Development Economics and Public Policy

WORKING PAPER SERIES

Paper No. 23

NETWORKS IN THE TRADITIONAL ECONOMY: EVIDENCE FROM INDIA

Vegard Iversen
Kunal Sen
Arjan Verschoor
Amaresh Dubey

IDPM, University of Manchester

June 2007

ISBN: 978 1 904143 95 6

Further details: Institute for Development Policy and Management
Published by: University of Manchester,
Harold Hankins Building, Precinct Centre,
Oxford Road, Manchester M13 9QH, UK
Tel: +44-161 275 2814 Fax: +44-161 273 8829
Email: idpm@manchester.ac.uk Web: www.manchester.ac.uk/idpm

Networks in the Traditional Economy: Evidence from India

Vegard Iversen[£], Kunal Sen[**], Arjan Verschoor[+] and Amaresh Dubey[++]*

Abstract

There is a broad consensus among economists that social networks impact on migration and labour market outcomes. Much of the empirical literature has focused on job search and thus supply-side explanations of network effects. This paper examines a particular demand-side explanation – the use of networks as an optimal recruitment strategy. Our theory of optimal recruitment predicts 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. It also predicts social identity to influence network access. Developing and implementing an empirical strategy to test for these relationships, we use migration data from an all-India Employment Survey to find strong support for demand-side explanations of the role of social networks in influencing migration behaviour. The negative association to skill-intensity suggests that demand-side driven labour market failure is likely to be a particularly severe problem in developing economies.

[£*] *International Food Policy Research Institute (IFPRI), New Delhi, CG Block, NASC Complex, Dev Prakash Shastri Road (Todpur), Pusa, New Delhi 110012. Ph: +91 11 25846565. e-mail:v.iversen@cgiar.org*

[**] *Institute for Development Policy and Management (IDPM), University of Manchester*

[+] *School of Development Studies, University of East Anglia, Norwich*

[++] *National Council of Applied Economic Research, Delhi, and North-East Hill University, Shillong, India*

1 Introduction

A large literature has addressed the impacts of social interactions and social networks on labour market outcomes (Ioannides and Loury 2004, Granovetter 2005, Wahba and Zenou 2005). A particular strand of this literature has emphasised the importance of network and peer emulation effects in influencing migration behaviour within or across countries (Hatton and Williamson 1998, Munshi 2003, Bayer et al. 2004).[1] The empirical literature has, however, found it hard to distinguish between network effects, which are transmitted to the migrant by virtue of the social connections he has access to, and peer emulation effects which are transmitted to the migrant through the local neighbourhood that he belongs to, whether at origin or destination. More importantly, the empirical literature has not distinguished between the case where employers actively recruit through employee networks from that where networks simply transmit information about job openings in a job search process. Thus, the literature has tended to conflate demand-side and supply-side explanations of the manner social networks impact on labour market outcomes and migration behaviour, treating them as one and the same thing.

To see why these effects are fundamentally different from each other consider the case of a male, Anand, residing in a village in India, and contemplating whether to migrate to Mumbai. Social interaction and network effects may influence Anand's decision to migrate in four ways. Firstly, Anand may migrate to Mumbai because many of his village friends have migrated to Mumbai earlier, and he would like to emulate their behaviour. Secondly, Anand may migrate to Mumbai because his co-villager Vijay in Mumbai has heard about and informs Anand (and possibly others in the village) about a vacancy. Thirdly, Anand may take a more proactive role and ask his relatives, Vijay and other of his city-based village friends to search for a job for him. Finally, Vijay's employer may ask Vijay to look for a suitable person to fill a new vacancy, and Vijay specifically *asks* Anand to take up the vacant position in Vijay's firm. The first of the mechanisms by which Anand migrates to Mumbai is the peer emulation effect, the support for which has been established in the empirical literature. The second and third mechanisms capture the information-dispersing role of social networks, where networks may resemble local public goods. In the latter case networks are non-market institutions that may correct market failures by providing job seekers with information about vacancies. The fourth mechanism involves the *explicit* use of networks in job recruitment, where network access is likely to be less open and restricted to some individuals and social groups because of the contrast between sharing information, which is an innocent act, and being someone's workplace guarantor. Bringing a new colleague to one's workplace is a sensitive and risky enterprise which changes the nature of the transaction, narrows the pool of potential recruits and is expected to strengthen the importance of social identity compared to other network mechanisms.

Which of these effects that dominate in determining migration behaviour will be crucial for ascertaining whether social networks enhance or diminish economic welfare in traditional societies. To see why, consider the existing literature on rural-urban migration which takes for granted that the urban labour markets that matter for the poor are level playing fields. However, if the propensity to find jobs through social networks is systematically higher in low or unskilled occupations, labour

[1] Two other studies that find the presence of peer emulation and network effects in job search and recruitment are Marmaros and Sacerdote (2002) and Antoninis (2006).

market failure would be of particular significance in poor economies where such jobs dominate. But any major concern would quickly dissolve if networks are local public goods that transmit information about job vacancies irrespective of the socio-economic background of the beneficiary.[2] In such a case, the role of public policy is straightforward: to strengthen the process of disseminating information about vacancies.

Policy conclusions are much less cut and labour market failures potentially more severe if network use is demand side driven and determined by optimal recruitment considerations by industrial and other employers. If such incentives are stronger for recruiting into low skill jobs, doubts about migration as an equitable route for escaping poverty would arise. So far, however, the literature on networks and labour markets has focused on supply-side explanations for the high incidence of network use in low-skill occupations in developing countries (e.g. Wahba and Zenou 2005)[3]. As a result, the prevalence and gravity of labour market failures, especially the implications for poor low- and unskilled workers, have been largely overlooked.

This paper presents the first attempt to disentangle demand-side and supply-side effects of social networks, and the effects of social networks from peer emulation in determining migration behaviour. We begin by sketching a theory of optimal network-based recruitment that offers new insights into and predictions about the relationship between network use and the skill or human capital intensity of jobs. Our demand-side analysis predicts more powerful network effects in low skilled

[2] Among a random sample of migrants in Delhi, Banerjee (1983) found the access to network services to transcend caste-boundaries with family and village ties serving similar functions in aiding migration. In such a case, networks fit the description of unrestricted, local public goods. Similarly, using Mexican data, Winters et al. (2001) found family or strong ties to provide no advantages over weak community ties in aiding migration, suggesting that cumulative public knowledge about migration becomes open to all migrants irrespective of their social or economic backgrounds. Munshi (2003;551) analysis of Mexican data makes a specific reference to the importance of belonging to the same origin community (paisanaje) as the dimension of social identity that establishes the right to network access which is consistent with Winters et al's (2001) observations.

[3] Iversen (2006) reports that more than 90 % of the jobs which migrants from Honnavara village in Mandya district in Karnataka moved into were arranged informally. See section 2.1 for more.

occupations and for less educated workers, and that group identity and social ties take on more significance when employers recruit through employee networks. In addition, network-based recruitment is predicted to be more prevalent during economic booms and less prevalent when legislation is worker-friendly.

We propose an innovative empirical strategy to distinguish neighbourhood effects from network effects, and the information dispersion hypothesis of network use from the optimal recruitment hypothesis. Using a large data-set of migrants to cities in India, we find strong evidence in favour of the optimal recruitment hypothesis specifically, and for demand-side explanations of social network use in migration behaviour in general.

In a path-breaking paper, Munshi (2003) is able to identify network effects in a study of Mexican migrants to the United States. Using variation within each origin-community's network over time, Munshi controls for selectivity in the migration decision and for the endogeneity of the network itself. However, Munshi's identification strategy cannot separate the information dispersion hypothesis from the optimal recruitment hypothesis. His empirical analysis also does not distinguish between workplace clustering caused by source community cohesion (peer emulation effects) and clustering attributable to employer preferences.

Wahba and Zenou (2005) provide a theory-based test of the presence of social network effects in job search. Their supply-side theory predicts social network based search to be more common among less educated workers. The mechanism creating this bias is that unskilled job searchers lack the ability to read job-ads and vacancy signs in the windows of prospective employers which provides a stronger incentive for searching through friends and contacts. With information spreading through networks of weak ties, Wahba and Zenou in effect test the equivalent of the information dispersion hypothesis of network effects in migration behaviour. They argue that Egyptian labour market data support the hypothesis of a negative association between skill-intensity and network-based job search, but use data and an empirical specification that is too coarse to persuasively support this conclusion. Moreover, their focus on weak ties suppresses social identity as a determinant of network access in notable contrast to the theoretical and empirical analysis presented below.

To identify the referral based and demand side type of network effects, we proceed in three steps. Firstly, we construct pairs of migrants working in a particular industry and in a particular city, and examine the likelihood that a migrant pair originates from the same source area. As in Munshi (2003), this provides an indirect test of network effects. Secondly, we control for peer emulation effects on migration behaviour by introducing city and industry fixed effects, and by examining whether the likelihood that the migrant pair is from the same source area is relatively higher for less educated migrants and migrants in low skilled occupations. Our identifying assumption is that the propensity to emulate peers should be equally strong across migrant categories. Next, following theoretical predictions and facilitating further distinction between information dispersion and optimal recruitment types of network effects, we test for the impacts of pro-worker legislation and economic activity on network use.

