Homesick or Homerun? Distance from Hometown and Employee Productivity – A Natural Experiment from India

Prithwiraj Choudhury[1] and Ohchan Kwon

Harvard Business School

Does distance of workplace from hometown help or hurt individual performance? In the short-term, distance from hometown could enhance individual performance by increasing the utility of working time vis-à-vis the utility of consumption time. However, in the longer-term, distance from hometown could hurt individual performance by triggering the psychic costs of being away from family and friends. Measuring the effect of distance from hometown on individual performance is beset with endogeneity and selection concerns. We exploit the randomized assignment of entry-level employees to eight production centers at an Indian technology firm to circumvent such concerns. Our results suggest that distance from hometown has a positive effect on individual performance in the short-term and a negative effect in the longer-term. We also identify variation in the longer-term effect based on the ability and social embeddedness of the individual.

Keywords – Distance from hometown, psychic costs, geographic preference, hiring, individual performance, employee productivity, family and friends, natural experiment

[1] Corresponding author. Email – pchoudhury@hbs.edu. We would like to thank Lauren Cohen, Shane Greenstein, Elena Kulchina, and the reviewers, discussants, and participants at seminars at DRUID, the Harvard Business School, and the Strategic Management Society for comments on a previous draft.

1. Introduction

There is a nascent literature that looks at the revealed preferences of scientists, engineers, entrepreneurs, and CEOs to choose work that is close to home (Dahl and Sorenson, 2010a, 2010b, 2012, Kulchina 2016, Yonker 2017). Dahl and Sorenson (2012) define "home" (for entrepreneurs) as "regions in which they have deep roots, the places where they have family and friends, their "home" regions" (Dahl and Sorenson, 2012, page 1059) and we adopt this definition. Using panel data on the Danish population, Dahl and Sorenson (2010a) estimate a strong revealed preference of scientists and engineers to live close to family and friends.[2] The same researchers (2012) find that entrepreneurs tend to locate in home regions and that ventures survive longer and generate greater cash flows when located in the home region of an entrepreneur. Kulchina (2016) finds evidence consistent with entrepreneurs' preference to live in locations that are personally attractive to them, even at the cost of firm profit. In a recent paper, Yonker (2017) finds evidence supporting the revealed preference of CEOs to find work close to home.[3]

Yet on the other hand, the literature in organizations and innovation has long established that firms benefit from hiring geographically distant individuals. This benefit accrues from mechanisms such as learning-by-hiring, knowledge flows, and the transfer of social capital embedded in mobile individuals (Rosenkopf and Almeida, 2003, Song et al., 2003, Singh and Agrawal, 2011, Dokko and Rosenkopf, 2010). As Rosenkopf and Almeida (2003) argue, in their search for new knowledge, firms are limited by their geographic context. Given this, hiring geographically distant individuals can create "bridges to distant contexts." Song et al. (2003) show that external hiring can be used to extend the geographical boundaries of interfirm knowledge transfer and find evidence that both domestic and international hiring are similarly conducive to learning-by-hiring.[4] Dokko and Rosenkopf (2010) demonstrate that mobile

[2] The authors find that the Danish technical workers who were surveyed value (in order from most to least important): (i) proximity to current homes; (ii) proximity to parents, (iii) proximity to high-school classmates, and (iv) proximity to college classmates.
[3] Relatedly, Kalnins and Lafontaine (2013) find that distance creates monitoring problems for distant owners.
[4] Singh and Agrawal (2011) find that hiring firms increase their use of new recruits' prior inventions by 219% on average compared to the time period when the new hires were not employees at the firm. These findings are situated in the rich prior literature on how inventor/employee mobility relates to knowledge transfer (Gilfillan 1935, Arrow,

individuals carry social capital, affecting the outcomes of the firms they join by altering the patterns of interactions between firms. In addition, the human capital literature has shown that being hired by a geographically distant firm can be beneficial for individuals, if such mobility enables them to receive training, leverage their skills, and develop their own human capital (Bidwell and Briscoe, 2010).

Given these two strands of literature, an important question relates to understanding how distance from hometown, i.e., distance from family and friends, affects the performance of individuals who work far away. However, we have very little empirical evidence of any productivity differences among knowledge workers when they are assigned to work either close or far away from home. We also do not know the mechanisms that might underlie such a relationship. In this paper, we attempt to address this gap in the literature and study the effect of distance of the workplace from an individual's hometown (hereafter called 'distance from home') on individual performance. Specifically, we study the effect of such distance over the short-term (one year after assignment), the longer-term (three years after assignment), the underlying mechanisms, and variation in the longer-term effect based on employee characteristics.

It is possible that distance from home has opposite effects on individual performance in the short and the longer-term. In the short-term, it is plausible that working far from home has a *positive* effect on individual performance, perhaps increasing the productivity and utility of 'working time' vis-à-vis the productivity and utility of 'consumption time' (Becker 1965, Gronau, 1976). However, it is also plausible that in the longer-term, working far away from home has a *negative* effect on individual performance. Here the underlying mechanism would relate to the "psychic costs" of being away from family and friends, a theoretical construct introduced by economists such as Sjaastad (1962) and Schwartz (1973). In the sociology literature, a similar construct. i.e., "social attachment to place" and the dissatisfaction of being away from family and friends has been discussed by Dahl and Sorenson (2010b, 2012).

1962). The contextual limitations of firms related to geography alluded to by Rosenkopf and Almeida (2003) is driven by the underlying geographic localization of knowledge (Jaffe et al., 1993)

There are several empirical challenges in estimating the effect of distance from home on individual performance. Firstly, in a conventional setting, workers might have strong preferences to choose a production center close by. It is also likely that firms hire locally and/or assign workers to production centers closer to their homes. In other words, there could be endogeneity in how distance from home for an individual knowledge worker is determined. Secondly, even if workers are employed at production centers far from home, there could be self-selection in how far away from home they choose to work based on individual characteristics that the econometrician is unable to observe. Finally, there could be informational constraints on workers choosing employment opportunities far away from home. This relates to the 'information hypothesis' proposed by Schwartz (1973) and reiterated by Dahl and Sorenson (2010), i.e., the fact that individuals are less likely to have information on job opportunities far away from home and are thus less likely to move to such jobs.

