

Discussion Papers in Economics

## Job Recruitment Networks and Migration to Cities in India

Vegard Iversen

Kunal Sen

Arjan Verschoor

Amaresh Dubey

January 2009

Discussion Paper 09-01



Indian Statistical Institute, Delhi  
Planning Unit

7, S. J. S. Sansanwal Marg, New Delhi 110016, India

## ***Job Recruitment Networks and Migration to Cities in India***

*Vegard Iversen<sup>+</sup>, Kunal Sen<sup>\*</sup>, Arjan Verschoor<sup>+</sup>, and Amaresh Dubey<sup>#</sup>*

### ***Abstract***

Economists have focused on job search and supply-side explanations for network effects in labour transactions. This paper develops and tests an alternative explanation for the high prevalence of network-based labour market entry in developing countries. In our theoretical framework, employers use employee networks as screening and incentive mechanisms to improve the quality of recruitment. Our framework suggests a negative relationship between network use and the skill intensity of jobs, a positive association between economic activity and network use and a negative relationship between network use and pro-labour legislation. Furthermore, social identity effects are expected to intensify compared to information-sharing and other network mechanisms. Using data from an all-India Employment Survey we implement a novel empirical strategy to test these relationships and find support for our demand-side explanation.

---

<sup>+</sup> School of Development Studies, University of East Anglia, Norwich. e-mail: v.iversen@uea.ac.uk

<sup>\*</sup> IDPM, University of Manchester.

<sup>#</sup> CSRD, Jawaharlal Nehru University and National Council of Applied Economic Research, New Delhi.

## 1. Introduction

An extensive literature has addressed the impacts of social interactions and networks on labour market outcomes (McEntarfer 2003; Ionnides and Loury 2004; Wahba and Zenou 2005). One strand of this literature has focused on how networks and peer emulation may affect migration within or across countries (Stark 1991; Hatton and Williamson 1998; Munshi 2003). Empirical research has, however, found it hard to distinguish between network effects, which are transmitted to a migrant by virtue of the social connections he has access to, and peer emulation effects which are transmitted through the local neighbourhood he belongs to, whether at origin or destination. Crucially, the empirical literature has failed to distinguish the case where employers recruit through employee networks from that where networks simply transmit information about vacancies during job search. The tendency, therefore, has been to conflate demand-side and supply-side explanations for network effects treating them as one and the same thing.

To see why these effects are fundamentally different, consider a male Anand, residing in village-India, and contemplating whether to migrate to Mumbai. Social interaction and network effects may influence his migration decision as follows: Firstly, Anand may migrate to emulate village friends already settled in Mumbai. Secondly, Anand may migrate because his Mumbai-based co-villager Vijay has learnt about and informs Anand (and possibly others) about a vacancy. Thirdly, Anand may ask his relatives, Vijay and other Mumbai-based village friends to find him a job.<sup>1</sup> Finally, Vijay's employer may ask Vijay to find a suitable person to fill a vacancy,

---

<sup>1</sup> Following Granovetter (1973), Dasgupta (2000;382) notes that "the reason weak ties are especially useful in job search is that they cover a greater range of links than strong ties do. Weak ties connect one to a wide variety of people and therefore a wider information base." Whether strong or weak ties matter – in search or recruitment – is an empirical question. Banerjee's (1983) data on migrants in Delhi show weak ties to be vital during search, while we hypothesise that employee referral narrows the recruitment pool. In principle, poor migrants may limit search to strong tie networks, but this requires the strong assumption that they can afford a strategy that will generate fewer job opportunities.

and Vijay *asks* Anand to fill this vacancy in Vijay's workplace. The first mechanism, the peer emulation effect, has received empirical support (e.g. Marmaros and Sacerdote 2002). The second and third mechanisms capture the information-dispersing role of networks, where networks are non-market institutions that may repair market failures. The fourth mechanism involves the strategic use of networks in recruitment. Acting as someone's workplace guarantor involves considerable reputational hazards and is likely to make an employee middleman cautious and selective. This will narrow the pool of potential recruits and strengthen the importance of social identity compared to information sharing and other network mechanisms.

Whether labour market failures should be a cause for concern depends on network outreach and the services networks provide. If networks mimic local public goods by dispersing information about job vacancies or are widely accessible to job seekers, such concerns would quickly dissolve.<sup>2</sup> Furthermore, the role of policy would be straightforward; to strengthen the dissemination of labour market information.

Policy conclusions are less cut and labour market failures potentially more severe if networks are screening and incentive mechanisms employers invoke to improve the quality of recruitment. If employer benefits from employee-referral are greater for low skill jobs, doubts about migration as an equitable route for escaping poverty would surface.

So far, the literature has favoured supply-side explanations for the high prevalence of network-based labour market entry among poor workers in developing countries (e.g. Wahba and Zenou 2005). The scale and adverse impacts of failures in lower end labour markets may therefore to a large extent have been overlooked.

---

<sup>2</sup> Banerjee (1983) found network access to transcend caste-boundaries with family and village ties serving similar functions in aiding migrants, and networks resembling local public goods. On Mexican data, Winters et al. (2001) found family or strong ties to provide no advantages over weak community ties, suggesting cumulative public

After sketching a theory of recruitment, we present the first attempt to filter out peer emulation effects and disentangle demand-side from supply-side explanations for how social networks affect migration behaviour. Our theoretical framework suggests more powerful network effects for less educated workers, more network-based recruitment during economic booms and less when legislation is worker-friendly. Finally, social identities gain in importance when employers recruit through employee networks.

In a path-breaking study of Mexican migrants to the US, Munshi (2003) is able to identify network effects. Using variation within each origin-community's network over time, he controls for selectivity in the migration decision and for network endogeneity. However, Munshi is unable to distinguish the information dispersion from the strategic recruitment explanation for network use. His empirical analysis also does not distinguish workplace clustering caused by source community cohesion (peer emulation effects) from clustering reflecting employer preferences.

Wahba and Zenou (2005) test for social network effects in job search. Their supply-side theory predicts a higher prevalence of network-based search among illiterate and semi-literate workers who are unable to read job-ads and vacancy posters and therefore have stronger incentives to search through friends and contacts. With vacancy information spreading through networks of weak ties, Wahba and Zenou in effect test a version of the information dispersion hypothesis of network effects. They argue that Egyptian labour market data support their main hypothesis but their empirical specification is too coarse to persuasively support this conclusion.<sup>3</sup>

---

knowledge about migration to be open to all irrespective of social or economic backgrounds. Munshi (2003;551) notes that belonging to the same origin community (*paisanaje*) entitles Mexico-US migrants to network access.

<sup>3</sup> Wahba and Zenou's (2005) network variable – which is a proxy for the number of random contacts per unit of time – is the population density of a region. This is a highly imperfect proxy and our analysis involves much more refined variables for identifying network effects, specifically the presence of similar migrants from a particular source area in a precisely defined city and industry.

After filtering out peer emulation effects, we adopt the following strategy to identify referral based network effects. We use dyads<sup>4</sup> or pairs of migrants as unit of analysis. Our binary dependent variable takes the value 1 when a pair of migrants working in a particular industry and in a particular city is from the same source area. In a perfect data set, we would be able to replace industry with firm, but industry at the high level of disaggregation we are able to offer represents an attractive proxy. Informed by our theoretical sketch, we test whether the likelihood that a migrant pair working in industry  $i$  and city  $j$  is from the same source area is higher for less educated migrants. Next and facilitating distinction between information dispersion and referral network effects, we test the impacts of pro-worker legislation and economic activity on our dependent variable. We proceed to test for social identity effects. If networks are information-dispersing local public goods (e.g. Winters et al 2001) or if information spreads through weak ties (Wahba and Zenou 2005), an aggregate migrant stock (e.g. Carrington et al 1996) should suffice, making disaggregation by social identity unnecessary. We perform additional robustness tests to distinguish referral based network effects from rival accounts.