A final identification strategy is to model the effects of social structure on migration behaviour. If network services are information-dispersing community level public goods (e.g. Winters et al 2001) or information spreads through weak ties (Wahba and Zenou 2005), a specification of an aggregate network effect would be sufficient to capture such a process (e.g Carrington et al 1996), making a more disaggregated analysis based on social identities redundant. Conversely, empirical support for disaggregation by social identity would tip the balance further in favour of referral and demand-side based network effects.[4]

The rest of the paper is organised as follows: Section 2 sketches the theory where employer incentives for using the rural networks of existing employees for recruitment are stronger in unskilled and low-skilled occupations, and where a switch point for ranking the use of village networks over market-based recruitment can be identified. The location of this switch point is sensitive to the skill- or human capital intensity of industrial production and jobs, as well as other employer advantages from network-based recruitment. Section 3 outlines our empirical strategy for identifying this optimal recruitment hypothesis. Section 4 describes the data and the empirical specification and provides some descriptive statistics. Section 5 presents the results of the empirical analysis. Section 6 concludes the paper.

2. Sketching a theory of recruitment

2.1 Skill, migration and social networks

In Montgomery (1991), employers resort to selective recruitment using the networks of talented employees. With assortative matching in contacts, these networks comprise similar, high quality individuals, but the theory provides no clear guidance about the prevalence of referral-based recruitment in high or low-skilled jobs or in developed or developing economies. Similarly, sociological research has

[4] Notice that social identity effects could creep in if the information dispersion process is exclusive in the sense that existing migrants share information about vacancies only with relatives, caste-fellows or co-villagers. It is, however, hard to envisage how such sharing could be sustained without being commensurate with employer interests: a key question is why an employer should condone such restrictive information sharing unless the employer benefits from relying on employee networks. In a rare study of India, the relevant dimension of identity determining access to network services among migrants in Delhi was sharing a native place (Banerjee 1983). Importantly, however, Banerjee's analysis does not distinguish between information sharing and providing referral. We expect the latter to follow social demarcations – whether kin, caste, religious affiliation or being a co-villager is the relevant dimension of a person's social identity that establishes entitlements to such a valuable network service is an important empirical question.

provided little guidance in explaining variation in the resort to referral use across different types of industries.

Using anthropological evidence from Bombay and focusing on motives for referral-based recruitment and the link to the skill-intensity of jobs, Holmstrom (1984;219) observes: *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, employers find the networks of their existing employees "*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.)*." In a study of migrants in Mumbai, Gore (1970) reported that 68 % of blue collar and 38 % of white collar workers received help with finding a job. These are powerful illustrations of what has widely been regarded as a key challenge of labour management in India, namely strategies for controlling turnover among low and unskilled workers (e.g. Myers 1958, Mazumdar 1973 and Newman 1979).

Echoing Holmstrom (1984) and the above references, theorists have proposed that referral based recruitment strengthens loyalty and attachment to jobs and workplaces (Fernandez and Castilla 2001). In addition, referrals reduce employer uncertainty about the productivity of prospective recruits (Marsden and Gorman 2001).

Labour turnover rates in India are notoriously high. In spite of a dearth of systematic and sufficiently detailed labour market statistics, the Annual Survey of Industries (2002-03), reports annual quit rates between 1.67 % and 120 % with an overall average of 18.67 % and a public enterprise average of 9.57 %. In contrast to absenteeism, which shows very little inter-industry variation and fluctuates around 10 %, labour turnovers vary considerably (σ=16.86). There are also striking contrasts in turnovers by state with the lowest private sector turnover rates among the main states are found in West Bengal (5.22 %) and Kerala (8.56 %), while the highest, 49.45 % is reported for Punjab.

A more detailed empirical illustration is provided in Table 1 where we report turnover rates for a small sample of un- and low-skilled young migrants (N=81), primarily to Mumbai and Bangalore, based on whether their first entry to primarily urban labour markets occurred through weak, strong or no referral. Strong referral implies taking up a job in the same workplace as the middleman, i.e. the person providing referral. Among others, the destination jobs include work in small eating places, in garage workshops and as assistants to lorry drivers. We also include information about the relationship between the migrant and the middleman. Strong social ties are defined to comprise kin including nuclear family members, uncles and aunts and where the boundary, somewhat arbitrarily, are set at cousins. Notice also that most cases of weak ties refer to co-villagers while almost all reported quits are voluntary (See appendix for a brief description of the data). Since the number of observations for each sub-category is small, we also report mean turnover rates after removing the most influential outlier.

Table 1: Type of referral and social ties involved in arranging the first job of migrants at migration destination[5]

Type of social tie	Strong	Weak	Strong	Weak	No referral
Type of referral	Strong	Strong	Weak	Weak	
n=	18	20	19	19	5
Total number of jobs	56	75	66	104	34
Total number of work years	234.5	363.5	240	281.5	57
Annual turnover rate	0.162	0.1513	0.1958	0.3019	0.5087
Annual turnover rate with 1 outlier dropped	0.13	0.1199	0.1681	0.1749	0.4473

The small number of observations makes the table indicative rather than conclusive about the interplay between the social distance between migrant and middleman, the type of referral provided and labour turnovers. Notice, in particular, the contrast in turnovers for migrants when no referral was provided. Table 1 highlights the need for theorising and for empirical analysis based on larger and more representative data-sets.

Turning, therefore, to theory, consider an urban, industrial firm comprising an owner and a migrant worker about to expand. Let $\overline{\theta}$ denote a high productivity and $\underline{\theta}$ a low productivity worker in a binary distribution of worker types.[6] The owner must

[5] The annual turnover rate for each category is calculated as: (Total number of jobs – n)/(Total number of workyears). Table 1 is based on worklife histories of a sample of migrants from purposively selected villages in Mandya (four villages) and Udupi (four villages) Districts, Karnataka, India. Households within each village were randomly selected and migrants selected for interview if they were below 15 years of age at the time of leaving home. The data were collected between October 2004 and November 2005. From a total number of interviews of 97, cases of migration into agricultural work were removed because agricultural destination work involves a sequence of short-term contracts rather than an employer-employee relationship with the potential to last. We also removed the small number of observations involving females migrating to work as domestic servants, which is characterized by lower turnovers than those reported above, reducing the number of observations to 81. More details are available from the authors.

[6] Following Munshi and Rosenzweig (2006), we assume that performance-contingent wage contracts are difficult to implement and distinguish between productivity in the execution of work tasks on the one hand and losses to the employer caused by high turnover rates on the other.

decide whether to recruit through the urban labour market or use the employee's village network. The first option offers the prospect (1):

$$p\overline{\theta}+(1-p)\underline{\theta}, \qquad (1)$$

where (1) is the expected value added in output from recruiting a new worker, and p is the probability that the new recruit is highly productive. When p=1, the employer's screening technology is perfect. Following Marsden and Gorman's (2001) reasoning, the employee has an informational advantage and can improve on (1) if recruiting through his village network provides (2):

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

To start with, let the distribution of labour or *task* productivities in the urban labour market and the worker's village network be identical.[7] The *informational* gain from network-based recruitment is now proportional to the gap in task productivity $V=\overline{\theta}-\underline{\theta}$, since the gain is $(q-p)V$. For most low skilled jobs, differences in task productivity, V are likely to be limited, thus suggesting a marginal informational gain.

Consider, next, the issue of loyalty, where improved discipline and lower turnovers represent the employer's principal gains from referral-based recruitment: Let

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

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

[7] This, we argue below, is plausible for unskilled jobs and jobs with low skill-intensity..

In terms of workforce stability, employer gains from referral-based recruitment, S, are given by:

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

where t_m is the labour turnover rate when recruiting through the market and t_n the turnover when recruiting through the employee's network. C is the unit cost of replacing a worker and s is skill-intensity. Cost of replacement are likely to increase with skill-intensity so that $dC/ds > 0$. [8]

Motivated by the references and evidence presented above, we assume that $t_m - t_n$ attains its maximum for un- and low-skilled jobs. The impact on S of a surge in skill-intensity becomes:

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

Since the gap between turnover rates is maximised for un- and low-skilled jobs, we have that:

$$\left|\frac{\partial t_n}{\partial s}\right| < \left|\frac{\partial t_m}{\partial s}\right|, \tag{6}$$

so that the first term in (5) is negative, while the second is positive. Hence, for increasing skill-intensity of jobs, whether referral-based recruitment becomes more attractive or not, depends on the relative strength of these two effects.

[8] Not only turnover rates but also the informational advantage from network based recruitment may depend on skill intensity (s) and social distance (SD). Starting with s, productivity differences between high and low productivity workers vary across jobs and professions. However, screening of more educated people is likely to be more accurate (because of proof of educational attainments), reducing the employee's informational advantage. Simultaneously, an increase in s means that the quality of the draw from the village network is negatively affected because of the higher prevalence of educated people in the urban labour market. Hence, skill intensity may have three effects on the informational advantage since recruiting through the village network (a) is more valuable for higher V, (b) less valuable if screening is better for more educated workers and (c) less valuable because the draw from the village network is negatively affected by the relative scarcity of rural, educated workers. (b) and (c) are likely to dominate (a), suggesting that the net effect on the informational advantage from recruiting through the migrant's village network is declining in s. The second point, regarding SD is more straightforward since p and thus the migrant worker's informational advantage is likely to depend on SD, i.e. the migrant's social proximity to the new recruit which unambiguously increases the advantage to the employer from referral-based recruitment and will *reinforce* our argument about the impact of social identity. Since the direction of the relationships of interest do not change, we prioritise expositional simplicity.

Consider, next, the effect of social distance on employer gains from referral-based recruitment. As argued overleaf, the act of bringing someone to the workplace is risky compared to sharing information about a vacancy. This is expected to narrow the pool of potential recruits. SD captures the intensity of the relationship between the employee and the new recruit, reflecting the gravity of sanctions that may be imposed on the latter should he decide to renege. In terms of loyalty and workforce stability, the employer is thus expected to benefit from closer social ties.