In this paper, we overcome such challenges by using a hand-collected personnel dataset from an Indian IT firm (hereafter TECHCO) to examine how distance from home affects individual performance. Importantly, we also exploit an employee-allocation protocol unique to our setting that helps us control for the endogeneity concerns summarized above. TECHCO recruits fresh college graduates from across the country, and most importantly for our empirical analysis, *randomly* assigns them to its production centers across eight different locations in India, without consideration to employee-level characteristics, including grades, performance during training, or distance from hometown. Though the motivation for such random assignment will be discussed in detail below, this assignment protocol helps us address the endogeneity concerns related to selection and/or the informational bias related to distant job opportunities and enables us to estimate a causal relationship between distance from home and individual performance.

We exploit this random assignment and estimate various fixed-effects models to understand how distance from home affects employee performance in the short and longer-term. Our focus is 443 newly hired college graduates recruited by TECHCO in 2007. The personnel dataset contains rich employee-level description, including demographic information, various proxies for ability measured during recruiting and training, and performance ratings measured one and three years after initial job assignment.

Importantly for our purpose, it also records an employee's hometown location and the location of the production center to which he or she is assigned, from which we are able to measure shortest travel time from workplace to hometown via train. Our field interviews indicate that almost all newly hired college graduates use trains to travel back to their hometown. We code the travel time back home for each employee based on hand-collected data from the Indian Railways timetable and a computation of shortest direct or indirect train path from their production center to their hometown. We then relate an employee's short- and longer-term performance to his or her travel time from workplace to home.

Interestingly, we find contrasting effects of distance from home on short- and longer-term individual performance. Our findings suggest that travel time positively affects short-term individual performance, that is, the farther away an employee is from his or her hometown, the more likely that his or her first-year performance rating is higher. However, the relationship reverses in the longer-term: employees with longer travel time tend to receive lower performance ratings three years after assignment. Further analysis leads us to believe that the longer-term negative influence is driven by the effect of travel time on on-the-job learning effectiveness. In the longer-term, employees who are far away from home are much less likely to improve their performance relative to peers. We control for several alternate explanations, including attrition and burnout. In secondary analysis, available with authors, we also run a non-linear (spline) regression and find that negative relationship between travel time and longer-term performance is particularly salient for employees who need to travel longer than 23 hours to visit their hometown.

We also leverage the human capital literature that theorizes how high-performing employees are motivated by challenge in the workplace (Taylor and Spence 1952; Farber and Spence 1953, Vroom, 1964, Katzell and Thompson, 1990, Chattopadhay and Choudhury, 2017) and identify variation in the longer-term effect based on whether or not the individual is a high- or low-ability employee. We additionally leverage the literature on social embeddedness (Granovetter 1985, Mitchell et al., 2001, Lee et al., 2004) and homophily (Marsden, 1987; Marsden, 1988, McPherson et al. 2001, Ruef et al., 2003, Vissa, 2011) to identify variation in the longer-term effect based on whether or not the individual is part of the majority language group in her cohort.

Our findings make a contribution to several literatures, including the literature on the geography of work, hiring, and migration, and have managerial implications for hiring managers and for individuals managing their careers.

2. Theory and Hypotheses

The theory literature on employee hiring is based on matching models, matching workers to jobs (Schein 1978, Heckman and Sedlacek 1985, Hall 1986). The rewards offered by the job, such as wages or personal happiness, might provide a good match for the preferences of the worker, leading to greater "horizontal fit" between worker preferences and traits of the job (Bidwell and Mollick, 2015). As Bidwell and Briscoe (2010) suggest, a job that offers greater flexibility, more autonomy, and better work-life balance might be a superior match with worker preferences and might lead to superior individual performance. In the following sections, we develop our theoretical arguments for how distance from home might lead not only to better or worse matches between the worker and the job, but also variation in individual performance.

2.1. Short-Term Effects of Distance from Home on Individual Performance

We build on the theory of allocation of time (Becker, 1965; Gronau, 1976) to theorize about the effect of distance from home on individual performance in the short-term. Becker (1965) provided a theory of allocation of time and defined the two elements of an individual's time: time spent at work (T_w) and time spent at consumption (T_c), where total time available to an individual equaled $T_w + T_c$. Modeling the allocation of time across work and consumption as a theory of choice, Becker argues that an individual's utility function is a combination of market goods and time. Each individual is also modeled as having a goods constraint that is a function of working time, T_w.

Becker (1965) additionally introduces the concepts ‘productivity of working time’ and ‘productivity of consumption time.’ The former is enhanced by market goods and technology related to the workplace, and the latter is affected by market goods and technology related to consumption. He also sets up the underpinnings of a substitution effect between working time and consumption time (‘leisure’). To quote the author, (individuals might) “forfeit money income in order to obtain additional utility, i.e. they exchange money income for a greater amount of *psychic income (italics added by authors)*. For example, they might increase their leisure time…..”(Becker 1965, page 498). The core idea here is that individuals might choose to *reallocate time* from home to work or vice versa, based on utility maximization. Similar to Becker (1965), Gronau (1976) states that the individual consumer maximizes utility according to time and budget constraints, where utility is a function of commodities that are “produced” using market goods and time. The author also extends Becker (1965) and theorizes about how productivity of time spent at work and productivity of time spent at home affect reallocation of time between the two. Gelber and Mitchell (2012) study the reallocation of time from home to work and vice versa based on the level of income taxes. Graff Zivin and Neidell (2014) study the reallocation of time from work to home based on exogenous change in climate and temperatures.

Extending Becker’s theory of allocation of time to our context, it could be argued that in the short-term, the employee who is moved far away from home faces incentives to allocate more time to work than to leisure. An employee working far away from family and friends has fewer consumption options in an unfamiliar new location. Thus, in the short-term, employees working far away from home might allocate disproportionate time to work relative to consumption, and this might positively affect their performance. On the other hand, employees working close to home have a greater number of consumption options that generate “psychic income,” such as spending time with family and friends. They might allocate more time to consumption and leisure than would distant employees. In summary, differences in allocation of time to work versus leisure can differentially affect employee performance in the short-term. We hypothesize:

Hypothesis 1: In the short-term, employee performance will be positively correlated to distance from home.

2.2. Longer-Term Effects of Distance from Home on Individual Performance

Scholars in the migration and productivity literature have long theorized about the effects of distance from home on individuals, and the construct of distance from home appeared prominently in the early gravity models in the migration literature (Carrothers, 1956; Isard, 1960; and Olsson, 1965).[5] As Greenwood (1969) summarizes, distance of the destination location from home acts as the single most important deterrent to migration. He also stated that migration decreases substantially with increased distance for two reasons: (i) distance serves as a proxy for transportation costs and psychic costs; and (ii) distance restricts flow of information related to opportunities available in the host region.