The paper is laid out as follows: Section 2 sketches the theory of recruitment and section 3 outlines our empirical strategy for identifying our recruitment explanation for network use. Section 4 describes the data and provides descriptive statistics. Section 5 presents the empirical results, while Section 6 concludes.

---

<sup>4</sup> The dyadic approach has gained popularity in network analysis by combining attractive statistical properties with intuitive appeal (Wasserman and Faust 1995; de Weerd 2004).

## 2. Networks as screening and incentive mechanisms

Based on a survey of 4,900 households in Mumbai, Munshi and Rosenzweig (2006) report that around 70 % of unskilled and blue collar workers relied on a referral for finding their first job. The corresponding figure for white-collar workers was around 44%. Like us, Munshi and Rosenzweig attribute the high prevalence of referrals in low and unskilled occupations to the more acute information and enforcement problems in such jobs. These problems are likely to intensify in connection with migration. A newly arrived migrant asking for a job confronts a prospective employer with the challenge of ascertaining the candidate's ability and his trustworthiness. The candidate's papers, if they exist, may not be reliable and his history may be costly to trace. Referral based recruitment may remedy the employer's problem by screening and reducing uncertainty about the candidate's labour productivity (Marsden and Gorman 2001) and by incentivising workplace loyalty (Fernandez and Castilla 2001).<sup>5</sup> The latter resonates with Holmstrom's (1984; 219) anthropological accounts of motivations for referral-based recruitment among Mumbai employers; *Above all he [a factory owner] wants a stable work force, people he has trained himself and who will resist the "temptation" to leave.* Moreover, networks of existing employees are described as *"useful as sources of unskilled labour, or people with enough basic education to learn semi-skilled tasks fast, but not for trained skilled workers (ibid.)."* Turnover rates in India, especially among low and unskilled workers, are high (Mazumdar 1973; Newman 1979). While absenteeism shows little inter-industry variation and fluctuates around 10%, the average turnover in private sector industries is 18.7 % (Annual Survey of Industries 2002-03). There are stark inter-state contrasts,

---

<sup>5</sup> Montgomery (1991) assumes assortative matching in contacts with employers recruiting through their most talented employees.

with West-Bengal reporting the lowest (5.2%) and Punjab (49.5%) the highest turnover rates.

## 2.1 Recruitment – a theoretical sketch

Consider a small urban employer already employing a migrant worker and about to expand. Let  $\bar{\theta}$  denote a high productivity and  $\underline{\theta}$  a low productivity worker in a binary distribution of worker types. The employer must decide whether to recruit through the local labour market or use the employee's village network. The first option offers the prospect:

$$p\bar{\theta} + (1-p)\underline{\theta}, \quad (1)$$

where (1) is the expected value added in output from a new recruit, and  $p$  is the probability that the recruit is highly productive. Following Marsden and Gorman (2001), the employee has an informational advantage if recruiting through his village network provides:

$$q\bar{\theta} + (1-q)\underline{\theta}, \quad q > p. \quad (2)$$

To start with, let the distribution of labour or *task* productivities in the urban labour market and the worker's village network be identical.<sup>6</sup> The *informational* gain from network-based recruitment is proportional to the heterogeneity in task productivity  $V = \bar{\theta} - \underline{\theta}$ , since the gain equals  $(q - p)V$ . Consider, next, loyalty, where better discipline and lower turnovers are prospective employer gains from referral-based recruitment as elaborated below:<sup>7</sup>

Let

$$t = t(s, SD, G, L) \quad (3)$$

---

<sup>6</sup> This, as suggested below, is plausible for un- and low-skilled jobs.



be the voluntary quit rate. Attrition depends on the human capital or skill-intensity of jobs,  $s$  while  $SD$  is a measure of the social distance between the employee middleman and a new recruit.  $G$  is an indicator of the state of the economy and labour market tightness (see Salop 1979), while  $L$  measures the degree of pro-labour legislation (see Besley and Burgess 2004).

The employer's workforce stability gain from referral-based recruitment,  $S$ , may be written:

$$S = [t_m(s, SD, G, L) - t_n(s, SD, G, L)]C(s), \quad (4)$$

where  $t_m$  and  $t_n$  are attrition rates when recruiting through the market and through employee networks, respectively.  $C$  is the unit cost of replacing a worker. Replacement costs are likely to increase with skill-intensity, so that  $dC/ds > 0$ . Consistent with evidence cited above, we assume that  $t_m - t_n$  attains its maximum for un- and low-skilled jobs.

The impact on  $S$  of an increase in skill-intensity becomes:

$$\frac{\partial S}{\partial s} = \left[ \frac{\partial t_m}{\partial s} - \frac{\partial t_n}{\partial s} \right] C + (t_m - t_n) \frac{\partial C}{\partial s}. \quad (5)$$

With the gap between attrition rates being higher in un- and low-skilled jobs, we have:

$$\left| \frac{\partial t_n}{\partial s} \right| < \left| \frac{\partial t_m}{\partial s} \right|, \quad (6)$$

so that the first term in (5) is negative, while the second is positive. For increasing skill-intensity of jobs, whether referral-based recruitment becomes more attractive depends, in part, on the relative strength of these two effects.<sup>8</sup>

---

<sup>7</sup> Hence referral-based recruitment cements employer-employee relations, thereby tackling a particular form of moral hazard.

<sup>8</sup> This refers to the workforce stability-gain from employee referral. Section 2.2 shows how incentives for recruiting through employee village networks are affected by human capital scarcities in village communities.

Consider, next, how social distance affects  $S$ . Compared to sharing information about a vacancy, bringing someone to the workplace exposes an employee middleman to considerable reputational risks and is likely to make him cautious and selective. This narrows the pool of potential recruits.  $SD$  measures the social proximity between employee and new recruit and therefore the intensity of the sanctions that may be imposed on a recruit renegeing on this implicit contract. Hence:

$$\frac{\partial S}{\partial SD} = \left[ \frac{\partial t_m}{\partial SD} - \frac{\partial t_n}{\partial SD} \right] C < 0, \quad (7)$$

$$\text{where } \frac{\partial t_m}{\partial SD} \cong 0, \frac{\partial t_n}{\partial SD} > 0.$$

Consider, next, the impact on  $S$  from a surge in economic activity:

$$\frac{\partial S}{\partial G} = \left[ \frac{\partial t_m}{\partial G} - \frac{\partial t_n}{\partial G} \right] C > 0. \quad (8)$$

For a given skill-level, workers recruited through employee referral are expected to be more loyal and less responsive to a heating up of the economy.<sup>9</sup> Hence,  $t_m$  is more sensitive to a change in  $G$  than  $t_n$ . In a tighter labour market, referral-based recruitment becomes more attractive for employers.

Labour protective legislation may achieve much the same as network-based recruitment in cementing employer-employee relationships. Hence,

$$\frac{\partial S}{\partial L} = \left[ \frac{\partial t_m}{\partial L} - \frac{\partial t_n}{\partial L} \right] C < 0, \quad (9)$$

$$\text{where } \frac{\partial t_m}{\partial L} < 0, \frac{\partial t_n}{\partial L} < 0 \text{ and } \left| \frac{\partial t_n}{\partial L} \right| < \left| \frac{\partial t_m}{\partial L} \right|.$$

As for an economic contraction, the gains from referral-based recruitment are expected to decline since better working conditions make workers more keen to hold

---

<sup>9</sup> Using US-data, Faberman (2005) reports a negative correlation between voluntary quits and unemployment.

on to their jobs. Hence, risks from recruiting through the market decline, lowering gains from referral-based recruitment.