The impacts on social distance of the gains from referral-based recruitment are given by:

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

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

In short, closer social ties between the employee and the new recruit will enhance the loyalty gains from network-based recruitment.

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

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

For a given skill-level, workers recruited through referral are expected to be more loyal and thus less inclined to respond to a heating up of the economy.[9] Hence, t_m is likely to be more sensitive to a change in *G* than t_n. In a tighter labour market, referral-based recruitment becomes more attractive to employers.

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

[9] On US-data, Fabermen (2005) finds a negative correlation between voluntary quits and unemployment. On the link between growth and attrition in India: quarterly turnover rates in outsourcing have been reported to be as high as 25 % in some firms (Computer Business Review 5th July 2005).

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

where $\frac{\partial t_m}{\partial L} < 0, \frac{\partial t_n}{\partial L} < 0$ 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 will now decline since people, because of better working conditions are more keen to hold on to their jobs, ceteris paribus. Hence, the risks from recruiting through the market decline, lowering the gains from referral-based recruitment.

2.2 Worker pools – the urban labour market, the village network and the skill-intensity of jobs

We now characterise the pools of workers in the urban labour market and the village network. To start with, consider an industry dominated by unskilled, type *j* jobs and let the binary distribution of individual productivities in the village network for type *j* jobs be X_j^R. The village network will contain many individuals capable of doing type *j* jobs well. The employer's alternative is to recruit through the urban labour market. Given the nature of *j*, we assume $X_j^R = X_j^U$. Notice that this does not rule out systematic differences in worker attributes 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 than his village counterpart. However, the manual and simple nature of *j* minimises labour productivity differences since productivity returns to education or skills in jobs of type *j* are close to zero.

Consider, instead, an industry with more complex and skill-intensive production and where recruitment is for a semi-skilled or more human capital-intensive job, *k*. In performing *k*, education raises labour productivity. This effect will depend on the educational levels of workers, as shown in table 2:

Table 2

	Example (a) Industry where manual, unskilled jobs dominate	**Example (b) Industry with more skill- and human-capital intensive production**
	Labour productivity in unskilled type j job: no labour productivity returns to education	Labour productivity in semi-skilled type k job: positive productivity returns to education
Low productivity, worker with low or zero education in village network	$\underline{\theta}$	$\underline{\beta}$
High productivity, worker with low or zero education in village network	$\bar{\theta}$	$\bar{\beta}$
Low productivity, worker with education level e_1 in urban unskilled labour pool	$\underline{\theta}$	$m(e_1)\underline{\beta}$
High productivity, worker with education level e_1 in urban unskilled labour pool	$\bar{\theta}$	$n(e_1)\bar{\beta}$

Column (a) displays type j jobs with zero productivity returns to education while column (b) summarises the productivity of each worker category in jobs of type *k*. Column (b) thus expresses the productivity of individuals in the urban (with education level e_1) labour pool as functions of the productivity of rural, unskilled workers in semi-skilled jobs. With zero labour productivity returns to education (or skills) in semi-skilled jobs, $m = n = 1$. With positive labour productivity returns to education in job *k*, $m(e_1), n(e_1) > 1$. Moreover, the more human-capital (or skill-) intensive production, the higher these productivity returns and thus $n(e_1)$.

The employer will now balance the informational advantage of the employee and the gains, *S*, from lower turnovers against the potentially higher average labour productivity among the more educated workers in the urban pool and rank village network over market-based recruitment whenever:

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

The right hand side measures the net productivity returns to e_1 among low productivity workers and is positive when $n > 1$. We normalise by using $\underline{\beta} = 1$ as "numeraire". Moreover, let $\bar{\beta} = a\underline{\beta} = a$ where $a > 1$ for obvious reasons. The condition for recruiting through the urban labour market can be written as:

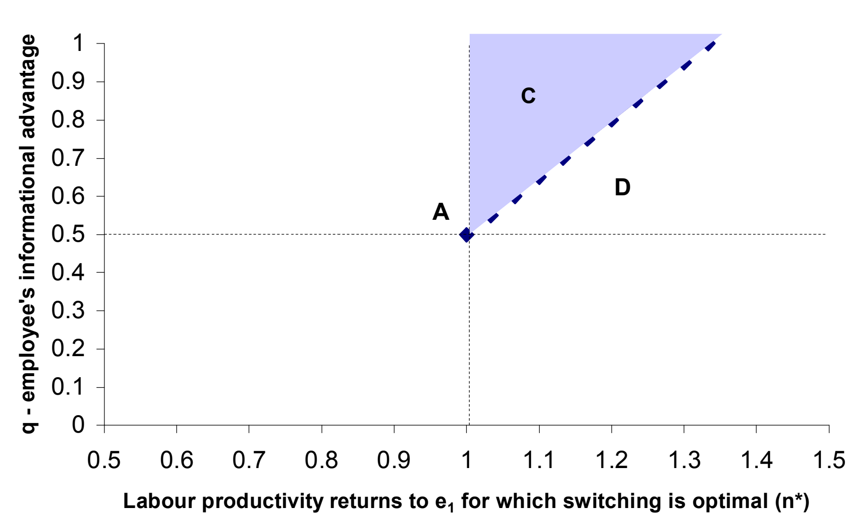
$$(np - q)(a-1) - S > (1-n). \qquad (11)$$

Reorganising and for given exogenous values of *q*, *p*, *a* and *S*, a switch point may be defined as:

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

To start with, consider the impacts of information and labour productivity returns to education, leaving *S* aside. Figure 1 displays the locus of points along which an employer is indifferent between recruiting through the employee's village network and the urban labour market. If other screening mechanisms are imperfect and the employee has a considerable informational advantage, positive productivity returns to the human capital of urban workers is required for employers to switch out of village-based recruitment. Figure 1 illustrates combinations of employee informational advantages and labour productivity returns to education which makes a particular recruitment strategy optimal. Southeast of point A, it is always preferable to recruit through the urban labour market since the informational advantages provided by the employee's village network are insufficient (and given that $n > 1$).

Figure 1: Switching points for network versus market-based recruitment (p=0.5)

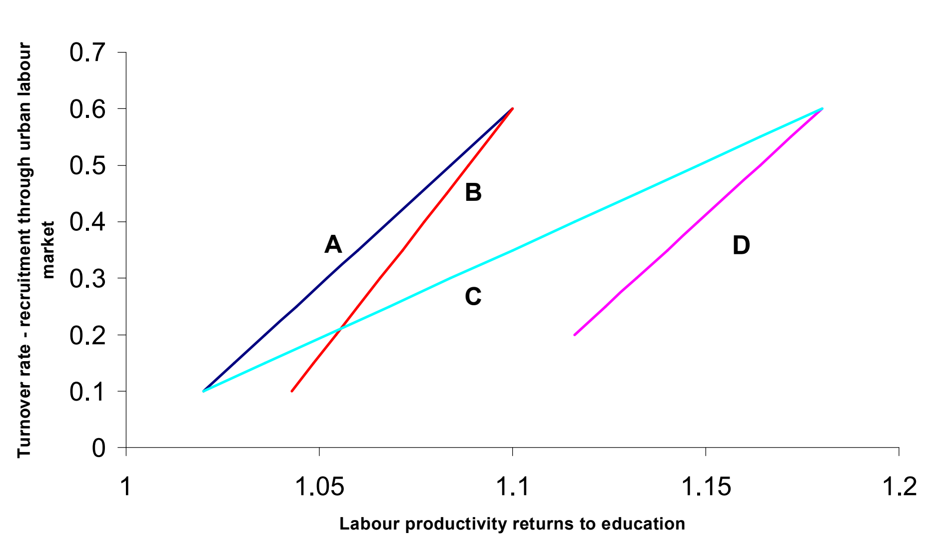


Region C, in contrast, provides combinations of informational advantages and labour productivity returns where recruiting through the village network is optimal. Region D comprises combinations for which employers recruit through the urban labour market.

2.3 Integrating employer savings on turnovers

Figure 2 now displays alternative switch-point loci for alternative parameter values.

Figure 2: Alternative switch point loci - complete model



Locus A displays the benchmark for given values of the informational gain ($q - p = 0.1$), the productivity difference ($V = 1.5$), a fixed turnover rate for recruitment through the village network ($t_n = 0.1$) and turnover costs ($C = 0.2$). The locus presents combinations of labour productivity returns to education and turnover rates when recruiting through the labour market for which the employer is indifferent between the two recruitment strategies. Along locus A, as turnover rates when recruiting through the labour market increase, the employer requires compensation through higher labour productivity returns to the human capital of urban workers to remain indifferent between the village network and the urban labour market. Locus B portrays the case of $V = 2.5$ and shows that a higher productivity difference between high and low productivity workers, increases the value to the employer of the employee's informational advantage. Locus C displays the effect of a doubling of unit turnover costs compared to A. For a given turnover rate for urban workers, a notable increase in labour productivity returns is required to "compensate", underscoring the importance of loyalty.

3. Empirical strategy

The key predictions of the theoretical sketch offered in the previous section are that, ceteris paribus and due to referral-based recruitment, network effects are stronger for hiring low and unskilled labour. Moreover, the effect of social identity is expected to be stronger, since referral-based recruitment narrows the pool of potential recruits. Our theory also predicts network effects to be stronger during economic booms and weaker when legislation is pro-labour. These predictions are distinct from an account of network effects whereby illiterate and semi-literate labourers are unable to access formal channels of job search (Wahba and Zenou 2005).

