004)	(0.005)		(0.008)	(0.009)
Foreign-born share of province population sq. ÷ 100 (%)	-0.003	0.008***	0.008***	0.007***	0.008***		-0.079***	0.047***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)		(0.010)	(0.008)
Labor market attributes								
Unemployment rate (%)	-0.122***			-0.072***	-0.115***	-0.091***	-0.173***	-0.121***
	(0.017)			(0.018)	(0.025)	(0.029)	(0.024)	(0.028)
Nonagricultural employment rate (%)		-0.009***						
		(0.002)						
Nonagricultural unemployment rate (%)			-0.049***					
			(0.012)					
Demographics and distance								
Population density (# of people/km^2)				-0.001**				
				(0.000)				
Land area (log, km^2)					0.659***			
					(0.090)			
Population size (log)	0.791***	0.920***	0.930***		0.625***	1.222***	0.989***	0.837***
	(0.059)	(0.056)	(0.057)		(0.076)	(0.050)	(0.050)	(0.057)
Dist. from origin prov. ('00 km)	-0.336***	-0.377***	-0.378***	-0.368***	-0.380***	-0.366***	-0.402***	-0.376***
	(0.036)	(0.030)	(0.030)	(0.031)	(0.030)	(0.030)	(0.031)	(0.064)
Dist. from origin prov. ('00 km) sq. ÷ 100	1.984***	2.154***	2.166***	2.117***	2.174***	2.151***	2.380***	1.878***
	(0.247)	(0.227)	(0.225)	(0.226)	(0.227)	(0.231)	(0.212)	(0.459)
Non-migration dummy								33.695***
								(0.545)
Pseudo-R^2	0.385	0.398	0.398	0.392	0.399	0.395	0.397	0.967
Number of observations	2,770,680	2,770,680	2,770,680	2,770,680	2,770,680	2,770,680	2,770,680	53,209,457

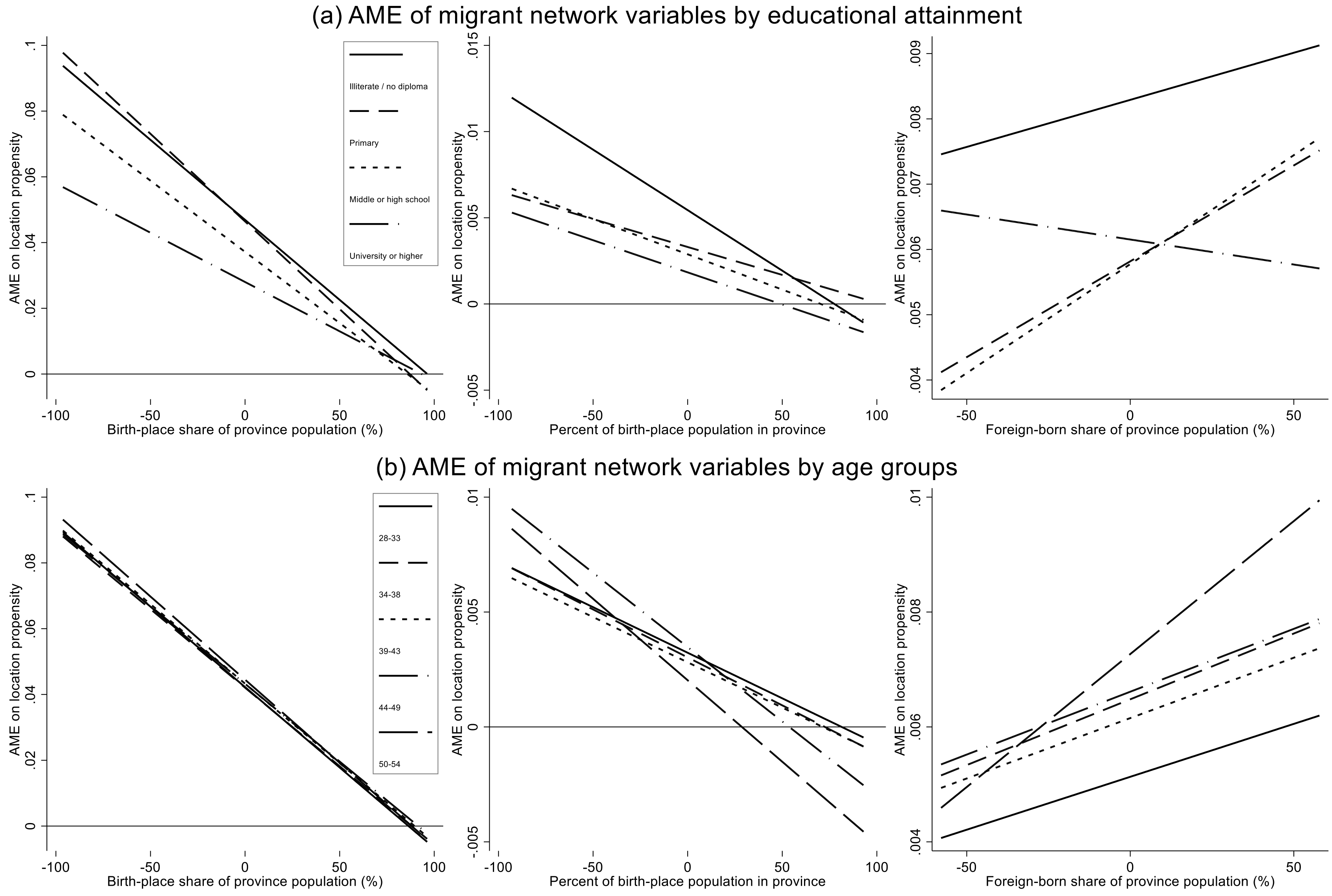
Notes: The table presents coefficient estimates from fixed effects conditional logit. We control for within-origin-province error correlation by implementing cluster robust variance estimator and the resulting standard errors are in parenthesis. All models also include province-group dummies, province of birth fixed effects, and dummy/dummies that accounts for the impact of province border change on migrants' location propensities. (1) omits province-group dummies. (2) and (3) test with alternative labor market measures. (4) uses population density as demographic control. (5) includes to the main specification land area of province. (6) omits foreign-born share of province networks. (7) incorporates as regressors the destination characteristics. (8) includes non-migrants to the sample. In all models except for (7) destination-source province differences in characteristics are used as regressors. Standard errors are clustered at source province level. * significant at 10%; ** significant at 5%; ***significant at 1%.