Before elaborating on when employee referral will be attractive, we briefly describe the two labour pools.

## 2.2 Urban labour markets versus village networks

Consider an industry where unskilled ( $u-s$ ) jobs dominate. Let the binary distribution of individual productivities for such jobs in the village network be  $X_{us}^R$ . The village network will contain many individuals capable of doing such jobs well. The urban employer's alternative is to recruit through the local labour market. Given the nature of  $u-s$  jobs, let  $X_{us}^R = X_{us}^U$ . This does not rule out systematic differences in the attributes of workers in the two pools. For instance, since village communities rarely are repositories of human capital, the average urban candidate is likely to be more educated. Yet, the simple nature of  $u-s$  jobs minimises labour productivity differences, driving productivity returns to education or skills in such jobs towards zero. In contrast and as table 1 suggests, for more human capital-intensive, semi-skilled ( $s-s$ ) jobs, education improves labour productivity:

[Insert table 1 about here]

Column (a) covers unskilled ( $u-s$ ) jobs with zero productivity returns to education. Column (b) summarises the productivity for each worker category in more human capital intensive ( $s-s$ ) jobs, expressing the productivity of workers (with education level  $e_I$ ) in the urban pool as functions of the productivity of rural, unskilled workers in  $s-s$  jobs. With zero labour productivity returns to education in  $s-s$  jobs,  $m = n = 1$ .

With positive returns,  $m(e_1), n(e_1) > 1$ . Let  $m=n$ . In choosing between referral- or market-based recruitment, an employer will balance the employee's informational advantage and the workforce stability gains  $S$  against the potential labour productivity gains from the more educated workers in the urban pool. Hence, the employer will rank the employee's village network over market-based recruitment whenever:

$$(q - np)(\bar{\beta} - \underline{\beta}) + S > (n - 1)\underline{\beta}. \quad (10)$$

The right hand side measures the net productivity returns to  $e_1$  for low productivity workers and is positive when  $n > 1$ . We normalise using  $\underline{\beta} = 1$  as “numeraire”. Moreover, let  $\bar{\beta} = a\underline{\beta} = a$  where  $a > 1$  measures worker heterogeneity.

The employer will recruit through the urban labour market whenever:

$$(np - q)(a - 1) - S > (1 - n). \quad (11)$$

Reorganising and for given exogenous values of  $q$ ,  $p$ ,  $a$  and  $S$ , a switch point is defined by:

$$n^*(e_1) = \frac{q(a - 1) + S + 1}{p(a - 1) + 1}. \quad (12)$$

Consider the employee informational advantage and labour productivity returns to education, leaving  $S$  aside.<sup>10</sup> Figure 1 displays a locus of switch points along which an employer will be indifferent between the two recruitment strategies.

**[Figure 1 about here]**

If other screening mechanisms are weak and the employee's informational advantage distinct, positive productivity returns to the human capital of urban workers are necessary to induce a switch from network to market-based recruitment. Southeast of point A, it is better to recruit through the urban labour market since the labour productivity of urban workers outweighs employee informational advantages. In

region C, labour productivity returns are insufficient to compensate for employee informational advantages making recruitment through the employee's village network optimal. In D, recruitment through the urban labour market is again preferred. In summary, the theoretical sketch captures information asymmetries about worker types, where loyalty and turnovers are endogenous and respond (i) to the social distance between employee middleman and the recruit, (ii) the state of the economy and (iii) pro-labour legislation. Furthermore, the relationship between skill-intensity and incentives for recruiting through employee village networks depends on S, but also crucially on the disparity between the labour productivity of urban and rural recruits as determined by the labour productivity returns to education in different types of jobs.

### **3. Empirical strategy**

Our main hypothesis is that referral-based recruitment makes network effects stronger for low- and unskilled jobs. Furthermore, network effects intensify during economic booms and weaken when legislation is pro-labour. In addition, social identity effects are expected to be stronger, since employee referral narrows the recruitment pool. This section outlines how we identify network effects (by filtering out peer emulation effects) and our strategy for distinguishing referral-based from information dispersion network effects.

To identify referral-based network effects we study the probability that a pair of randomly selected migrants who work in the same city and industry, with industry defined as narrowly as the data permits, are from the same source area. This benchmark empirical specification is given by:

---

<sup>10</sup> Including S is straightforward.

$$D_{ij} = \alpha_1 X_{ij} + \alpha_2 F_{ij} + \alpha_3 S_{ij} + \gamma_i + \delta_j + \varepsilon_{ij} \quad (13)$$

where  $D$  is a dummy variable that takes the value one if a pair of migrants working in the same city and industry are from the same source area and zero otherwise;  $i$  and  $j$  denote industry and city respectively,  $X$  is a vector of control variables and  $F$  a vector of relevant migrant pair attributes such as education and social identities;  $\gamma_i$  are industry dummies,  $\delta_j$  city dummies and  $S$  a vector of stock network variables. We include individuals who have migrated from the same source region in the past five years in the migrant pairs in the dependent variable.

Our choice of dependent variable enables us to *uniquely map* the probability that a migrant pair who work in the same industry in the same city and are from the same source area to the shared social characteristics of the matched migrant pair and therefore to test whether network effects, as our theory suggests, are stronger when migrants share characteristics such as low or zero literacy and social identity. We add controls for sex, age and other attributes expected to affect labour market outcomes. A distinction is made between recent and past network effects; ‘stock’ effects are exhibited when the dependent variable responds to past migration. Specifically, the stock variable comprises the number of migrant pairs working in the same industry and city that are both from the same source area, where  $t > 5$  years.<sup>11</sup>

The identification of network effects requires the filtering out of peer emulation and of so-called ‘correlated effects’ (Manski 2000). The latter refers to the concentration of worker characteristics in particular source areas that are relevant to employment in specific cities and/or industries which could make it harder to identify network effects. For instance, if skilled weavers are disproportionately demanded in Mumbai and supplied from a particular part of Gujarat, there is a risk of attributing

the clustering of migrated skilled weavers in Mumbai to network effects when the real cause is a concentration of supplied and demanded worker characteristics at source and destination. This problem is not likely to be particularly severe here, since we are forced to work with rather broad definitions of ‘same source area’. We do, however, check this possibility by testing the robustness of our results to the introduction of source area dummies which control for variation in such concentrations of worker characteristics across source areas.

Peer emulation effects imply that new migrants emulate past migration strategies and may appear identical to network effects.<sup>12</sup> Emulation could result in clustering of migrants from the same source area in the same city and, possibly, industry which may be wrongly attributed to network effects. We adopt the following strategy to filter out peer emulation effects and identify network effects. We first construct our dependent variable such that the more obvious peer emulation effects are not considered. By restricting the analysis to employed migrants, those who have failed to find work are left out. Peer emulation effects may feature in this group, but referral-based recruitment does clearly not. Furthermore, in constructing our dependent variable, we adopt a definition of ‘same industry’ that is as narrow as possible, thus utilising the highly refined disaggregation of industry that our data permits (see Section 4). The more narrow the definition of ‘same industry’ in which a migrant pair from the same source area work, the more likely it is that referral-based network effects as opposed to peer emulation effects operate. While both emulation and referrals may get migrants to the same city as their peers, the likelihood of ending up in the same (narrowly defined) industry is greater for referral-based recruitment.

---

<sup>11</sup> We experimented with different values of  $t$  in the construction of the network and stock variables with no change in our results.

<sup>12</sup> Thus exemplifying so-called ‘endogenous interactions’ (Manski 2000), whereby the propensity to behave in a certain way varies with the average behaviour of one’s social reference group.