In this section we outline how we identify network effects (by filtering out peer emulation effects) and how we propose to detect the presence of referral-based effects as opposed to information-dispersion effects. It is worth noting at the outset that Manski's (1993) reflection problem does not apply in this set-up to the extent that work-related migration is indeed due to referral-based effects. The reflection problem arises if an individual's behaviour responds to mean current (or almost current) group behaviour. This (almost) simultaneous occurrence rules out the possibility of identifying cause and effect. If there are compelling reasons for believing that an individual's behaviour responds to past group behaviour, the reflection problem does not apply (Manski 2000). Since referral-based-recruitment-induced migration naturally implies that current migration responds to past migration, a plausible account of migration emphasising referral-based effects circumvents the reflection problem.

The empirical focus will be on the probability that a pair of randomly selected migrants who work in the same city and industry, with industry defined as narrowly as our data permit, are from the same source area. Our basic empirical specification is given by:

$$D_{ij} = \alpha_1 X_{ij} + \alpha_2 N_{ij} + \gamma_i + \delta_j + \varepsilon_{ij}, \qquad \text{(I)}$$

where D is a dummy variable that takes the value one when two individuals in the pair are from the same source region, zero otherwise; i and j denote industry and city respectively, X is a vector of control variables, γ_i are industry fixed effects and δ_j are city fixed effects, and N is a vector of our measures of network effects. We elaborate on these measures below.

The dependent variable is thus a dummy that takes the value one if two migrants working in the same city and industry are from the same source area. The reason that we use pairs of migrants to construct our dependent variable is best explained by considering the alternative. The dependent variable could also be the *percentage* of migrants in the same city and industry who are from the same source area. However, our preferred dependent variable allows us to uniquely map the LHS variable to the shared social characteristics of the matched pair. By doing so, we are able to investigate whether network effects are stronger for certain shared social characteristics. The specific shared social characteristics that we are interested in are the literacy levels of the pair, whether network effects are stronger for migrant pairs where both have little or no education versus those with some education and whether and if so which dimensions of social identity that matter for the labour market outcomes of migrant pairs.

We distinguish between recent and past network effects, which we perhaps rather fancifully term 'flow' and 'stock' effects, respectively. 'Flow' effects are exhibited when the probability that two recent migrants working in the same city and industry are from the same source area is related to *other* characteristics of that migrant pair, where the most important of such characteristics for our purposes is the skill-intensity of the jobs they are recruited into. 'Stock' effects are exhibited when the probability just referred to is related to the number of similar pairs that have migrated in the past. For the identification of stock effects, we construct an independent variable that takes the value of the number of migrant pairs working in the same industry and city that are both from the same source area. The coefficient on this variable is expected to be positive until (or unless) a network is congested. We define recent, as distinct from past migration, somewhat arbitrarily, as migration that occurred within the past five years.[10]

[10] The intention here is to construct stock variables such that serially correlated demand shocks are not likely to give rise to spurious correlation between migration from the same source area and the stock variables. For

The identification of network effects within this empirical set-up requires that we have successfully filtered out peer emulation effects, and the effects of the corresponding concentration of supplied and demanded worker characteristics in source and destination areas, respectively. The latter exemplify what Manski (2000) calls 'correlated effects' and are arguably rather innocuous in our context. The concentration of worker characteristics in particular source locations relevant to employment in specific cities and/or industries could pose a problem for identifying network effects. For example, if skilled weavers are disproportionately demanded in a certain city and supplied by a certain source area, we are at risk of falsely attributing the clustering of migrated skilled weavers in a particular destination location to the operation of network effects while the clustering is caused by a concentration of supplied and demanded worker characteristics in source and destination areas, respectively. This problem is not likely to be particularly severe in our application because, for reasons detailed in Section 4, we are obliged to work with rather broad definitions of both 'same source area' and 'same destination area'. However, we do test for the presence of this problem by checking robustness to the introduction of source area dummies.

Peer emulation effects in our context refer to the emulation by new migrants of successful past work-related migration strategies by migrants with similar characteristics, and may appear identical to network effects in the empirical analysis. They are an example of what Manski (2000) calls 'endogenous interactions', in which the propensity of an agent to behave in a certain way varies with the average

---

example, suppose that demand shocks in particular industries in particular cities are positively serially correlated and that these industries make disproportionate use of particular worker types. Positive correlation will then be found between stocks of worker types and migration from the same source area that is not attributable to network effects. A larger period that covers the definition of 'recent migration' reduces the likelihood of this problem. We elaborate below on the identification challenge posed by this problem, insofar as it remains.

behaviour of the group as a whole[11]. In our application, emulation by new migrants of previous migrants would lead to a clustering of migrants from the same source area in the same destination city and, possibly, industry. In that case, we are at risk of falsely attributing such clustering to the operation of network effects (be it referral-based network effects or information-dispersing network effects). We deploy the following strategy to filter out peer emulation effects so as to identify network effects.

We first construct our dependent variable such that the more obvious peer emulation effects are not considered in the analysis by limiting the analysis to employed migrants only. In that way, all migrants that have come in search of work but have so far been unsuccessful are excluded from the analysis. Peer emulation effects are likely to be at work in this group but referral-based recruitment by definition is not. Moreover, we adopt a definition of 'same industry' that is as narrow as possible. Here we take advantage of the fact that our data permit a highly refined disaggregation of the industry level at which we construct our dependent variable (for details see Section 4). Arguably, the more narrow the definition of 'same industry' in which a pair of migrants from the same source area work, the more likely it is that referral-based network effects rather than peer emulation effects are at work. The picture we have in mind here is that both emulation and referrals get migrants to the same city as their peers from the same source area, but that the likelihood of ending up in the same (narrowly defined) industry is greater for referral-based recruitment. If that picture is broadly correct, it follows that the more precisely we are able to define destination industry, the more successful we will have been in filtering out peer emulation effects.

[11] As noted above, their identification is problematic when this propensity is concurrent with group behaviour, due to the reflection problem.

Having constructed the dependent variable such that the more obvious peer emulation effects have been filtered out, we obtain a lower bound on the size of network effects, as follows. The inclusion of industry and city fixed effects, γ_i and δ_j respectively, provides us with a baseline probability for each city and industry that pairs of migrants, who work in the same city and industry, are from the same source area. The coefficients on dummies in the vector N_{ij} denoting the skill category that the individuals in the migrant pair belong to are therefore marginal effects associated with each skill category. To the extent that the construction of the dependent variable has not filtered out all peer emulation effects, we need the identifying assumption that a tendency to emulate peers is equally strong across migrant categories, in particular across the range from low to high-skilled labourers. This allows us to interpret differences across the coefficients on skill-related characteristics of migrant pairs as a lower bound on a measure of the size of a network effect. Following the same logic, differences across the coefficients on stocks of migrants disaggregated by skill type in the vector N_{ij} may be interpreted in the same manner.[12]

Let us return to our key predictions, then, and the elimination of rival accounts. According to our theoretical sketch, referral based effects are more likely when individuals are uneducated or with low levels of education. We construct four flow network variables – F_{ij}^{ILL}, F_{ij}^{LOWLIT}, F_{ij}^{MEDLIT}, and $F_{ij}^{HIGHLIT}$ which are dummy variables which take the value one when both the individuals in the pair 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, educated to primary

[12] To be precise, to the extent that the construction of our stock variables has not filtered out spurious correlation between recent migration from the same source area and particular stock variables due to positively serially correlated demand shocks, as argued above, the identifying assumption deployed here that a tendency to emulate peers is equally strong across migrant categories allows us to interpret differences across coefficients on stocks of migrants of particular skill types as a lower bound on the magnitude of network effects.

level, educated to secondary level and educated to graduate level respectively. Thus, our empirical specification can be re-written as:

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

Our theory of referral-based networks predicts that α_2 and α_3 are both positive and greater than α_4 and α_5. We then disaggregate the stock network measure by level of education of the migrant pairs that constitute the stocks. Thus, we disaggregate S_{ij} into S_{ij}^{ILL}, S_{ij}^{LOWLIT}, S_{ij}^{MEDLIT}, and $S_{ij}^{HIGHLIT}$ which are stocks of migrant pairs who have migrated from the same source region in time period $t > 5$ years, and where the superscripts have the same meaning as before. In this case, our empirical specification becomes:

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

We would expect that the presence of referral based networks will lead to α_6 and α_7 being positive and greater than α_8 and α_9.

However, a significantly lower probability for low and unskilled migrant-workers that a pair of them, working in the same city and industry, are from the same source area would not on its own rule out an interpretation along the lines of networks performing primarily an information-dispersing role, which substitutes for formal advertisement of jobs where ads are unlikely to be read and understood by illiterate and semi-literate would-be workers (Wahba and Zenou 2005). In the Wahba-Zenou hypothesis, the probability of hearing about a job vacancy increases, at a decreasing rate, as the number of weak ties increases (which in their formalisation is identical to the number of random contacts per unit of time). Their network variable – their 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 we are able to work with much more refined network variables, related to the presence of similar migrants, both past and present, in a precisely defined industry. Since our network variables are defined much more precisely than the Wahba-Zenou one, we have a chance of detecting referral-based effects. The challenge remains, however, to distinguish a referral-based interpretation from a job search-based one, even at the level of disaggregation at which we initially investigate the matter.