Table 4: Heterogenous effects by last finished schooling, age groups and migration motives

	Number of observations	Birth place share of province population (%)	Percent of birth place population in province	Foreign-born share of province population (%)	Unemployment rate (%)	Dist. from origin prov. ('00 km)
Education Groups						
Illiterate / no completed degree	240,042	0.314*** (0.030)	0.036*** (0.009)	0.056*** (0.008)	-0.120*** (0.045)	-0.348*** (0.044)
Primary	1,577,928	0.311*** (0.022)	0.022*** (0.007)	0.039*** (0.007)	-0.164*** (0.030)	-0.397*** (0.033)
Middle or high school	666,798	0.249*** (0.025)	0.019** (0.008)	0.039*** (0.004)	-0.128*** (0.031)	-0.416*** (0.035)
Tertiary education	284,790	0.188*** (0.020)	0.012 (0.008)	0.041*** (0.006)	-0.121*** (0.038)	-0.294*** (0.027)
Age Groups						
28-33	1,183,116	0.283*** (0.022)	0.022*** (0.007)	0.034*** (0.005)	-0.118*** (0.028)	-0.395*** (0.034)
34-38	672,144	0.282*** (0.020)	0.020*** (0.007)	0.043*** (0.006)	-0.110*** (0.029)	-0.393*** (0.031)
39-43	420,024	0.290*** (0.025)	0.019** (0.008)	0.041*** (0.004)	-0.190*** (0.032)	-0.334*** (0.034)
44-49	324,324	0.290*** (0.022)	0.023*** (0.008)	0.044*** (0.006)	-0.205*** (0.037)	-0.369*** (0.037)
50-54	171,072	0.298*** (0.036)	0.014 (0.009)	0.049*** (0.008)	-0.211*** (0.061)	-0.321*** (0.038)
Migration Motives						
Job search/start a job	1,099,230	0.286*** (0.021)	0.011 (0.007)	0.030*** (0.005)	-0.098** (0.048)	-0.381*** (0.034)
Transfer	217,338	0.264*** (0.026)	-0.000 (0.007)	0.012*** (0.004)	-0.005 (0.036)	-0.377*** (0.030)
Tied mover	88,044	0.336*** (0.050)	-0.004 (0.017)	0.027*** (0.008)	-0.020 (0.081)	-0.391*** (0.054)
Education	27,390	0.349*** (0.045)	-0.000 (0.014)	0.039** (0.017)	0.093 (0.159)	-0.466*** (0.088)
Earthquake	86,856	0.225*** (0.019)	0.035*** (0.012)	0.022*** (0.007)	-0.055 (0.053)	-0.263*** (0.090)
Safety issues	13,200	0.467*** (0.085)	-0.029 (0.025)	0.060* (0.033)	0.024 (0.204)	-0.318*** (0.105)