Hence, the more precisely we are able to define destination industry, the more successful the filtering out of peer emulation effects will have been. Finally, we introduce industry and city dummies, denoted  $\gamma_i$  and  $\delta_j$  respectively. The city effects control for the possibility that workplace clustering of migrants from a particular region reflects migrant preferences for being with friends and family in the same city. The industry effects net out peer emulation effects whereby migrants emulate past migrants by seeking employment in industries where these past migrants found jobs.

We next describe how we distinguish referral from information-dispersion network effects. In our theoretical sketch, referral based effects are more likely for individuals with no or low levels of education. We construct four variables capturing migrant pair attributes –  $F_{ij}^{ILL}$ ,  $F_{ij}^{LOWLIT}$ ,  $F_{ij}^{MEDLIT}$ , and  $F_{ij}^{HIGHLIT}$  which are dummy variables that take the value one when both migrants in industry  $i$  and city  $j$  are educated to the same level and where superscripts *ILL*, *LOWLIT*, *MEDLIT* and *HIGHLIT* denote migrant pairs who are illiterate or educated to the primary, secondary or graduate level, respectively. The empirical specification becomes:

$$D_{ij} = \alpha_1 X_{ij} + \alpha_2 F_{ij}^{ILL} + \alpha_3 F_{ij}^{LOWLIT} + \alpha_4 F_{ij}^{MEDLIT} + \alpha_5 F_{ij}^{HIGHLIT} + \alpha_6 S_{ij} + \gamma_i + \delta_j + \varepsilon_{ij}. \quad (14)$$

Our theory suggests that  $\alpha_2$  and  $\alpha_3$  are positive and greater than  $\alpha_4$  and  $\alpha_5$ .

The next step is to disaggregate the stock network measure by the level of education of migrant pairs. Hence, we disaggregate  $S_{ij}$  into  $S_{ij}^{ILL}$ ,  $S_{ij}^{LOWLIT}$ ,  $S_{ij}^{MEDLIT}$ , and  $S_{ij}^{HIGHLIT}$  which represent stocks of migrant pairs who migrated from the same source region in time period  $t > 5$  years, and where superscripts have the same interpretations as before.

The empirical specification becomes:



$$\begin{aligned}
D_{ij} = & \alpha_1 X_{ij} + \alpha_2 F_{ij}^{ILL} + \alpha_3 F_{ij}^{LOWLIT} + \alpha_4 F_{ij}^{MEDLIT} + \alpha_5 F_{ij}^{HIGHLIT} \\
& + \alpha_6 S_{ij}^{ILL} + \alpha_7 S_{ij}^{LOWLIT} + \alpha_8 S_{ij}^{MEDLIT} + \alpha_9 S_{ij}^{HIGHLIT} \\
& + \gamma_i + \delta_j + \varepsilon_{ij}.
\end{aligned} \tag{15}$$

If networks were information dispersion mechanisms, the stock of past migrant pairs in a particular industry and city should positively affect the availability of information about job opportunities in this city and industry. Accordingly, the *aggregate* stock of past migrants in city  $i$  and industry  $j$  from a particular source area should affect recent migration from this area. If so,  $\alpha_6$  in equation (14) would be positive and significant. In contrast, with referral based recruitment, migrants in low skilled occupations use their networks to recruit new workers. Hence for our theory to represent the more plausible account,  $\alpha_6$  and  $\alpha_7$  in equation (15) should be positive and significant, and greater than  $\alpha_8$  and  $\alpha_9$ .

To further distinguish referral from supply side explanations, we test the effects of city-wide unemployment (as an inverse proxy for the level of economic activity) and pro-labour legislation on our dependent variable. We then explore whether belonging to the same social group is among the migrant attributes that more precisely identify network effects. Variation in network effects across social groups from the same source area, render the information-dispersion hypothesis less plausible since weak ties are necessary and sufficient for vacancy information to reach would-be workers. In Wahba and Zenou (2005), such information spreads through weak ties and social identities do not enter the frame. Our hypothesis of social identity effects intensifying under referral based recruitment is consistent with specific social groups obtaining strongholds in particular industries, restricting job access to group members (Holmstrom 1984; Chandavarkar 1994).<sup>13</sup>

---

<sup>13</sup> Such clustering is also, in principle, consistent with information sharing networks where labour market information is restricted to specific groups. While this is a prevalent explanation in the literature, there are two

## 4. Data and descriptive statistics

### 4.1 Data

We derive our dataset from a large nationally representative employment survey implemented by the Indian National Sample Survey Organisation (NSSO) in 1999-2000: the 55<sup>th</sup> which at the time of writing was the most recent available “thick” (i.e. comprehensive) survey round. The survey contains detailed information on a range of demographic, socio-economic and employment characteristics of every individual household member, and basic information (such as location, size, caste/religion, income and assets) for households as a whole.<sup>14</sup>

We limit the dataset to individual working migrants in the seven largest urban agglomerates (UAs): Ahmedabad, Bangalore, Calcutta, Chennai, Delhi, Hyderabad and Mumbai. We reconfigured the dataset according to the location codes that make up UAs rather than cities narrowly defined to ensure that individuals commuting from outside a city’s boundaries are covered. We only use the seven largest UAs so that identical industry dummies may be deployed for each; in smaller UAs some industries are absent.

The NSSO data provides seven possible locations of last usual residence: migrants may originate from the district in which they currently reside, either from a rural (1) or an urban (2) area; they may have come from the state in which they currently reside but from a different district, again either from a rural (3) or an urban (4) area; they may have come from a rural (5) or an urban (6) area from another state; or from another country (7). For inter-state migrants, the NSSO data provides

---

objections. As Winters et al (2001) demonstrate on data from Mexico, keeping information about migration opportunities under wraps may be very difficult. Secondly, clustering is only plausible if consistent with employer interests. In contrast to referral-based recruitment, spreading vacancy information through networks does not resolve the screening and incentive problems confronting employers.

<sup>14</sup> The NSSO Employment Survey covers more than 120,000 households and more than 600,000 individuals, 37,146 of whom are living in the seven largest urban agglomerates, 12,284 are in addition currently working, and 4,723 are in addition migrants.

information on the name of the state. We decided to keep urban-urban migration despite that our theoretical sketch applies more naturally to rural-urban migration. Areas classified in the NSSO data set as ‘urban’ include small towns in predominantly rural areas, which should not be excluded. We drop the relatively small categories from abroad and from the same district (both rural and urban).<sup>15</sup>

We construct all possible pairs of individuals working in the same industry in the same UA (henceforth, we call UAs cities). However, we eliminate reflexive pairs – that is, for two individuals A and B in the same industry and the same city, we include pair AB but not pair BA. The industries that these individuals work in are at the ISIC five-digit level of aggregation, which is the highest level of disaggregation the data permits.<sup>16</sup> Because of this high level of disaggregation, not all industries have more than one observation for each city; in total, we are able to construct 2737 pairs.

#### *4.2 Descriptive Statistics*

Figure 2 plots the distribution of migrant pairs across the seven cities in our sample.

**[Figure 2 about here]**

The distribution of pairs is uneven across these cities, with Mumbai comprising over 47 per cent of the total sample, and Delhi only about 1 per cent. The proportion of migrants from the same region ranges from 32 per cent in Bangalore to 80 per cent in Delhi. Figure 3 presents the distribution of pairs across the top 10 industries in our sample.

**[Figure 3 about here]**

---

<sup>15</sup> The NSSO employment survey also asks about migration motives. We omit individuals who migrated to enrol in an educational institution or for marriage. In these cases, migration was clearly not linked to recruitment through referral.