In order to distinguish the information-dispersing account from the referral-based recruitment account, we test whether belonging to the same social group is among the low-skilled migrant characteristics that more precisely identify network effects than the attribute 'low-skilled' on its own. Variation in network effects across social groups from the same source area renders the information-dispersion hypothesis less plausible. This is because weak ties are both necessary and sufficient for information about vacancies to reach would-be workers, whereas strong ties – which we measure as 'belonging to the same social group' – are sufficient but not necessary[13]. In the Wahba-Zenou account, knowledge about vacancies is dispersed through weak ties. This sits awkwardly with anthropological and other evidence which suggests that certain social groups, united through strong ties, obtain strongholds and cluster in certain industries, restricting job access to group members (Chandavarkar 1994; Holmstrom 1984, Panini 1996). Such chains are likely to reflect referral-based phenomena. To test whether such chains can be detected, we modify (III) by interacting social-group dummies with low-skill dummies. If we find that network effects are stronger among unskilled workers (which confirms both the

[13] The distinction between weak and strong ties was originally made by Granovetter (1973).

information-dispersion account and the referral-based recruitment account) and vary for unskilled workers across social groups, the more plausible interpretation is that access restrictions apply, thus supporting the referral-hypothesis. Furthermore, we also test directly the theoretical predictions of the effect of city-wide unemployment (as an inverse proxy for the level of economic activity) and the Besley and Burgess (2004) measure of pro-labour legislation on the incidence of referral-based recruitment. Empirical support for these predictions would tip the balance further in favour of the hypothesis of referral-based network effects at the expense of the information-dispersion hypothesis.

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 55th and most recent available "thick" (i.e. comprehensive) survey round.[14] Detailed information has been collected 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 the household as a whole. We briefly describe below the manner in which we reconfigured the NSSO data for our purposes.

Firstly, we decided to limit the dataset to individual working migrants in the seven largest urban agglomerates (UAs): Ahmedabad, Bangalore, Calcutta, Chennai, Delhi, Hyderabad, and Mumbai. The focus is on individuals rather than households because individuals may migrate alone and join a (semi-)permanent household. The

[14] For details on sampling design, selection procedure and other related issues, see GOI (1999).

limitation to currently working migrants follows naturally from our research interest in the processes through which individuals have acquired jobs, and filters out some peer emulation effects, as argued above. We reconfigured the dataset according to the location codes that make up UAs rather than cities narrowly defined, in order to ensure that individuals commuting from outside a city's boundaries are included. We only use the seven largest UAs so that identical industry dummies may be deployed for each (see Section 3 on why industry dummies are required for our purposes): in smaller UAs some industries are absent. The reduction in sample size that results from our limitation to individual working migrants in the seven largest urban agglomerates is as follows. 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 UAs, 12,284 are in addition currently working, and 4,723 are in addition migrants.[15]

Secondly, we construct a dummy variable that takes the value one for each pair of migrants, constructed as per above, from the same source area. The NSSO data provides us with 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). We decided to keep urban-urban migration despite the fact 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 from the analysis. We drop the relatively small categories

[15] We noted that individuals are classified as migrants when they migrated at any point during their life time, even as long as 70 years ago.

from abroad and from the same district (both rural and urban), so that our final dataset comprises 3,988 individuals.

Thirdly, we construct all possible pairs of individuals working in the same industry in the same UA (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 that the data will permit. Because of this high level of disaggregation, not all industries have more than one observation; the total number of pairs that we are able to construct is equal to 2,968.

Fourthly, we differentiate migrants by the skill level of their occupation. The NSSO data provide the occupation of the migrant at the NCO three-digit level. We code the 900 occupations present in the data to three skill levels – low-skilled occupations, medium-skilled occupations, and high-skilled occupations. This allows us to experiment with a different way of disaggregating stocks by the level of skill as opposed to the educational level of the migrant pair.

4.2 Descriptive Statistics

Figure 3 plots the distribution of migrant pairs across the seven cities in our sample. The distribution of pairs is uneven across these cities, with Mumbai comprising over 45 per cent of the total sample, and Delhi only about 1 per cent.[16] The proportion of migrants from the same region in total migrants across the seven cities range from 32 per cent in Bangalore to 80 per cent in Delhi. Table 3 presents the distribution of migrant pairs across the top 25 industries measured by the total number of migrant

[16] This echoes the perception of Mumbai as a city of migrants (Zachariah 1968).

pairs. Industry 45201 (Construction of residential buildings) contains as much as 23 per cent of all migrant pairs. However, migrant pairs are more evenly distributed for the other 116 industries, with Industry 18101 (manufacturing of textile garments and clothing accessories) ranking second with 7.9 per cent of all migrant pairs. Apart from Industry 45201, no other industry dominates with regard to migrant pairs that are either from the same source region or not.

In Tables 4 and 5, we present cross tabulations of migrant pairs by educational level, and social group, separately for migrants who come from the same source region, and for migrants who are not from the same source region. Table 4 presents a cross tabulation of migrant pairs by educational levels. Two findings are note-worthy. Firstly, for migrant pairs from the same source region, it is more likely that they have the same educational level – the diagonal elements of the first half of Table 4 generally dominate the off-diagonal elements. Secondly, for migrants with the same educational level, it is more likely that such migrants are from the same region than from different regions – the diagonal elements in the first half of the table dominate the diagonal elements in the second half of the table, except for the category 'secondary and below'. Table 5 presents cross tabulation of migrant pairs by social group. Among the Hindu castes, we see that both migrants are more likely to be from the same source region if both are SCs or OBCs, or where one is an SC and the other is an OBC. The likelihood of both migrants coming from the same source region is less strong when one or both of the migrant pair is from the non SC/ST/OBC category. This provides indirect evidence that network effects may be stronger among scheduled castes and other backward castes compared to groups more highly ranked in the caste hierarchy. Moreover, the "other" category is relatively large in numbers, and it is of interest to provide a more detailed breakdown of social identity and thus the caste and religious characteristics, of the migrant pairs. Table 6 presents odds ratios of the likelihood of being from the same source region for different combinations of characteristics of the migrant pairs. We see that the odds ratios exceed unity when both migrants are illiterate or educated to primary level, when at least one of the migrant pair is SC or from the non SC/ST/OBC social group and when both are from the same religion (for the three major religious groups considered). Interestingly, when we compute odds ratios for the different combinations of social groups and educational levels, we find that the odds ratio exceeds one for SCs who are illiterate, and for OBCs who are illiterate, and educated to primary and graduate levels. The data thus seems to suggest that migrants are more likely to come from the same source region if they are less educated, and within the less educated, from the backward castes. The odds-ratios also indicate the presence of spillover effects across some social groups, suggesting that Munshi and Rosenzweig (2006) may be wrong to interpret sub-caste as the relevant boundary of labour market networks. According to our data-set, the boundaries of labour market networks among Hindu castes appear more porous than among different religious communities such as Christians and in particular, among Muslims.[17] We will next examine whether these patterns evident in the cross tabulations are supported by more formal econometric analysis.

5. Empirical Results

We model the likelihood of a migrant pair in a given industry and city to have migrated from the same source region. We estimate probit models where the dependent variable is a binary variable which takes the value one when the migrant

[17]On village data from South-India, Iversen (2006) found referrals mainly to cover family and fellow-caste members, but also to extend beyond caste boundaries.

pair is from the same source region, and zero when they are not from the source region. A possible econometric problem with the pair measure is that it may induce spatial correlation between observations involving the same individual, and consequently bias standard errors. Since we construct the 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 the 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 in this section (Wooldridge 2002, pp. 409-410) where we allow for cluster correlations within a industry-city cell.

We begin by estimating flow network effects. The results are presented in Table 7. In the Cols (1), (2) and (3), we estimate flow network effects without industry and city fixed effects, with city effects only, and with both industry and city effects. We introduce three demographic control variables – the product of the age of the two migrants in the pair, and its square, and a dummy variable which takes the value one when both migrants are male (zero otherwise). Our key variables of interest are the flow network variables related to the level of education. As we have argued in Section 2, if referral based networks are the mechanism by which migrants move to cities and find jobs, then the flow network variables that characterise both migrants as being with no or little education will increase the probability of the migrant pair originating from the same location. We find support for this hypothesis in the data. With industry and city fixed effects included, the probability that both migrants are from the same source region increases by 13% if both migrants are illiterate.. This variable remains significant when we estimate the probits without city and industry fixed effects, with city fixed effects only, and with both city and industry fixed effects. When both migrants are educated to the primary level, the probability that the migrant pair is from the same source region increases by a range of 5-7% across the three estimates. However, this variable is not significant when industry fixed effects are included. In contrast, the flow network variables corresponding to both migrants being educated to the secondary level, and to the graduate level are not statistically significant for any of the estimates.

Our aggregate stock network variable is not significant in any of the three estimates. With respect to the control variables, the product of age and its square is significant, and age seems to have a non-linear effect on the likelihood of migration from the same source region, first increasing and then decreasing. The characteristic that both migrants are male increases the probability that they are from the same source region, and the coefficient on this variable is significant at the 10% level when both city and industry fixed effects are included.

The theory of optimal recruitment also predicts that for a higher level of economic activity (and thus lower unemployment) at destination the higher is the likelihood that employee networks will be used in recruiting workers. We measure unemployment for the seven cities using the indicator for structural unemployment from the NSS 55 Round (1999-2000) Employment and Unemployment Data Set.[18] Our theory also predicts that more pro-labour legislation will reduce the loyalty gains from referral-based recruitment. We capture the degree of pro-labour legislation using the Besley-Burgess (2004) measure of labour regulations, which is available for India's major states.[19] We use the Besley-Burgess measure for the states in which

[18] The unemployment rate in India is calculated in three different ways, daily, weekly and yearly. The latter refers to structural unemployment based on what is reported as the principal status of an individual.