The construct *psychic costs of migration* dates back to two seminal studies – Sjaastad (1962) and Schwartz (1973) – both of whom treat migration as an investment increasing the productivity of human resources, which has costs and renders returns. Both authors also view the costs as having some nonpecuniary components, prominent among which are psychic costs.

Schwartz (1973) describes psychic cost as follows: "this cost is a result of the departure from family and friends. The longer the distance migrated, the lower will be the frequency of reunion; hence the higher will be the psychic cost" (Schwartz, 1973; page 1160). Sjaastad (1962) argues that since people are genuinely reluctant to leave familiar surroundings, migration involves a psychic cost that influences the private cost of migration to an individual.[6] It can also be argued that the construct of 'psychic costs'

[5] The gravity models hypothesized migration to be directly related to the size of the relevant origin and destination populations and inversely related to the distance. A more recent exposition of the gravity model in migration is offered by Lewer and Van den Berg (2008).

[6] As Sjaastad (1962) describes, given the earnings levels at all other places, there is some minimum earning level at location i which will cause a given individual to be indifferent between migrating and remaining at i. For any higher earnings at i, he collects a surplus, in the sense that part of his earnings could be taxed away and that taxation would not cause him to migrate. The maximum amount that could be taken away without inducing migration represents the value of the surplus. By perfect discrimination, it would be possible to take away the full amount of the surplus. The psychic cost of migration is analogous to this lost consumer surplus.

(of being away from family and friends) is the mirror image of the construct ‘psychic income’ (the benefits of being close to family and friends) articulated by Becker (1965).

The subsequent empirical literature offers some evidence of the existence of psychic costs. Lansing and Mueller (1967) conduct a survey of 723 moves between 1962 and 1963 and found that a large fraction were made to be closer to family members. Other studies such as Fabricant (1970), Nelson (1959), and Greenwood (1969) found evidence suggestive of psychic costs, a construct similar to “social attachment to place” in the sociology literature. To quote Dahl and Sorenson (2010b), “one commonly cited reason for why people do not move more often is that they value being near family and friends, or at least the more frequent and more extended interactions that propinquity allows” (page 637).

In the more recent literature, two important studies document that individuals are motivated to work at locations more personally attractive to them. Using a sample of foreign entrepreneurs in Russia, Kulchina (2016) provides evidence that entrepreneurs who view a host country as an attractive location are more likely to relocate and manage their firms personally. Yonker (2017) studies whether a preference for living and working closer to home explains why firms are five more times likely to hire a local CEO and finds strong evidence of the same. His results indicate that local CEOs have a lower turnover than nonlocal CEOs, a finding is driven by unforced turnover, not forced turnover.

Beyond psychic costs, as Schwartz (1973) outlines, there is a second mechanism behind how distance affects migrants: the greater the distance from home, the less information flow regarding opportunities in the host region. As we describe in detail later, the firm in our study randomly assigns employees across production centers, circumventing the possibility of this alternative mechanism affecting our results. In our setting, individuals are still affected by psychic costs related to separation from friends and family, which especially in the longer-term could adversely affect employee performance. We hypothesize:

Hypothesis 2: In the longer-term, employee performance will be negatively correlated to distance from home.

2.3. Heterogeneity of Longer-Term Effect based on Employee Characteristics

We next theorize how the negative effect of distance from home on employee performance in the longer-term could be attenuated by certain employee characteristics. We specifically theorize how such a negative effect could be attenuated by employee ability and/or her social embeddedness in the production center she is assigned to.

The literature on human capital has long documented that individuals will be highly motivated to exert extra effort in adverse situations, if they believe that doing so will lead to greater career rewards (Vroom, 1964; Katzell and Thompson, 1990).[7] In our context, all distant employees, irrespective of their short-term performance, are affected by psychic costs, especially in the longer-term. However, distant employees with higher ability might react differently to psychic costs in the longer-term. Higher-ability employees, even if they are distant from home, might anticipate greater eventual career rewards than distant employees with weaker performance in the short-term. Relatedly, the literature on careers has shown that employees value skill-development opportunities early in their careers and that larger organizations (such as the organization we study) are better placed to offer such opportunities (Bidwell and Briscoe, 2010). Distant employees with higher abilities might respond to the challenge of being far away from friends and family by seizing opportunities to acquire new skills (Eichinger, Lombardo and Ulrich 2004; DeRue and Wellman 2009, Dragoni et al. 2009) and might continue to perform better in the longer-term.

There is also a rich literature in sociology, human capital, and strategy on how social embeddedness affects employee performance (Granovetter 1985, Mitchell et al., 2001, Lee et al., 2004). As Granovetter (1992) states, the relational embeddedness argument stresses the role of strong dyadic ties in generating trust and cooperation among individuals. Mitchell et al. (2001) define embeddedness from several perspectives such as the extent to which people have links to other people (structural

[7] Relatedly, the Hull-Spence theory in psychology would suggest that stress and anxiety arising from threats in an adverse environment increase individual motivation and effort, especially if individuals perceive eventual greater rewards (Taylor and Spence 1952; Farber and Spence 1953; Spence and Farber 1953).

embeddedness). Extending this argument, Lee et al. (2004) state that the more an individual is socially embedded in an organization, the more likely he or she should be to display citizenship behaviors (i.e., mutually beneficial actions based on interdependent relationships), plausibly leading to higher productivity. There is also a rich literature on social embeddedness via homophily (Marsden, 1987; Marsden, 1988, McPherson et al. 2001, Ruef et al., 2003, Vissa, 2011). In particular, Vissa (2011) found that Indian entrepreneurs are more likely to have intentions of forming an interpersonal tie with new people who speak the same regional language.[8] In our context, high social embeddedness achieved through common regional language ties might act as a substitute for being close to family and friends in the longer-term, and might attenuate the negative relationship between distance from home and employee performance in the longer-term. Given this we hypothesize:

Hypothesis 3: The negative relation between distance from home and employee performance in the longer-term will be attenuated for (a) employees who are of ex ante higher ability and (b) employees with high social embeddedness within their assigned production center.