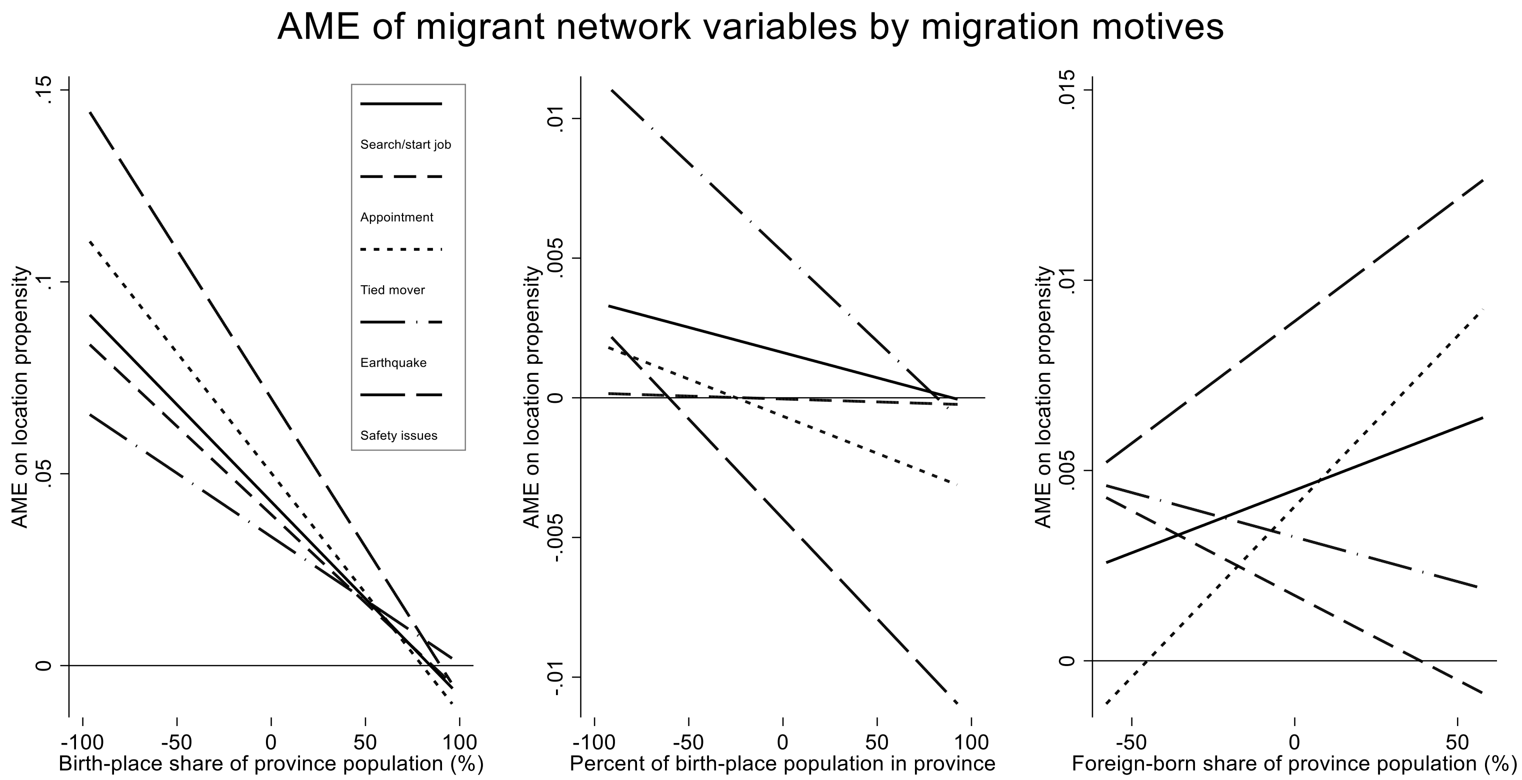
Notes: The table presents heterogeneous effects by highest achieved schooling, age group and migration motives of 28-54 years-old male work migrants. For results concerning highest level of schooling and age groups pooled sample of migrants from 1990 and 2000 censuses is used. Heterogeneity by migration motives is examined using data from 2000 census. Each row is a fixed effects conditional logit regression of migrants' location choices on a sample stratified by groups in the first column. We suppress the coefficient estimates of population size and of quadratic terms for migrant network and distance variables to save space. All models also include province-group dummies, province of birth fixed effects, and dummy/dummies that accounts for the impact of province border change on migrants' location propensities. In all models destination-source province differences in characteristics are used as regressors. Standard errors are clustered at source province level and are reported in parenthesis. * significant at 10%; ** significant at 5%; ***significant at 1%.

Figure 1: AME of migrant network variables for samples stratified by highest educational attainment and age



Notes: Average marginal effect of a 10% increase in each migrant network variable on migrants' location propensities for samples stratified by education and age is presented in the figure.

Figure 2: Average marginal effect of migrant network variables for samples stratified by migration motives



Notes: Average marginal effect of a 10% increase in each migrant network variable on migrants' location propensities for samples stratified by different migration motives is presented in the figure.