[19] The major labour legislation at the central level in India is the Industrial Disputes Act of 1947, which sets out the conciliation, arbitration and adjudication procedures to be followed in the case of an industrial dispute. The Act has been amended by different state governments during the post-independence period. Besley and Burgess code the amendments at the state level to the Act with each amendment coded as being pro-worker, neutral or pro-employer. Each pro-worker amendment is given a value of 1, each neutral amendment a value of 0, and each pro-

Ahmedabad, Bangalore, Chennai, Hyderabad, Kolkata and Mumbai are located, and for Delhi, we use the value for the adjoining state of Haryana.[20] A higher value of this indicator implies labour legislation is more pro-labour. We include the unemployment rate and the Besley-Burgess labour regulation measure in the basic specification as in Col. (3) of Table 7. We first include the unemployment rate in Col. (1) of Table 8.[21] We find that a 10 % rise in unemployment leads to a 4.3% decrease in the likelihood that both migrants are from the same source region, and that the coefficient on the unemployment rate is significant at 5% level. Including the labour legislation variable by itself in Col. (2) of Table 8, we find that a 10% increase in the pro-labour legislation measure leads to a fall in the likelihood that both migrants are from the same source region by 0.35%, and that the coefficient on labour legislation variable is significant at the 10% level. When including both variables at the same time, the labour legislation variable becomes statistically insignificant. This is not surprising as the two variables are expected to be correlated – a higher degree of pro-worker labour legislation would be expected to increase unemployment (Besley and Burgess 2004). These results provide further support for our theory of optimal recruitment.

We next present four different robustness tests of our basic results on flow network effects. Firstly, to control for the effects of source region clustering on destination region clustering, we introduce separate dummies for the twenty four different states from which migrants originate in our sample. This we do in Col. (1) of Table 9. In Col (2), we only include individuals who have migrated from a different

---

employer amendment a value of -1. These scores are then cumulated over time. We use the cumulative scores in 1999, the year of the employment survey used in the paper.

[20] 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 the state of Haryana.

[21] We omit city fixed effects in all specifications in Table 6 as they are expected to be highly collinear with both the unemployment rate and the measure of labour flexibility as the latter two variables vary only across cities, and not within cities.

state than the one they are currently employed in. This we do 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 explained by an omitted variable – the distance that the migrant pair had to travel 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 more migrants from the same (nearby) source area, especially when the source area has a concentration of such less-educated individuals. By confining our sample to individuals who have migrated from one state to another (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.[22] In Col (3), we omit migrant pairs working in the city of Mumbai, which we have seen in the previous section, comprises 40 per cent of our sample. Finally, in Col (4), we omit industry 45201, where a quarter of our migrant pairs work. In all these specifications, we include industry and city fixed effects and the same control variables as in the previous specification presented in Table 7.

The finding that the probability of the migrant pair being from the same source increases when the paired migrants are both uneducated remains remarkably robust to the differences in the samples of migrant pairs that we use in the estimates presented in Table 9. This flow network variable is significant at the 1 per cent level whether state source dummies are used, at the 5 per cent level when the city of Mumbai is excluded and at the 10 per cent level when industry 45201 is omitted. Both migrants

[22] It should also be pointed out that by construction of our dependent variable where we exclude individuals who have migrated from the same district, we reduce 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 in the paper works to our advantage 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.

being uneducated increases the probability that the migrant pair are from the same source region by 15-30%.

When we confine our sample to migrants from another state, we find that our results are actually stronger than in the comparable baseline specification (3) in Table 7. That is, the probability that the migrant pair come from the same source region increases by 30% when the migrants are both uneducated, and the coefficient is significant at the 5 per cent level. Also, the coefficient on the migrant pair being both educated to primary level is now statistically significant at the 10 per cent level. The results are strongly supportive of our maintained hypothesis that referral based network effects dominate other factors in explaining the clustering of less educated migrant pairs from same source regions in destination industries and cities.

We now turn to the estimation of stock network effects. We first disaggregate our stock variable by level of literacy of migrants pairs included in this variable. Thus, we have stocks of illiterate workers (stock_illiterate), stocks of workers educated to primary level (stock_low literacy), stocks of workers educated to secondary level (stock – medium literacy) and stocks of workers educated to graduate level (stock – high literacy). We augment our basic specification (as in Col. (3) of Table 7) with these four stocks. The results are presented in Table 10, Col. (1). As predicted by a referral based network perspective on migration behaviour, we find that the coefficient on stock-illiterate is positive and significant at the 5 per cent level. An increase in the stock of illiterate workers from the same source region by 1% leads to a 0.15% increase in the probability that the migrant pair is from the same source region. Interestingly, the effect of the stock – medium literacy is negative on the probability that the migrant pair is from the same source region. This may suggest a congestion effect of the increase in the stock of medium literate workers from the same source region. The other stock variable is statistically insignificant. The flow network variables for both migrants being uneducated remain statistically significant.

When we disaggregate the stock by the skill of occupation of the migrant pair, so that we have three stocks – stock_low skill, stock_medium skill and stock_high skill, we find that our estimates of stock network effects in Col (2) are similar to our estimates in Col (1). The stock of low skilled workers from the same source region has a positive and significant (at the 10% level) effect on the probability that the current migrant pair is from the same source region. Neither the stock of medium-skilled workers nor the stock of high-skilled workers has a significant effect on current migration from the same source region. Thus, these results establish the presence of network effects, and in particular, those predicted by our referral-based-recruitment perspective.

Finally, we try to capture the effects of social identities on the probability of migration from the same source region. We examine in particular the influence of caste and religion on migration behaviour. We argued in Section 3 that if caste or religion mediates network effects for low-skilled workers for whom the effects of social identity are predicted to be more pronounced because of the risks associated with providing referrals, our perspective on networks will be more plausible than one based on information-dispersion. To distinguish referral based effects from information dispersion effects, we disaggregate the two network variables for low skilled workers that matter in our theory and were found to be significant in previous estimates – the variables corresponding to when both migrants are illiterate and when both migrants are educated to primary level. These two variables are disaggregated by caste and major religion – within Hindus, we have scheduled castes (SC), other backward castes (OBC) and non SC non OBCs (corresponding to the upper castes).[23] The other religious affiliation we consider is Muslim, as there are few observations for other religions such as Christianity, Buddhism and Sikhism. The interaction between four social groups and two education variables produces eight variables to include in the empirical specification. We estimate flow network effects as in Col (3) of Table 7, with the disaggregated variables corresponding to the two levels of education – illiterate and educated to primary variable. Since we have few observations for some of the disaggregated variables, they drop out of the final estimates which feature in Table 11. The interaction term between both individuals being Muslim and when both migrants are illiterate is positive and statistically significant. Similarly, the coefficient on the upper Hindu caste and when both migrants are illiterate is positive and statistically significant. All other interaction terms are insignificant. This suggests that individuals from upper Hindu caste backgrounds and Muslims are more likely to be chain migrants and access referral based networks of their own social groups than individuals from SC and OBC backgrounds. This supports the anthropological literature suggesting that certain social groups regulate access to work, thereby favouring others from the same caste or religious background while excluding others. Such phenomena can best be accommodated by a referral-based perspective on networks.

6. Conclusion

This paper contributes to the empirical literature that examines the role of social networks in determining migration behaviour and labour market outcomes. In contrast to previous studies, the paper pays particular attention to the demand-side of network

[23] We drop the scheduled tribes from our sample as there are few observations for this group.

use. It sketches a theory of optimal recruitment applicable to traditional economies where employers use their employees to recruit individuals into low-skilled jobs. We test the negative relationship between network effects and skill-intensity of the migrant that our theory predicts. We use a nationally representative large data-set of migrants to the largest cities in India for this purpose and find strong support for our main hypothesis. In addition, our results for pro-labour legislation and economic activity concur with theoretical predictions and provide further support for this demand-side story. Additional support is provided by the evidence suggesting variation in the strength of network effects across social groups. The results are robust to alternate specifications and changes in sample size, and are more strongly suggestive of a referral-based perspective than an information-dispersion one. Finally, the descriptive statistics provided in table 6 suggest that the boundaries of labour market networks transcend social category among Hindu castes, but that the spillover effects across religious groups are more limited. Future research should aim to shed more light on the all-India picture regarding the social location of these boundaries.

References

Antoninis, Manos (2006), "The Wage Effects from the Use of Personal Contacts as Hiring Channels", *Journal of Economic Behaviour and Organisation*, 59 (1): 133-146.

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

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa (2004), "Place of Work and Place of Residence: Informal Hiring Networks and Labour Market Outcomes", mimeo.

Bertrand, Marianne, Erzo Luttmer and Sendhil Mullainathan (2000), "Network Effects and Welfare Cultures", *Quarterly Journal of Economics*, 115(3): 1019-1055.

Besley, T. and R. Burgess (2004), "Can Labour Regulations Hinder Economic Performance? Evidence from India?", *Quarterly Journal of Economics*, 19(1): 91-134.

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

Gore, M. S. (1970): Immigrants and neighbourhoods: Two aspects of life in a metropolitan city. Tata Institute of Social Sciences, Bombay.

Government of India (1999), NSSO, 55th Round: Household Schedule 10: Employment and Unemployment, SDRD.

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

Granovetter, Mark (1973), "The Strength of Weak Ties", *American Journal of Sociology*, 78: 1360-1380.

Granovetter, Mark (2005), "The Impact of Social Structure on Economic Outcomes", *Journal of Economic* Perspectives, 19 (1): 33-50.

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

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

Ioannides, Yannis M. and Linda D. Loury (2004), "Job Information Networks, Neighbourhood Effects and Inequality", *Journal of Economic Literature*, 42(4): 1056-1093.