3. Empirical Setting: An Indian Technology Firm

To understand the effect of distance from home on employee productivity, we examine a unique administrative dataset from an Indian technology firm (hereafter TECHCO). It contains rich employee-level data, including demographic information, various proxies for ability measured during hiring and during training, and two performance ratings: one and three years after job assignment. In this section, we describe the institutional details behind hiring entry-level employees at TECHCO, particularly its human resource management practice related to randomly assigning entry-level employees across production centers, the sample construction process, and the description of variables.

[8] As Vissa (2011) states, the 2001 census of India reports 29 different regional languages each spoken by more than a million native speakers.

3.1. Hiring and Training Entry-Level Employees

We investigate the effect of distance from home on employee performance using data on newly hired entry-level employees. Such employees, hired from college campuses, are a suitable sample in which to study this question for several reasons. First, newly hired entry-level employees maintain strong social ties to family and friends in their hometown but often lack social ties with colleagues and new friends in the new work location. Therefore, distance from home is likely to affect their performance. Second, measuring employee performance is more objective and reliable for entry-level employees. Tasks assigned to them tend to be homogeneous, and objective performance measures are available, allowing comparisons across employees. Moreover, we are able to control for employees' innate abilities using proxies from various tests conducted during recruitment and training.

Every year TECHCO hires about 10,000 undergraduates from more than 250 colleges across India. These new hires are mostly from engineering colleges and have no prior full-time employment experience. TECHCO tends to hire a larger number of graduates from smaller towns in India than several of its peer IT firms in India; this hiring strategy implies that TECHCO is not hiring locally, within close proximity to its headquarters or production centers.

After being recruited, new hires receive an intensive four-month induction training. After recruitment and before training, TECHCO assigns new employees to one of several technological areas such as .NET, Java, or Mainframe. New employees then receive induction training at a centralized training center before being assigned to one of the production centers. Located in the southern Indian city of Mysore, the corporate training center has a 337-acre campus, 400 instructors, and 200 classrooms. Employees are trained in batches of about 100, and starting dates range from May to November. According to our field interviews, TECHCO spends around $3,500 training each new college graduate.

Upon completion of the four-month training, employees are assigned to one of the eight production centers located across India. TECHCO has more than 120,000 employees spread over those production centers and serves clients from all around the world. Most importantly for our empirical analysis, individual-level characteristics do not affect the assignment decision. As described in detail later, assignment is automated following pre-determined algorithms in the centralized enterprise resource planning system, which prevents employees from exerting any influence on the process. It is also very uncommon for an employee to transfer to a different location after initial assignment. Since understanding these institutional details is critical for our empirical purpose, we will describe them in depth later.

3.2. Sample Construction

We begin our sample construction process with 1,696 undergraduates who are hired by TECHCO in 2007 and assigned to the .NET technological area. We focus on only one technological area to minimize bias arising from temporal demand and supply fluctuations that may differently affect the performance ratings of employees working in other technology areas. For instance, we are less concerned about the possibility that an average employee assigned to the .NET area receives a higher performance rating simply because more foreign clients demand services related to this technological area over time. About 17% of all newly hired undergraduates in 2007 are assigned and trained in the .NET area. They are trained in 14 batches of 94 employees each and four batches of 95 employees each.

Using this sample, we aim to investigate how distance from home affects employee performance one and three years after assignment. Unfortunately, not all employees in the sample received a performance rating in the first year, so we further narrow our sample to those who did receive a first-year performance rating. We should be concerned about potential sampling bias if receiving the first-year rating is correlated with an employee's performance or any factors affecting it; however, this is not the case in our setting, because it is mostly determined by the "nine-month work rule."

The "nine-month work rule" dictates that an employee receives a performance rating if and only if he or she has worked on a coding/testing project for at least nine months in the calendar year. Our field interviews with HR managers at TECHCO suggest that the two most important factors that determine whether an employee worked for at least nine months in 2008 (i.e., the first full year after they are hired in 2007) are: 1) the timing of induction training completion, and 2) the availability of new coding/testing projects at the production center to which an employee was assigned. Factors that might affect an employee's performance, such as ability, project team quality, or the technical difficulty of a project, were not considered in determining which employees will work for nine months or more in 2008.

Given that whether an employee receives a first-year performance rating is orthogonal to individual-level characteristics, we are assured that our estimates would not be biased, even though we still drop observations with missing performance ratings in 2008. We validate this point in Table A3 by comparing individual-level observables between employees with and without a 2008 rating. As we expected, there is no systematic difference between the two groups.

Our final sample consists of 443 employees who are hired and trained in 2007 and then assigned into one of the eight production centers across India in 2008 (Table 1). This is after we further drop a few employee-level observations, given that travel-time data for some employees is unavailable because their hometowns are in foreign countries, in locations that are not accessible via train, or hometown information is missing from the firm's personnel database. We also drop the one employee among the cohort assigned to Chandigarh, for whom within-center comparisons would have been impossible.

Table 1 about here

3.3. Data and Summary Statistics

Table 2 presents summary statistics for our final sample of 443 employees. All variables are constructed at the employee level. Table 3 presents correlations among the variables. As described before, our main data source is the administrative employee database from TECHCO, which includes an employee's gender, various ability proxies, hometown location at the district level, and assigned production center location. We supplement that by hand-collecting the shortest travel time from workplace to the hometown by train. We also use the same personnel database to observe employee productivity.

Tables 2 and 3 about here

***Dependent variable*.** To measure individual performance, we use an employee's yearly performance ratings in 2008 (the first year since being assigned to a production center) and 2010 (three years after assignment). An appealing feature of this dependent variable is that it is constructed based on objective measures and thus less prone to measurement errors. At the end of each year, managers enter an initial performance rating for each employee. Field interviews with the head of talent development at TECHCO, a senior manager in HR, and several employees in the sample indicate that the performance ratings for entry-level employees are based on objective measures, including quality of coding and/or testing, measured using "mistakes" in code and timeliness/completeness in coding/testing and documentation, all tracked by automated software. HR managers check the rating against underlying scores to correct any errors made by the manager of the employee in computing the overall rating.

In the first year, newly hired employees receive one of three performance rating scores: one (high), two (average), or four (low). The final performance rating represents an employee's relative performance ranking compared to his or her peers. The left panel in Figure 1 presents the distribution of performance rating scores. In 2008, an employee receives the highest rating (one) if he or she falls in the top approximately 35% in the relative performance distribution; the second-highest rating (two) if he or she falls in the top 96%. The lowest rating (four) is given rarely, i.e., only if an employee falls in the bottom 4% in the relative performance distribution.