The total numbers of pairs in these 10 industries comprise 71 per cent of our sample. Construction of Residential Buildings comprise 21 per cent of the number of migrant pairs and 75 per cent of the migrant pairs in this industry originate from the same source region.

Table 2 presents odds ratios of the likelihood of being from the same source region for different migrant pair attributes, including education, caste (Scheduled Caste-SC, Other Backward Caste – OBC, and Others), and religion.<sup>17</sup>

**[Table 2 about here]**

The odds ratios exceed unity when both migrants are illiterate or educated to primary level or to graduate level or when the individuals in the migrant pair are from backward castes (SC/OBC). The odds ratios also exceed unity when both migrants are from the same religion or from different religions, except when one migrant is a Hindu and the other a Muslim (for the three principal religious groups considered). The odds ratios are very high (over 2.0) when both migrants are illiterate, when one is illiterate and the other educated to primary level, when both are SC, when both are OBC, when one is SC and the other is OBC, and when both are Muslim or Christian. The odds ratios, thus, suggest that migrants are more likely to come from the same source region if they are less educated, from backward castes, and when both belong to minority religions. We next examine whether these observations receive econometric support.

## **5. Empirical Results**

A possible econometric problem with the pair measure is that it may induce spatial correlation between observations involving the same individual, and

---

<sup>16</sup> There are 117 industries in our sample which also include the service sector.

<sup>17</sup> Others are individuals who are neither SC, OBC nor Scheduled Tribe (ST).

consequently bias standard errors. Since we construct pairs at the industry-city level, the spatial correlation may be present between observations in the same industry-city cell, but not across industries in the same city, or across cities for the same industry. There is also a possibility of heteroskedasticity across industries and cities. Therefore, we estimate a cluster-correlated estimate of the variance-covariance matrix in all the probit models presented (Wooldridge 2002) where we allow for cluster correlations within an industry-city cell.

We estimate maximum likelihood probit models and first use the specification in equation (14) to identify the impacts of migrant pair attributes and the general stock variable on the dependent variable with results presented in Table 3.

**[Table 3 about here]**

In Cols (1), (2) and (3), we report effects of migrant pair attributes without industry and city dummies, with city dummies only, and with both industry and city dummies. We introduce three demographic controls – the product of the age of the two migrants in the pair, and its square, and a dummy variable valued one when both migrants are male (zero otherwise). The main variables of interest are the migrant pair attributes related to education. As argued in Section 2, if referral based networks are the mechanism by which migrants move to cities and jobs, having zero or little education should increase the probability that a migrant pair working in industry  $i$  and city  $j$  are from the same source area. We find support for this hypothesis. With industry and city dummies included, the probability that both migrants are from the same source area increases by 15% if both are illiterate. This variable remains strongly significant at the one per cent level when the city and industry dummies are omitted, with city dummies only, and with both city and industry dummies. When both are educated to the primary level, the probability that the migrant pair is from the same source region

increases by a range of 8-10% across these three specifications; the coefficients are significant at the one per cent level.<sup>18</sup> In contrast, when both migrants are educated to the secondary or the graduate level, coefficients are not statistically significant in any of these equations. This supports the idea that chain migration is a low education phenomenon (McEntarfer 2003).

The aggregate stock network variable is also not significant in any of these three specifications. As argued in Section 3, if the information dispersion hypothesis is correct, we would expect a positive and statistically significant coefficient on the aggregate stock. That it is not suggests that the process whereby migrants from the same source region end up in the same city and industry is not driven by the use of networks to access vacancy information. For the control variables, the product of age and its square is significant, and age has a non-linear effect on the dependent variable, first decreasing and then increasing. The coefficient on both migrants being male is insignificant in most specifications.

Our theory suggests that higher economic activity (and lower unemployment) at destination increases the likelihood of referral-based recruitment. We measure unemployment for the seven cities using the indicator of structural unemployment (defined as being unemployed throughout the year) from the NSSO employment survey and include this unemployment rate in Col. (4) of Table 3.<sup>19</sup> We find that a 10 % rise in unemployment leads to a 3.0% decrease in the likelihood of both migrants being from the same source region, with the coefficient on the unemployment rate significant at the one per cent level.

Our theoretical sketch also suggests that more pro-labour legislation will reduce the loyalty gains from referral-based recruitment. We capture the degree of

---

<sup>18</sup> These are marginal effects estimated from the probit models.

pro-labour legislation using the Besley-Burgess (2004) measure of labour regulations, which is available for India's major states.<sup>20</sup> We use this measure for the states in which Ahmedabad, Bangalore, Chennai, Hyderabad, Kolkata and Mumbai are located, and for Delhi, we use the value for the adjoining state of Haryana.<sup>21</sup> A higher value implies that labour legislation is more pro-labour. The Besley-Burgess measure is included in Col. (5) of Table 3. We find that a 10% increase in the pro-labour legislation measure reduces the likelihood that both migrants are from the same source region by 0.30%, and that the labour legislation coefficient is significant at the five per cent level. These results further support our theoretical approach.

Table 4 presents two robustness tests of the basic results.

**[Table 4 about here]**

Firstly, in Col (1) we introduce separate dummies for the twenty four states from which migrants in our sample originate.<sup>22</sup> This is to control for source region clustering of worker characteristics, as explained in Section 3. In Col (2), we only include individuals who have migrated from a different state than where they are currently employed. This is to address the possibility that the positive relationship between the migrant pair originating from the same source region and both migrants being uneducated may be caused by an omitted variable – the distance from the source region to the destination industry/city. If uneducated migrants are more likely to travel to nearby cities and certain industries, less skill-intensive industries will have

---

<sup>19</sup> To avoid co-linearity, we omit city dummies in specifications that include the unemployment rate and labour legislation variables.

<sup>20</sup> The Industrial Disputes Act of 1947 sets out the conciliation, arbitration and adjudication procedures to be followed during industrial disputes. The Act has been amended by state governments post-independence. Besley and Burgess code state level amendments to the Act as being pro-worker, neutral or pro-employer. Each pro-worker amendment is given the value 1, each neutral amendment the value 0, and each pro-employer amendment the value -1. These scores are cumulated over time and we use the cumulative scores in 1999, the year of the employment survey used in this paper.

<sup>21</sup> Several manufacturing plants in the UA of Delhi are located around Gurgaon, an industrial region in the state of Haryana, and one can assume that the city of Delhi has similar labour laws as Haryana.

more migrants from the same (nearby) source area, especially if the source area has a concentration of less-educated individuals. By confining our sample to inter-state migrants (and keeping in mind that most Indian states are large geographical areas), we control for the effect of distance to travel on the clustering of migrants from the same source region.<sup>23</sup>

When we include source region dummies, the effects of the two variables of interest – both migrants are illiterates and both migrants are educated to primary level – are positive and significant at the one per cent level. Thus, the effect of a positive relationship between lower skill intensity of networks and a higher likelihood that the migrant pair is from the same source region cannot be attributed to the possible concentration of illiterate and semi-literate migrants in a few source regions. When confining the sample to inter-state migrants, both migrants being semi-literate increases the likelihood that both migrants are from the same source region compared to the baseline specification in Table 3 (Col (3)). However, the coefficient on the migrant pair being both illiterate is not statistically significant.

Following equation (15), we next replace the aggregate stock variable with three disaggregated stock variables - stocks of illiterate workers (stock\_illiterate), stocks of workers educated to the primary level (stock\_low literacy), and stocks of workers educated to the secondary level and beyond (stock – medium and high literacy). The results are presented in Table 5, Col. (1). As our theory suggests, the coefficient on stock-illiterate is positive and significant at the one per cent level, while the coefficients on stock-low literacy and stock-medium and high literacy are

---

<sup>22</sup> There are forty-six source region dummies in all – twenty-three each for the two individuals in the migrant pair (we drop one dummy in each case to avoid multicollinearity). The dummy for a particular state takes the value one when the migrant is from that state, zero other wise.