Iversen , V (2006): Segmentation, spillovers and network multipliers: A theory of rural urban migration for a traditional economy, mimeo, School of Development Studies, University of East Anglia.

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

Massey, Douglas S. (1988), "Economic Development and International Migration in Comparative Perspective", *Population and Development Review*, 14 (3): 383-413.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60 (3): 531-542.

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

Marmaros, David and Bruce Sacerdote (2002), "Peer and social networks in job search", *European Economic Review*, 46 (4-5): 870-879.

Marsden, P. and E. Gorman (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): 477-96

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

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

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

Myers, C.A. (1958): Labour Problems in the Industrialization of India, Cambridge, Massachusetts.

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.

Panini, M. N. (1996): The Political Economy of Caste, in M.N. Srinivas (ed.): Caste: Its Twentieth Century Avatar, *Viking Publishing House*, New Delhi.

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

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

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

Zachariah, K. C. (1968): Migrants in Greater Bombay, *Asia Publishing House*, Bombay.

Table 3. Distribution of Migrant Pairs by Industries

Industry Code	Not from Same Region	Per cent of all Migrant Pairs not from same region	Per cent of all migrant pairs in same industry	From Same Region	Per cent of all Migrant Pairs from same region	Per cent of all migrant pairs in same industry	Total Migrant Pairs	Per cent of all Migrant Pairs
45201	165	12.5	24.3	514	31.3	75.7	679	22.9
18101	148	11.2	63.5	85	5.2	36.5	233	7.9
95009	110	8.3	52.4	100	7.9	47.6	210	7.1
75220	64	4.8	33.7	126	7.7	66.3	190	6.4
60221	107	8.1	59.1	74	4.5	40.9	181	6.1
52201	106	8.0	59.2	73	4.4	40.8	179	6.0
55202	100	7.5	64.5	55	3.3	35.5	155	5.2
36911	77	5.8	55.4	62	3.8	44.6	139	4.7
72200	69	5.2	60.0	46	2.8	40.0	115	3.9
45402	22	1.7	32.8	45	2.7	67.2	67	2.3
55101	24	1.8	51.1	23	1.4	48.9	47	1.6
75232	25	1.9	54.3	21	1.3	45.7	46	1.5
55209	11	0.8	26.8	30	1.8	73.2	41	1.4
93090	13	1.0	31.7	28	1.7	68.3	41	1.4
65191	22	1.7	62.9	13	0.8	37.1	35	1.2
60231	23	1.7	69.7	10	0.6	30.3	33	1.1
52110	14	1.1	45.2	17	1.0	54.8	31	1.0
52202	19	1.4	63.3	11	0.7	36.7	30	1.0
85110	8	0.6	29.6	19	1.2	70.4	27	0.9
95001	19	1.4	70.4	8	0.5	29.6	27	0.9
36912	16	1.2	66.7	8	0.5	33.3	24	0.8
15549	6	0.5	28.6	15	0.9	71.4	21	0.7
17292	0	0.0	0.0	21	1.3	100.0	21	0.7
60100	15	1.1	71.4	6	0.4	28.6	21	0.7
36101	5	0.4	31.3	11	0.7	68.8	16	0.5
Total Migrant Pairs in Full Sample	1325	--		1643	--		2968	--

Note: Top 25 Industries in terms of total migrant pairs. There are 117 industries in total.

Table 4. Cross Tabulation of Migrant Pairs by Educational Level

If Migrant Pair is from the Same Source Region						
Second Migrant	First Migrant					
	Educational Level	Not Literate	Primary and Below	Secondary and Below	Graduate and Above	Total
	Not Literate	166	169	96	2	433
	Primary and Below	135	212	173	6	526
	Secondary and Below	49	123	313	45	530
	Graduate and Above	2	5	62	84	153
	Total	352	509	644	137	1,642
If Migrant Pair is not from the Same Source Region						
Second Migrant	First Migrant					
	Educational Level	Not Literate	Primary and Below	Secondary and Below	Graduate and Above	Total
	Not Literate	43	53	74	8	178
	Primary and Below	65	125	114	11	315
	Secondary and Below	61	188	355	55	659
	Graduate and Above	4	16	61	76	157
	Total	173	382	604	150	1,309

Table 5. Cross Tabulation of Migrant Pairs by Social Group

If Migrant Pair is from the Same Source Region						
Second Migrant	First Migrant					
	Social Group	Scheduled Tribe	Scheduled Caste	Other Backward Caste	Others	Total
	Scheduled Tribe	2	2	7	9	20
	Scheduled Caste	0	55	124	53	232
	Other Backward Caste	6	72	371	92	541
	Others	11	37	147	655	850
	Total	19	166	649	809	1,643
If Migrant Pair is not from the Same Source Region						
Second Migrant	First Migrant					
	Social Group	Scheduled Tribe	Scheduled Caste	Other Backward Caste	Others	Total
	Scheduled Tribe	0	1	3	15	19
	Scheduled Caste	1	17	30	71	119
	Other Backward Caste	4	37	146	133	320
	Others	12	66	183	606	867
	Total	17	121	362	825	1,325

Table 6. Odds Ratios of the likelihood being from the same source region

Characteristics of Migrant Pair	Odds Ratio
Both Illiterate	**3.08**
Illiterate and educated to primary level	**2.1**
Illiterate and educated to secondary level	0.84
Illiterate and educated to graduate level	(0.3) n=16
Both educated to primary level	**1.35**
Educated to primary level and educated to secondary level	0.86
Educated to primary level and educated to graduate level	(0.34) n=38
Both educated to secondary level	0.7
Educated to secondary level and educated to graduate level	(0.73) n= 9
Both educated to graduate level	(0.88) n=42
SC-SC	**(2.61)** n=72
SC-OBC	**2.07**
SC-Others	**1.32**
OBC-OBC	0.81
OBC-Others	**3.41**
Others-Others	**1.08**
Hindu-Hindu	**1.05**
Hindu-Muslim	0.55
Hindu-Christian	0.16
Muslim-Muslim	**3.49**
Muslim-Christian	(0.03) n=11
Christian-Christian	**(2.42)** n=12
SC-Illiterate	**1.61**
SC-Primary	0.81
SC-Secondary	0.61
SC-Graduate	(0.81) n=26
OBC-Illiterate	**2.90**
OBC-Primary	**1.61**
OBC-Secondary	0.94
OBC-Graduate	**(1.77)** n=71
Others-Illiterate	0.84
Others-Primary	0.72
Others-Secondary	0.86
Others-Graduate	0.70

Note: Odds Ratio is calculated as ratio of the proportion of migrant pairs with a particular set of characteristics in total migrant pairs from the same source region to the proportion of migrant pairs with the same set of characteristics in total migrant pairs not from the same source region. Odds ratios greater than one are highlighted in bold, and those based on a number of pairs with particular characteristics that is smaller than 100 are in parentheses.

Table 7. Flow Network effects – Basic Specification

	(1)	(2)	(3)
Age	-0.0002 (0.0002)	-0.0003 (0.0002)	-0.0005 (0.00009)***
Ageprodsq	0.0000001 (0.00000004)	0.0000001 (0.00000004)	0.0000002 (0.0000004)***
Maleprod	0.0157 (0.0433)	0.0243 (0.0435)	0.0519 (0.0433)*
Stock	0.0006 (0.0005)	0.0003 (0.0004)	0.0002 (0.0005)
Both not literate	0.2299 (0.0658)***	0.1819 (0.0635)***	0.1257 (0.0682)*
Both educated to primary level	0.0772 (0.0405)*	0.0982 (0.0343)***	0.0551 (0.0500)
Both educated to secondary level	-0.0587 (0.0597)	-0.0262 (0.0512)	0.0002 (0.0371)
Both educated to graduate level	0.0621 (0.0939)	-0.0548 (0.1173)	0.0845 (0.0834)
City fixed effects?	No	Yes	Yes
Industry fixed effects?	No	No	Yes
Pseudo-R square	0.0357	0.0636	0.1386
Number of Observations	2968	2968	2840

Note: a) *, ** and *** significant at the 10, 5 and 1 percent level respectively; b) Cluster-adjusted heteroskedasticity consistent standard errors in parenthesis.
Cols (1), (2) and (3) are with no industry and city fixed effects, with only industry fixed effects, and with both industry and city fixed effects.

Table 8. Flow Network effects – The Effects of Unemployment and Labour Regulations

	(1)	(2)	(3)
Age	-0.0005 (0.00001)***	-0.0005 (0.00001)***	-0.0005 (0.00001)***
Ageprodsq	0.0000001 (0.00000001)***	0.0000001 (0.00000001)***	0.0000001 (0.00000001)***
Maleprod	0.0684 (0.0418)*	0.0633 (0.0418)	0.0689 (0.0418)*
Stock	0.0013 (0.0006)*	0.0010 (0.0006)*	0.0013 (0.0006)*
Both not literate	0.1643 (0.0798)**	0.1659 (0.0793)**	0.1645 (0.0797)*
Both educated to primary level	0.0384 (0.0545)	0.0405 (0.0538)	0.0393 (0.0543)
Both educated to secondary level	-0.0245 (0.0371)	-0.0274 (0.0381)	-0.0256 (0.0372)
Both educated to graduate level	0.1118 (0.0725)	0.1132 (0.0741)	0.1114 (0.0729)
Unemployment	-0.434 (0.0209)**	---	-0.0347 (0.0154)**
Degree of Labour Regulations	---	-0.0354 (0.0201)*	-0.0094 (0.0141)
City fixed effects?	No	No	No
Industry fixed effects?	Yes	Yes	Yes
Pseudo-R square	0.1221	0.1209	0.1222
Number of Observations	2840	2840	2840

Note: a) *, ** and *** significant at the 10, 5 and 1 percent level respectively; b) heteroskedasticity consistent standard errors in parenthesis.