In the third year, the 2007 new hires similarly receive performance rating scores based on their relative performance ranking. However, the rating in the third year is based on a five-point scale: one (highest) to five (lowest). The right panel in Figure 1 displays the distribution. About the top 13 percent of employees receive the highest rating of one, and only 8 employees whose performance falls in the bottom 2 percent receive the lowest rating.

Figure 1 about here

In the regression analysis, we multiply the original performance ratings by -1 and use this transformed variable as our dependent variable. Originally the lower an employee's performance rating, the higher his or her relative performance. After the transformation, a numerically higher rating score indicates higher performance. This transformation makes the interpretation of the regression results more intuitive. As an example, we are able to interpret a positive coefficient as positive association between an independent variable and performance. It should be noted that the magnitude of estimated coefficients remains the same before and after the transformation.

Additionally, we collect data on whether an employee leaves the company by 2011 and code the variable *left the firm.* About 28% of employees in the sample have left the firm by 2011, when we completed the data collection exercise.

Independent variables. As our measure of distance from home to workplace, we manually construct a variable (*travel time*) that indicates shortest travel time (in hours, one way) from an employee's hometown to workplace via train. Our field interviews indicate that almost all newly hired college graduates use trains to travel to their hometown. We first identify an employee's hometown and production center from the personnel database, and code the shortest travel time manually from the Indian Railways official timetable. When there is no direct train path connecting the two locations, we calculate travel time by including minimum connection time for transfer. On average, it takes about 15.5 hours for an employee in the sample to travel from hometown to production center. Figure 2 plots the distribution of travel time to home.

Figure 2 about here

Controls. We also construct other employee-level controls, which are included in the following analyses to control for other factors affecting productivity. We first create a dummy variable to indicate the gender of the employee ($Male$). About 66% of the sample are male. Additionally, three measures are developed to capture an employee's ability. First, we record an employee's logical and verbal scores from standardized multiple-choice tests during recruitment. Second, we store the cumulative grade point average each employee receives by the end of the four-month induction training ($CGPA\ Training$) to represent productivity and is expected to be positively correlated with productivity after job assignment.

As proxies for cultural similarity between hometown and production center location, we develop a language similarity measure. Based on the recent Indian linguistics literature (Sengupta and Saha, 2015), we create a dummy variable $Similar\ Language$ that is equal to one if the official languages for an employee's hometown and his production center location belong to the same language family. Using various machine-learning techniques, the authors provide a method to classify Indian languages into a few families based on similarity

Finally, we measure whether a newly hired employee has prior migration experience in a dummy variable $Migration\ Experience$. To construct this, we compare employees' hometown location and university location at the district level, and code the variable as one if the locations are different. We include this as an additional control to rule out the alternative explanation that distant employees tend to have prior migration experience, which may help them perform better in the long run because of superior ability to adjust to new environments.

4. Identification Strategy

One might be tempted to simply regress individual performance on distance from home to characterize the relationship; however, such an approach suffers from two empirical issues. First, in reality, most firms hire employees from neighboring regions due to lower search costs. This is particularly the case for entry-level employees, as their skills are mostly homogeneous. When attempting to run the naïve regression, one is unlikely to find significant results, because there is not much variation in the travel time variable among employees.

More importantly, the naïve regression framework is highly likely to generate biased estimates when some unobservable is correlated with both the assignment decision and individual performance. For instance, it is possible that employees hired from Bangalore are of high quality because of knowledge spillovers from the many technology firms in that region. If TECHCO also tends to assign employees to the production center in Bangalore simply because it is close to their hometown, we would see the spurious correlation between travel time and individual performance in the naïve regression framework.

Luckily for our purpose, TECHCO adopted a computerized central talent-assignment system in which an employee's distance from home as well as other individual-level characteristics is *not* considered when deciding which production center the employee will be assigned to. In the following subsections, we describe qualitative and quantitative evidence elucidating this quasi-random assignment of employees, and how we exploit this institutional feature in the regressions.

4.1. Employee Assignment Protocol

We first discuss qualitative insights on why and how TECHCO implements the quasi-random assignment process for newly hired employees. We further validate this maintained assumption by conducting two supplementary quantitative analyses.

Understanding how each employee is allocated to a production center is central to our empirical analysis. Allocation is performed by a computer application called 'Talent Planning,' which is part of the firm's enterprise resource planning software. Talent Planning matches two factors: 1) individual production center requirements (HR at each center provides data on the number of employees needed in various technological areas); and 2) data from HR at the training location. Two weeks prior to the completion of training batches, HR at the training location releases data on which employees are expected to complete training. The two variables that the Talent Planning team considers while doing the matching on an automated system are details on the technology on which the employee was trained and the estimated date of training completion.

Most importantly for our econometric analysis, allocation of trainees to production centers is *not* correlated with distance from their home, their background, or their test scores before or after training. Field interviews with the head of talent development at TECHCO reveal that the primary motivation of this random, computer-driven talent allocation policy is to ensure that the end customers of TECHCO are indifferent to the location of the development center that executes their projects. The secondary motivation is to avoid regional and ethnic cliques at the production centers. To quote the head of talent development at TECHCO, "We do not want all Tamils to join the Chennai center or all Punjabis to join Chandigarh and start conversing in their regional language rather than in English. If that happens, both our clients and employees from other parts of the country are adversely affected."

We also offer quantitative support for our claim regarding the exogeneity of distance from home. First, we conduct Monte Carlo simulations to determine whether the realized mean value of distance from home is in fact not different from hypothetical distance-from-home values one would expect to see if the employee assignment is truly random. We randomly draw (with replacement) from the entire employee sample the same number of employees actually assigned to one of the eight locations. We conduct 1,000 random draws and present the sampling distribution of mean travel time values in the histogram. By comparing the sampling distribution with the realized mean value of distance from home, we are able to evaluate how similar or different the realized assignment results are from a truly randomized employee-assignment protocol. Figure 3 presents the sampling distribution of mean travel time when employee assignment is completely random. We also plot the realized mean travel time as a dashed line for comparison purposes. The realized mean value of travel time (i.e., mean value of travel time observed in our data) is not statistically different from the hypothetical mean value of travel time (i.e., where employee assignment is entirely random). This strengthens our belief in the validity of the random-assignment protocol.