<sup>23</sup> In constructing our dependent variable we exclude individuals who have migrated from the same district, thus reducing the possible role of distance to travel in explaining why migrant pairs in a particular city are from the same source region. Furthermore, the coarseness of the source area definition that we use works to our advantage



insignificant. The attribute variables for both migrants being illiterate and being educated to primary level remain statistically significant. That the coefficient on the aggregate stock variable is insignificant along with the finding that among the disaggregated stock variables only stock-illiteracy is significant further supports our referral hypothesis.

Finally, we study social identity effects. We examine in particular the influence of caste and religion and estimate the impacts on the dependent variable by introducing the caste or religious background of migrant pairs in Cols. (2) and (3) of Table 5. In Col. (2), we include three dummy variables for the three main caste backgrounds on which we have information and in Col (3), three dummy variables capturing the three main religions in India. For illustration, consider the first dummy variable *Both SC* which takes the value one if the migrant pair are both SC and zero otherwise. Similar dummy variables are constructed for *Both OBC*, *Both Others* as well as for *Both Hindu*, *Both Muslim* and *Both Christian*. We also include source region dummies in the probit equations augmented with caste and religion variables to control for possible source area concentrations of individuals from the same caste or religious background.

In Col. (2), all three dummy variables pertaining to the same caste are positive and statistically significant. Among these three ‘same caste’ dummy variables, the characteristic that both individuals are of SC background has the strongest effect on the probability of being from the same source region. For the ‘same religion’ dummies in Col (3), ‘Both Muslim’ and ‘Both Christian’ are positive and statistically significant, while the ‘Both Hindu’ dummy is insignificant.

---

as the more broadly defined the source area, the less likely that source areas will be markedly different in terms of concentration of worker characteristics.

These strong social identity effects are consistent with the idea that social groups regulate access to work, favouring others from the same caste or religious background. This may not prove but is strongly suggestive of referral based networks being the main mechanism through which clusters of individuals from the same source region in the same city and industry form.

## **6. Conclusion**

This paper contributes to the growing literature on how social networks affect migration and labour market outcomes but focuses on demand-side explanations for network-based labour market entry. We sketch a theory of recruitment where employers use employee networks as screening and incentive mechanisms when recruiting into low-skilled jobs. Our empirical strategy involved firstly the filtering out of peer emulation effects from network effects; this guards our results from criticisms suggesting that migrants cluster in particular industries because of preferences for working with others resembling themselves or because of source community cohesion. We tested and found support for our main hypothesis of a negative relationship between network effects and the educational level of migrants. While this is also consistent with Wahba and Zenou's (2005) prediction of a higher incidence of network based search among illiterate and semi-literate workers, the strong social identity effects that we observe are not. Furthermore, our theoretical perspective, in contrast to theirs, suggests that employers would be reluctant to advertise such jobs. This is because advertising and broad dispersion of information through networks are unable to remedy the information and incentive problems confronting employers and that referral based recruitment help employers to overcome. We also tested the effects of pro-labour legislation and economic activity

and found further support for our demand-side account. Two other robustness tests provided vital additional support for our referral-based perspective.

## References

Banerjee, B. (1983) Social Networks in the Migration Process: Empirical Evidence on Chain Migration in India. *The Journal of Developing Areas*, 17, pp. 185-96.

Besley, T. and Burgess, R. (2004) Can Labor Regulation Hinder Economic Performance? Evidence from India. *Quarterly Journal of Economics*, 119(1), pp. 1-41.

Chandavarkar, R. (1994): *The Origins of Industrial Capitalism in India: Business Strategies and the Working Classes in Bombay, 1900-1940*, (Cambridge: CUP).

Dasgupta, P. (2000): Economic Progress and the Idea of Social Capital, in Dasgupta, P. and Serageldin, I. (eds) *Social Capital – A Multifaceted Perspective*, (Washington DC: World Bank).

de Weerd, J. (2004) Risk-Sharing and Endogenous Network Formation, in S. Dercon (ed.) *Insurance Against Poverty*, (Oxford: OUP).

Government of India (2002-03) Annual Survey of Industries.

Granovetter, M. (1973) The Strength of Weak Ties. *American Journal of Sociology*, 78, pp. 1360-1380.

Hatton, T. J. and Williamson, J. G. (1998) *The Age of Mass Migration – Causes and Impacts*, (Oxford: OUP).

Holmstrom, M. (1984) *Industry and Inequality – The social anthropology of Indian labour*, (Cambridge: CUP) .

Ionnides, Y. M. and Loury, L. D. (2004) Job Information Networks, Neighbourhood Effects and Inequality. *Journal of Economic Literature*, 42(4), pp. 1056-1093.

Faberman, R. J. (2005) Studying the Labor Market with the Job Openings and Labor Turnover Survey. Working Paper 388, US Department of Labor.

Manski, C. F. (2000) Economic Analysis of Social Interactions. *Journal of Economic Perspectives*, 14 (3), pp. 115-136.

Marmaros, D. and Sacerdote, B. (2002) Peer and social networks in job search. *European Economic Review*, 46 (4-5), pp. 870-879.

Marsden, P. and Gorman, E. (2001) Social Networks, Job Changes and

Recruitment, in Berg, I. and Kalleberg, A. L. (eds): *Sourcebook of Labor Markets: Evolving Structure and Processes*, (Kluwer Academic Publishers).

Mazumdar, D. (1973) Labour Supply in Early Industrialization: The Case of the Bombay Textile Industry, *ECHR Second Series*, 25(3), pp. 477-96.

McEntarfer, E. (2003) Three Essays on Social Networks in Labor Markets, Doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Montgomery, J. D. (1991) Social Networks and Labour Market Outcomes: Toward an Economic Analysis. *American Economic Review*, 81 (5), pp. 1408-1418.

Munshi, K. (2003) Networks in the Modern Economy: Mexican Migrants in the U.S. Labour Market, *Quarterly Journal of Economics*, 118(2), pp. 549-599.

Munshi, K and Rosenzweig, M. R. (2006) Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy. *American Economic Review*, 96(4), pp. 1225-1252.

Newman, R. (1979): Social Factors in the Recruitment of the Bombay Millhands, in K.N. Chaudhury and C.J. Dewey (eds): *Economy and Society: Essays in Indian Economic and Social History*, (New Delhi: ).

Salop, S. C. (1979) A Model of the Natural Rate of Unemployment. *American Economic Review*, 69(1), pp. 117-125.

Stark, O. (1991): Migration Incentives, Migration Types: The Role of Relative Deprivation, in Stark, O (ed) *The Migration of Labor*, (Cambridge MA: Blackwell).

Wahba, Jackline and Yves Zenou (2005) Density, social networks and job search methods: Theory and application to Egypt. *Journal of Development Economics*, 78 (2), pp 443-473.

Wasserman, S. and Faust, K. (1995) *Social Network Analysis*, (Cambridge: CUP).

Wooldridge, J. (2002) *The Econometric Analysis of Cross Section and Panel Data*, (Cambridge, MA.: MIT Press).