Table 9. Flow Network Effects - Robustness Tests

	(1)	(2)	(3)	(4)
Product of Ages of Migrant Pair	-0.0005 (0.0001)***	-0.0004 (0.0003)	-0.0004 (0.0002)**	-0.0004 (0.0001)**
Square of Product of Ages of Migrant Pair	0.0000002 (0.00000004)***	0.0000001 (0.0000001)	0.0000001 (0.00000005)**	0.00000001 (0.0000003)**
Both are male	0.0361 (0.0398)	0.3050 (0.1297)**	-0.0013 (0.0402)	0.0398 (0.0633)
Stock	0.0004 (0.0005)	0.0008 (0.0010)	0.0003 (0.0005)	-0.0012 (0.0005)**
Both not literate	0.1814 (0.0543)***	0.2988 (0.1311)**	0.1514 (0.0839)*	0.1823 (0.1001)*
Both educated to primary level	0.0548 (0.0452)	0.0974 (0.0517)*	0.0209 (0.0573)	0.0711 (0.0601)
Both educated to secondary level	-0.0015 (0.0399)	0.0593 (0.0501)	0.0623 (0.0828)	0.0015 (0.0376)
Both educated to graduate level	0.0562 (0.0873)	0.1910 (0.1606)	0.0469 (0.0862)	0.1257 (0.0837)
City fixed effects?	Yes	Yes	Yes	Yes
Industry fixed effects?	Yes	Yes	Yes	Yes
Pseudo-R square	0.1845	0.1618	0.1865	0.0767
Number of Observations	2833	1176	1541	2161

Note: a) *, ** and *** significant at the 10, 5 and 1 percent level respectively; b) heteroskedasticity consistent standard errors in parenthesis.
Col (1): Dummies for source regions included.
Col (2): Restricted to migrants who are from a different state than where they are working
Col (3): City of Mumbai excluded.
Col (4): Industry 45201 (Construction of residential buildings) excluded.

Table 10. Stock Network Effects

	(1)	(2)
Product of Ages of Migrant Pair	-0.0004 (0.0001)***	-0.0002 (0.0002)
Square of Product of Ages of Migrant Pair	0.0000002 (0.00000001)***	0.00000008 (0.0000004)
Both are male	0.0671 (0.0431)*	0.0334 (0.0472)
Stock-illiterate	0.0015 (0.0007)**	---
Stock-low literacy	-0.0123 (0.0040)***	---
Stock-high literacy	-0.0007 (0.0015)	---
Stock-low skilled	---	0.0009 (0.0004)*
Stock-medium skilled	---	-0.0002 (0.0002)
Stock-high skilled	---	-0.0003 (0.0003)
Both not literate	0.1114 (0.0644)*	0.1990 (0.0673)***
Both educated to primary level	0.0682 (0.0497)	0.0665 (0.0420)
Both educated to secondary level	-0.0053 (0.0368)	-0.0365 (0.0564)
Both educated to graduate level	0.1041 (0.0825)	0.0655 (0.0941)
City fixed effects?	Yes	Yes
Industry fixed effects?	Yes	Yes
Pseudo-R square	0.1456	0.0514
Number of Observations	2840	2968

Note: a) *, ** and *** significant at the 10, 5 and 1 percent level respectively; b) Heteroskedasticity consistent standard errors in parentheses.

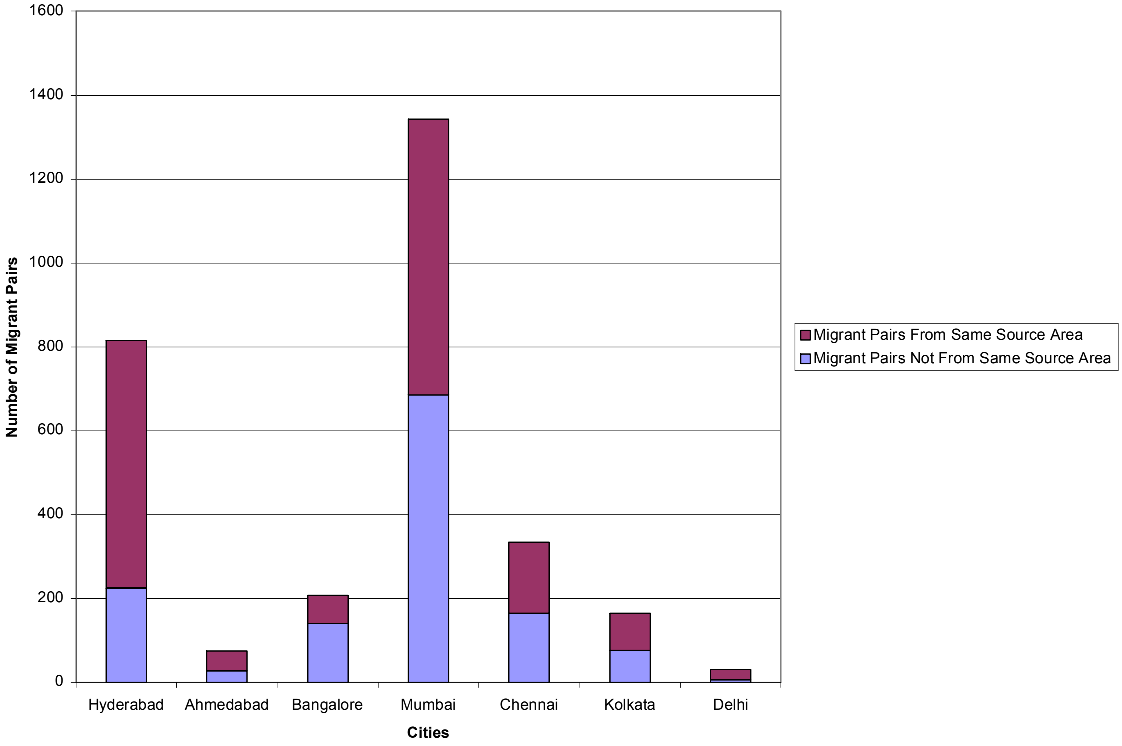
Col (1): Stock split into stocks of illiterate workers, stocks of workers educated to primary level, and stocks of workers educated to secondary level and beyond.
Col (2): Occupations split into three groups: less skilled, medium skilled and high skilled. Interaction variables constructed which are the three occupational skill dummies interacted with the stock variable.

Table 11. Social structure and Network Effects

	(1)
Product of Ages of Migrant Pair	-0.0005 (0.0001)***
Square of Product of Ages of Migrant Pair	0.0000002 (0.00000004)***
Both are male	0.0473 (0.0436)
Stock	0.0002 (0.0004)
Both educated to secondary level	0.0111 (0.0368)
Both educated to graduate level	0.0818 (0.0837)
Both SC and illiterate	-0.0532 (0.1821)
Both SC and educated to primary level	-0.0295 (0.3262)
Both OBC and illiterate	0.2760 (0.1534)
Both OBC and educated to primary level	0.0803 (0.0965)
Both non SC/non OBC Hindu and illiterate	0.3412 (0.0966)**
Both non SC/non OBC Hindu and educated to primary level	-0.0533 (0.0833)
Both Muslim and educated to primary level	0.3482 (0.0759)**
City Fixed Effects?	Yes
Industry fixed effects?	Yes
Pseudo-R square	0.1440
Number of Observations	2839

Note: a) *, ** and *** significant at the 10, 5 and 1 percent level respectively; b) Heteroskedasticity consistent standard errors in parentheses.

Figure 3. Distribution of Migrant Pairs by Cities



Appendix 1. Skill Classification of Occupations in India using NCO Occupation Codes

High Skilled Workers

Division 0-1 Professional, Technical and Related Workers
Division 2 Administrative, Executive and Managerial Workers
Division 3 Clerical and Related Workers

Medium Skilled Workers

Division 4 Sales Workers
Division 71 Miners, Quarrymen, Well Drillers and Related Workers
Division 72 Metal Processors
Division 74 Chemical Processors and Related Workers
Division 83 Blacksmiths, Tool Makers and Machine Tool Operators
Division 84 Machinery Fitters, Machine Assemblers and Precision Instrument Makers
Group 85 Electrical Fitters and Related Electrical and Electronic Workers
Division 88 Jewellery and Precious Metal Workers and Metal Engravers
Division 90 Rubbers and Plastic Product Makers
Division 91 Paper and Paper Board Products Makers
Division 92 Printing and Related Workers
Division 96 Stationary Engines and Related Equipment Operators
Division 97 Material Handling and Related Equipment Operators
Division 98 Transport Equipment Operators

Low Skilled Workers
Division 5 Service Workers
Division 73 Wood Preparation Workers and Paper Makers
Division 75 Spinners, Weavers, Knitters, Dyers and Related Workers
Division 76 Tanners, Fellmongers and Felt Dressers
Division 78 Tobacco Preparers and Tobacco Product Makers
Division 79 Tailors, Dress Makers, Sewers, Upholsterers and Related Workers
Division 80 Shoemakers and Leather Goods Makers
Division 81 Carpenters, Cabinet and Related Wood Workers
Division 82 Stone Cutters and Carvers
Division 87 Plumbers, Welders, Sheet Metal and Structural Metal Preparers and Erectors
Division 89 Glass Formers, Potters and Related Workers
Division 93 Painters
Division 94 Production and Related Workers
Division 93 Bricklayers and other Construction Workers
Division 99 Labourers not elsewhere classified