Figure 3 about here

Second, we estimate the logit choice model with all covariates, including CGPA training, male, logical score, and verbal score, to test whether any of the covariates is correlated with the likelihood of being assigned to Bangalore. The production center in Bangalore is the largest and is regarded as the most important development center. If TECHCO strategically assigns newly hired employees based on individual-level characteristics, it probably wants to assign employees with higher underlying ability to the production center in Bangalore to maximize the center's performance, or assign employees with the shortest distance from home to lower the likelihood of attrition.

Table 4 contains the estimation results from the logit choice model. It shows that none of the individual-level observables is systematically correlated to assignment to Bangalore. Any ability measures, such as CGPA at the end of training or standardized test scores at recruitment, are not significantly related to assignment to Bangalore. The decision whether to allocate an employee to Bangalore is also not correlated with other observable individual characteristics such as gender. This validates our maintained assumption that no individual-level characteristics are considered in the employee-assignment process.

Table 4 about here

4.2. Model Specification

To examine how the travel time of employees from their current production center location to their hometown affects individual performance, we estimate the following equation separately for short- and longer-term performance:

$$Performance_{ij} = \alpha + \beta \cdot Travel\,Time_i + \gamma' X_i + \delta_j + \epsilon_{ij}$$

Here, $Performance_{ij}$ indicates performance rating for an employee i working in production center j. We use two performance ratings, measured at the end of 2008 and 2010 respectively, to shed light on both the short- and longer-term effect of distance from home on performance. The main independent variable $Travel\,Time_i$ is the minimum time (in hours) an employee would expect to spend traveling from the production center to his or her hometown by train. Our main coefficient of interest is β, which measures how an employee's performance is systematically related to distance from home. We include employee-level observables X_i to control for other factors that may affect productivity, such as gender and proxies for ability such as cumulative grade point average at the end of training and test scores during the standardized recruitment tests. In some specifications we include measures of language similarity between hometown and production center location to capture cultural differences between the two locations. In the base case, we estimate ordered logit models using Maximum Likelihood Estimation (MLE), given that performance rating is measured in normalized bands.

We also include location fixed effects for two reasons. First, they capture any production center–level differences across locations. While various management practices at TECHCO are designed to reduce quality differences across production centers, it is still highly plausible that some quality differences remain. For instance, a production center located near India's major technology cluster, such as Bangalore, is likely to have a higher concentration of knowledge because of agglomeration economies. By comparing employees within the same production center, we make sure that such external forces do not affect our analysis. Second, and specifically for our research design, we include center fixed effects so that distance from home is not systematically different across centers. Even though employees are randomly assigned to production centers, certain production centers located in central India are likely to have employees with shorter travel time than production centers in remote areas. Including center fixed effects rules out that possibility.

5. Results

In this section, we present our findings on the relationship between distance from home and individual performance. We begin by focusing on individual performance in the short-term (one year after recruitment) and repeat the analysis on employee productivity in the longer-term (three years after recruitment).

5.1. Distance from Home and Employee Performance in the Short-term

Table 5 presents the estimation results on the effect of distance from home on individual performance in the short-term (H1). Hypothesis 1 states that in the short-term, employee performance will be positively correlated to distance from home. In all specifications across Table 5, we find a positive and significant relationship between travel time and performance rating in the first year. That is, the farther away an employee is from his or her hometown, the more likely his or her short-term performance is higher. Column 1 reports baseline results, which control for an employee's gender, CGPA training, and location fixed effects. CGPA training is most strongly correlated with first-year performance, which is consistent with our field interviews. Column 2 additionally controls for employees' innate ability, which we think is captured in logical and verbal test scores during recruitment. Even after controlling for these strong predictors of employee productivity in the short-term, we still find a positive and statistically significant relationship between distance from home and individual performance in the first year.

Our proposed relationship between distance from home and short-term individual performance is robust after controlling for alternative explanations. More specifically, we consider two scenarios in which our main independent variable $Travel\ Time$ captures dimensions other than distance from home. The first is the language similarity between employees' hometowns and workplaces. To tease out the effects of cultural similarity (measured using language similarity) that could be captured in $Travel\ Time$, we include language similarity between employees' hometown and workplace region as an additional control variable. Column 3 presents the results. While the coefficient on language similarity is not statistically significant, the coefficient on $Travel\ Time$ is still significant and similar in magnitude.

Second, we deal with the possibility that distance from home is correlated with employees' prior migration experience. One might think that employees with such prior experience may be better positioned to perform better in the first year, as they might adjust more quickly to a new environment. To control for such effects, we include another control variable, $Migration\ Experience$, which is a dummy variable equal to one if an employee migrated from his or her hometown when he or she entered college. We present the results in Column 4. Unexpectedly, we find that an employee's prior migration experience is negatively associated with short-term performance; however, we still find a similar coefficient for $Travel\ Time$.[9] In addition, we include all control variables we considered previously and re-run the same analysis (Column 5); the hypothesized positive relationship between distance from home and short-term individual performance remains robust.

Table 5 about here

Given that we rely on an ordered logit model to establish the statistical significance of the effects of travel time on short-term individual performance, interpretation of the estimated coefficients from this specification is not straightforward. To understand its economic significance more intuitively, we calculate the predicted probability of receiving the highest performance rating in the first year, holding other independent variables at their mean but varying only travel time. We use the specification in Column 5 in calculating the predictive probability. The "average" employee here is male, has a CGPA training score of 4.55, a logical score of 5.07, a verbal score of 4.32, has prior migration experience but his native language is not similar to the language in the workplace, and works in Bangalore.

The top panel in Figure 4 presents the results graphically. As we saw in the regression results, there is a positive relationship between travel time and the likelihood of receiving the highest rating in the first year. Interpreting its economic significance, we find that for an average employee, an additional 10-hour increase in travel time leads to about 6 percent increase in the likelihood of receiving the highest performance rating in the first year. The probability that the average employee receives the highest rating

[9] A plausible explanation of this negative prior migration experience and performance relates to selection. It is possible that the individuals who never migrated for college, are disproportionately from smaller town colleges; such individuals might belong to the extreme right tail in the distribution of ability for individuals in their towns.

in the first year is only about 13 percent if his travel time is zero. In contrast, the probability increases to 43 percent if his travel time is 50 hours.

Figure 4 about here

5.2. Distance from Home and Employee Performance in the Longer-term

Table 6 presents the effects of distance from home on employee productivity in the longer-term (H2). Hypothesis 2 states that in the longer-term, employee performance will be negatively correlated to distance from home. Across all specifications in Table 6, we find a negative and statistically significant relationship between travel time and performance rating in the third year, a stark contrast to our previous findings on short-term performance. Our baseline specification results are presented in Column 1, in which we include gender, CGPA training, and location fixed effects as controls. We also include other measures of employee abilities as additional controls in Column 2 and find an almost identical relationship between travel time and longer-term individual performance.