**Table 1: Worker types, labour pools and labour productivity returns to education**

	<b>(a) Industry with manual, <math>u</math>-<math>s</math> jobs</b>	<b>(b) Industry with <math>s</math>-<math>s</math> jobs.</b>
	Labour productivity in $u$ - $s$ jobs: zero labour productivity returns to education	Labour productivity in $s$ - $s$ jobs: positive productivity returns to education
Low productivity worker in village network with no or little education	$\underline{\theta}$	$\underline{\beta}$
High productivity worker in village network with no or little education	$\bar{\theta}$	$\bar{\beta}$
Low productivity worker in urban pool with $e_1$ education	$\underline{\theta}$	$m(e_1)\underline{\beta}$
High productivity worker in urban pool with $e_1$ education	$\bar{\theta}$	$n(e_1)\bar{\beta}$

**Table 2. Odds ratios of the likelihood of being from the same source region**

<b>Characteristics of migrant pair</b>	<b>Odds ratio</b>	<b>Per cent of total number of pairs</b>
<b>Both illiterate</b>	<b>3.04</b>	6.4
<b>Illiterate and educated to primary level</b>	<b>2.13</b>	14.0
Illiterate and educated to secondary level	0.81	9.8
Illiterate and educated to graduate level	0.32	0.5
<b>Both educated to primary level</b>	<b>1.45</b>	11.6
Educated to primary level and educated to secondary level	0.71	21.2
Educated to primary level and educated to graduate level	0.37	1.0
Both educated to secondary level	0.63	23.9
Educated to secondary level and educated to graduate level	0.87	7.0
<b>Both educated to graduate level</b>	<b>1.07</b>	4.5
<b>SC-SC</b>	<b>2.50</b>	2.4
<b>SC-OBC</b>	<b>2.29</b>	8.3
SC-Others	0.50	6.0
<b>OBC-OBC</b>	<b>2.56</b>	17.0
OBC-Others	0.49	18.7
Others-Others	0.86	43.1
<b>Hindu-Hindu</b>	<b>1.27</b>	76.0
Hindu-Muslim	0.49	13.7
<b>Hindu-Christian</b>	<b>1.32</b>	4.9
<b>Muslim-Muslim</b>	<b>3.58</b>	4.6
<b>Muslim-Christian</b>	<b>1.41</b>	0.4
<b>Christian-Christian</b>	<b>5.63</b>	0.8

Odds ratios are calculated as follows. First, for all migrant pairs from the same source region, the number of migrant pairs with a particular set of attributes is expressed as a ratio of the number of migrant pairs that do not have those attributes. Second, an analogous ratio is computed for all migrant pairs not from the same source region. The first ratio is then divided by the second ratio. Pairs are constructed of migrants that work in the same city and industry (see Section 4.1).

**Table 3. The impacts of migrant pair attributes and migrant stock on dependent variable: benchmark and augmented specifications**

	(1)	(2)	(3)	(4)	(5)
Product of Ages of Migrant Pair	-0.00024 (0.0002)	-0.0004 (0.0005)	-0.0011 (0.0004)***	-0.0051 (0.0001)***	-0.0004 (0.0003)
Square of Product of Ages of Migrant Pair	0.00000002 (0.00000082)	0.0000001 (0.0000001)	0.0000003 (0.0000001)***	0.0000001 (0.00000008)***	0.00000007 (0.000000001)
Both male	0.09 (0.10)	0.11 (0.12)	0.18 (0.11)*	0.03 (0.05)	0.17 (0.12)
Stock	0.0015 (0.0014)	0.0006 (0.0008)	0.0002 (0.0008)	0.0007 (0.0003)*	0.0019 (0.0010)**
Both illiterate	0.66 (0.16)***	0.51 (0.17)***	0.45 (0.17)***	0.20 (0.06)***	0.66 (0.18)***
Both educated to primary level	0.27 (0.09)***	0.28 (0.10)***	0.29 (0.11)**	0.09 (0.05)*	0.33 (0.10)***
Both educated to secondary level	-0.13 (0.12)	-0.04 (0.10)	-0.06 (0.08)	-0.07 (0.04)*	-0.14 (0.09)*
Both educated to graduate level	0.30 (0.24)	0.005 (0.29)	0.31 (0.27)	0.15 (0.08)*	0.42 (0.24)*
Unemployment	--	--	--	-0.04 (0.01)***	--
Labour Regulation	--	--	--	--	-0.08 (0.04)**
City dummies?	No	Yes	Yes	No	No
Industry dummies?	No	No	Yes	Yes	Yes
Log Pseudo-Likelihood	-1821.5	-1771.3	-1703.0	-1742.9	-1760.9
N	2737	2737	2737	2737	2737

a) Equations are ML probits.

b) \*, \*\* and \*\*\* significant at the 10, 5 and 1 percent level respectively;

c) Standard errors in parentheses are robust to heteroskedasticity and clustered residuals within industry-city cells.

**Table 4. Migrant pair attributes - tests of robustness**

	(1)	(2)
Product of Ages of Migrant Pair	-0.0008 (0.0003)***	-0.0008 (0.0005)
Square of Product of Ages of Migrant Pair	0.0000002 (0.0000001)**	0.00000008 (0.0000002)
Both male	0.15 (0.1)	0.67 (0.23)***
Stock	0.0004 (0.001)	-0.0009 (0.002)
Both illiterate	0.57 (0.13)***	0.33 (0.32)
Both educated to primary level	0.28 (0.09)***	0.41 (0.13)***
Both educated to secondary level	-0.02 (0.09)	-0.13 (0.11)
Both educated to graduate level	0.35 (0.29)	0.25 (0.44)
City fixed effects?	Yes	Yes
Industry fixed effects?	Yes	Yes
Log Pseudo-Likelihood	-1617.5	-794.5
N	2727	1324

a) Equations are ML probits.

b) \*, \*\* and \*\*\* significant at the 10, 5 and 1 percent level respectively;

c) Standard errors in parentheses are robust to heteroskedasticity and clustered residuals within industry-city cells.

d) Col (1): Dummies for source regions included– each region is assigned a value one if the migrant is from that region, zero otherwise – see Footnote 21 for more details on the construction of the dummies; Col (2): Restricted to inter-state migrants.



**Table 5. Stock network and social identity effects**

	(1)	(2)	(3)
Product of Ages of Migrant Pair	-0.0008 (0.0003)***	-0.0011 (0.0003)***	-0.0010 (0.0004)***
Square of Product of Ages of Migrant Pair	0.0000002 (0.00000001)**	0.0000003 (0.0000001)***	0.0000003 (0.0000001)***
Both male	0.21 (0.1152)*	0.018 (0.0009)	-0.001 (0.1112)
Stock		0.0004 (0.0008)	0.0008 (0.0008)
Stock-illiterate	0.0054 (0.0023)***	--	--
Stock-low literacy	-0.0068 (0.0071)	--	--
Stock-medium and high literacy	-0.0029 (0.0032)	--	--
Both illiterate	0.40 (0.19)**	0.56 (0.14)***	0.53 (0.13)***
Both educated to primary level	0.29 (0.12)**	0.18 (0.09)**	0.18 (0.09)*
Both educated to secondary level	-0.06 (0.08)	-0.04 (0.08)	-0.02 (0.10)
Both educated to graduate level	0.35 (0.27)	0.38 (0.31)	0.39 (0.30)
Both SC	--	0.72 (0.34)**	--
Both OBC	--	0.46 (0.12)***	--
Both Others	--	0.33 (0.09)***	--
Both Hindu	--	--	0.20 (0.15)
Both Muslim	--	--	0.99 (0.47)**
Both Christian	--	--	1.39 (0.66)**
City dummies?	Yes	Yes	Yes
Industry dummies?	Yes	Yes	Yes
Source Area dummies?	No	Yes	Yes
Log Pseudo-Likelihood	-1689.4	-1585.8	-1585.1
N	2737	2727	2727

a) Equations are ML probits.