As with the previous analysis on short-term individual performance, we also take potential alternative explanations into consideration, but the baseline finding on the negative relationship between travel time and longer-term individual performance does not change. First, in Column 3 we consider the possibility that travel time may capture language similarity between an employee's hometown and workplace location. Second, we separate the effects of employees' prior migration from the effects of distance from home, and present the results in Column 4. Finally, we include all control variables together and report the estimates in Column 5. In all specifications, we consistently find a negative relationship between travel time and longer-term individual performance.

Table 6 about here

Again, to help interpret the substantive meaning of the estimates, we estimate the probability that an average employee receives the highest rating in the longer-term. The graphical result is presented in

Figure 4 (bottom panel). For comparison purposes, we calculate the likelihood of receiving the highest (1) or second-highest rating (2) in 2010. Together, these two ratings capture the likelihood that one's performance falls in the top 25 percent of the relative performance distribution. This distribution is similar to the likelihood of receiving the highest rating (1) in 2008 when a three-value scale is used.

We find a negative relationship between travel time and the likelihood of receiving the highest rating in the longer-term. On average, a 10-hour increase in the travel time between hometown and workplace location is associated with 3.4 percent decrease in the likelihood of receiving the highest rating. If the average employee's travel time is zero, his likelihood of receiving the highest rating is predicted to be 29 percent. However, the likelihood decreases to 12 percent if his travel time is 50 hours.

So far, we have modeled the effect of travel time on longer-term performance as linear. However, it is possible that the relationship between travel time to hometown and longer-term performance is non-linear. For instance, visiting the hometown during the weekend may be equally feasible for employees whose travel time is 1 hour and 4 hours. However, employees who face much longer travel times back home may not be able to plan such frequent trips. In fact, during our field interviews, one of the employees in our sample noted that "if travel time each way is close to 24 hours, individuals are unable to visit their hometowns on long weekends".

We examine whether such a nonlinear relationship exists between travel time and longer-term performance. To do so, in the base case, we estimate a simple spline linear regression model with one "knot," or a breaking point around which the effect of travel time on performance differs significantly. Note that our goal here is not to estimate complex nonlinear functions with best predicting power, but simply to investigate the existence of a nonlinear relationship.

We follow the exploratory approach used in Gulati and Sytch (2008). First, we specify a spline regression specification in which the value of breaking point t_{bp} is to be determined. We then estimate the specification by varying t_{bp} by one unit for the entire range or travel time in the dataset, and choose t_{bp} that gives us the best model fit within this simple spline specification framework.

$$E\left[Performance_{ij,2010}\,\middle|\,X_{ij}\right]$$
$$= \alpha + \beta_1 \cdot Travel\,Time\ \times\ 1\left(Travel\,Time < t_{bp}\right) + \beta_2 \cdot Travel\,Time$$
$$\times\ 1(Travel\,Time \geq t_{bp})$$

Following this approach, we first find that the optimal t_{bp} is 23 hours. Results available with authors upon request use this optimal breaking point value and a spline linear regression, and demonstrate the existence of a nonlinear relationship between travel time and performance in the longer-term. We find strong evidence that travel time negatively affects longer-term performance if travel time exceeds 23 hours. This finding validates the insight from field interviews reported earlier. In robustness checks, we create four bins of travel time based on its quartile values. Results available with the authors suggest that the negative effect of travel time on longer-term performance exists only if travel time exceeds 22 hours (4^{th} quartile).

5.3. Heterogeneity of Longer-Term Effect based on Employee Characteristics

Hypothesis 3 stated that the negative relation between distance from home and employee productivity in the longer-term will be attenuated for (a) employees with high ability and (b) employees with high social embeddedness in the production center to which they are assigned. We examine that hypothesis by estimating a series of interaction models and testing whether the effect is heterogeneous based on the individual-level characteristics mentioned earlier.

To test H3a, we include the interaction term between travel time and whether an employee belongs to a low-ability group to determine whether the negative effect of travel time is of larger magnitude for employees with low ability. We create a dummy variable $Low\ Ability$ that is equal to one if an employee's CGPA training is below median, and zero otherwise. As we discussed earlier, CGPA training is the strongest predictor of employee performance in the short-term, so employees with low CGPA could plausibly be classified as having *ex ante* lower ability. More importantly, CGPA training is measured before an employee is assigned tasks at the production center and does not affect the assignment (of employee to production center) outcome, because of the random-assignment protocol described earlier. Our low-performer sub-sample consists of 223 employees, and the high-performer sub-sample consists of 220 employees (Table 2).

We then estimate the following interaction model using an ordered logit model to investigate whether the effect of distance from home on *ex ante* low-ability employees' longer-term productivity is different from the effect on high-ability employees. H3a is supported if the coefficient on the interaction term, β_3, is estimated as negative.

$$Performance_{ij,2010} = \alpha + \gamma' X_i + \delta_j +$$

$$\beta_1 \cdot Travel\,Time_i + \beta_2 \cdot Low\,Ability_{\mathrm{i}} + \beta_3 \cdot Travel\,Time_i \times Low\,Ability_{\mathrm{i}} + \epsilon_{ij}$$

Table 7 presents the regression results. Column 1 replicates our previous finding that on average, longer travel time is negatively associated with longer-term employee performance. Column 2 tests the validity of H3a. The impact of travel time on longer-term performance is not significant for high-ability employees (β_1). However, the estimate on the interaction term (β_3) is negative and statistically significant at the 10-percent significance level, suggesting that the effect of travel time on longer-term performance differs between high- and low-ability employees. For a representative low-ability employee, for instance, a 10-hour increase in travel time is associated with a 2.6 percent decrease in the likelihood of receiving the highest performance rating in the longer-term (Figure 5, top panel, dotted line). This would be a substantial difference, given that only 13 percent of employees receive the highest rating in 2010.