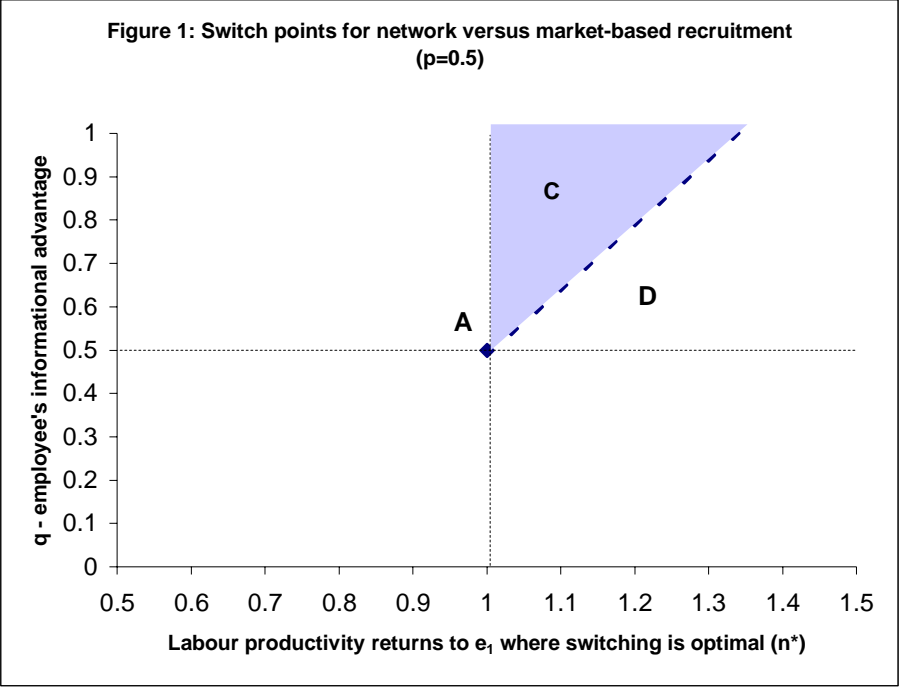
b) \*, \*\* and \*\*\* significant at the 10, 5 and 1 percent level respectively;

c) Standard errors in parentheses are robust to heteroskedasticity and clustered residuals within industry-city cells.

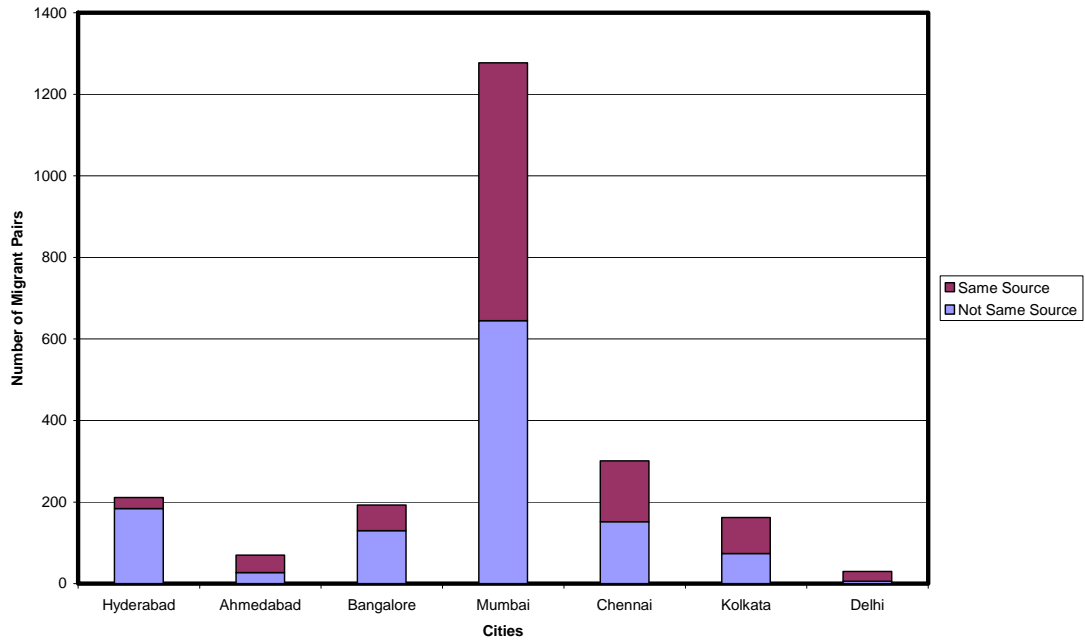
d) Col (1): Stock split into stocks of illiterate workers (stock-illiterate), stocks of workers educated to primary level (stock-low literacy), and stocks of workers educated to secondary level and beyond (stock-medium and high literacy).

Col (2): Includes 'same caste' variables (see Section 5).

Col (3): Includes 'same religion' variables (see Section 5).

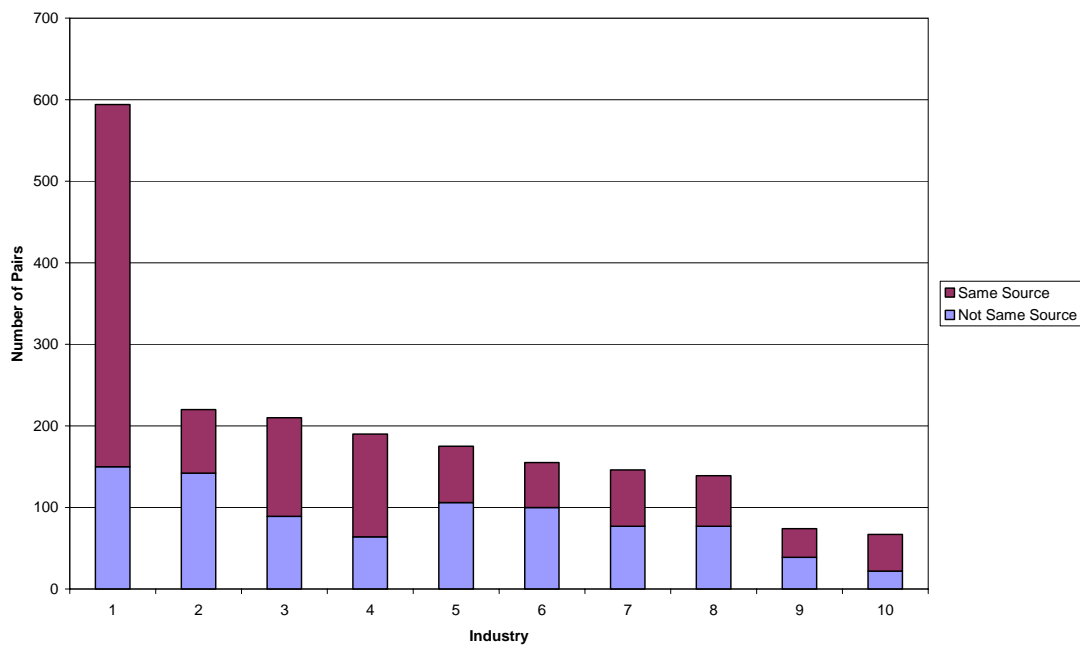


**Figure 2. Distribution of migrant pairs by cities**



**Source:** authors' calculations.

**Figure 3. Distribution of migrant pairs by industries**



**Source:** authors' calculations.

Top Ten Industries in terms of numbers of pairs; Industries: 1 – Construction of Residential Buildings; 2 – Textile and Clothing Accessories; 3- Employment of Domestic Help; 4 – Defence-related Activities; 5 – Retail Sale of Food Items; 6 –Restaurants without Bars; 7- Taxi Services; 8- Manufacture of Gold Jewellery; 9 – Software Publishing; 10- Painting and Decorating. The total number of pairs in these ten industries comprises 71 per cent of total sample.