Table 7 and Figure 5 about here

Now we test H3b, i.e., whether the negative effect of distance on longer-term performance is mitigated by the social embeddedness of the employee. To do this, we first need to construct a measure of social embeddedness in the production center to which an employee is assigned, since we do not have a direct measure of such embeddedness. Instead, relying on the homophily literature (Marsden, 1987; Marsden, 1988; Vissa, 2011), which suggests that individuals tend to form social ties when they share similar cultural backgrounds, including languages, we hypothesize that employees whose native language is frequently used among their cohort will tend to have a greater number of social ties in their production center than employees whose native language is less frequently used.[10] Employees who are members of the 'majority' native language group among the cohort in each production center are less likely to suffer from the psychic costs of being distant from home.

To operationalize this idea, we first define a dummy variable $Low\ Social\ Embeddedness$. This variable is zero for employees whose native language is spoken by the majority of other employees within each cohort at each location; and one otherwise. We use employees' hometown location to infer their native language, and then identify for each production center the distribution of native language among 2007 intakes. For instance, there are 105 intakes assigned to Bangalore in 2007, with Telugu and Kannada the two most heavily spoken native languages. Together, about 63 percent of the 2007 intakes in Bangalore speak either language. We coded $Low\ Social\ Embeddedness$ as zero for employees using either Telugu or Kannada, as together these employees account for more than 50 percent of all 2007 intakes assigned to Bangalore, and one for other employees in Bangalore. To identify "heavily spoken native languages" more generally, we use the sub-set of languages for which the cumulative number of speakers exceeds at least 50% of the cohort population at the center for 2007.
$Low\ Social\ Embeddedness$ is coded as one for 223 employees, and zero for others.

We then estimate the following interaction model using an ordered logit model to investigate whether the effect of distance from home on employees' longer-term productivity is heterogeneous according to the size of friendship networks. H3b is supported if the coefficient on the interaction term, β_3, is estimated as negative.

$$Performance_{ij,2010} = \alpha + \gamma' X_i + \delta_j +$$

$$\beta_1 \cdot Travel\ Time_i + \beta_2 \cdot Low\ Social\ Embeddedness_i +$$

[10] For a more comprehensive review on homophily and the formation of social networks, see McPherson and Smith (2001).

$$\beta_3 \cdot Travel\ Time_i\ \times\ Low\ Social\ Embeddedness_i + \epsilon_{ij}$$

Column 3 in Table 7 presents the results. We find evidence supporting H3b. For employees with high social embeddedness, we find that the relationship between travel time and long-term productivity is in fact positive, although not statistically significant (β_1). In contrast, the coefficient on the interaction term (β_3) is negative and statistically significant. This implies that the negative effect of travel time on longer-term performance is much more salient for employees with lower social embeddedness. The bottom panel in Figure 5 displays the finding graphically. For an average employee with low social embeddedness, a 10-hour increase in travel time is associated with about a 3 percent decrease in the likelihood of receiving the highest performance rating in the longer-term.

5.4. Ruling out Alternative Explanations

Our empirical results present the contrasting effects of travel time on short and longer-term employee performance. We reconcile this seemingly puzzling finding by relying on the concept of the mirror-image concepts of 'psychic income' (from being close to friends and family) (Becker, 1965) and 'psychic costs' of being away from friends and family (Sjaastad, 1962; Schwartz, 1973). In this section, we consider other potential theoretical frameworks that might explain our empirical results. Specifically, we evaluate the role of burn-out and endogenous attrition.

The first alternative explanation is based on the possibility of burn-out. Distant employees perform better than local employees in the short-term, potentially because distant employees exert greater effort in the short-term that may not be sustainable in the longer-term. As a result, distant employees burn out more easily in the longer-term, and their performance decreases faster than that of local employees.

However, our data suggest that this is not the case in our setting. If the burn-out mechanism is a dominant explanation, then the effects of the negative relationship between travel time and performance changes over time should be larger for our high-performer sub-sample. To conduct the sub-sample analysis, we divide the sample by performance in 2008 into low performers and high performers. Because of statistical power concerns, we group two ratings, “Below” and “Average,” into one. As a result, our low-performer sub-sample consists of 234 employees, and our high-performer sub-sample consists of 151 employees. Then we separately estimate the relationship between travel time and performance changes between 2008 and 2010, using the same specification introduced earlier. For all specifications, we find a negative and significant relationship between travel time and performance changes between 2008 and 2010 among the low-performer sub-sample, but do not find similar results in the high-performer sub-sample. This suggests that the burn-out mechanism fails to explain our findings. (These results are available by request from the authors.)

An important alternate explanation that could confound results related to H2 and H3 is the endogenous attrition of employees.[11] To rule out this possibility, we first develop testable necessary conditions under which this attrition-based mechanism would explain the changes in the relationship between distance from home and employee productivity over time. We then rely on our data to test whether this explanation is satisfactory.

Theoretically, it is possible that distant employees are more likely to leave the company than comparable nearby employees. We consider two explanations here (note that these explanations are not mutually exclusive, and it is not our goal to evaluate which explanation is more plausible in our empirical setting). First, from a demand-side perspective, distant employees are more likely to leave the company because of their higher productivity in the short-term, possibly because of more and better outside career opportunities. Second, from a supply-side perspective, distant employees are more likely to leave, as they may suffer from various psychic costs of being distant from home. It could be that distant employees do not enjoy working at distant locations, as they miss friends and family in their hometown, and so they are

[11] To be clear, not all employees with missing performance rating in 2010 have left the firm. We have 58 employees with missing performance rating in 2010, among which only 51 employees have left the firm. The performance rating for the other 7 employees is missing because of the “nine-month rule”. In Table A4, we compare observables of employees with 2010 performance rating to the 7 employees with missing performance rating in 2010. While the sample size is too small to reach any concrete conclusion, we find that these two groups are mostly comparable, except in logical score.

more likely than nearby employees to leave the company and search for career options closer to their hometown.

If employee composition changes explain the contrasting effects of distance from home on employee performance over time, we should observe that distant employees with a high-performance rating in the short-term are more likely to leave the firm.

We test whether this condition is observed in the dataset. To reiterate, the necessary condition is that high-performing distant employees are more likely to leave the firm. We use the information on whether an employee leaves the firm as a binary dependent variable, and examine whether it is correlated with the interaction term between travel time and performance in the short-term. As the dependent variable is binary, we estimate the following logit model specification using MLE:

$$Left\ the\ firm_i = \alpha + \beta_1 \cdot Travel\ Time_i + \beta_2 \cdot Travel\ Time_i \times High\ Performer\ in\ 2008_i\ + \gamma' X_i + LOCATION_j + \epsilon_i$$

The dummy variable *High Performer in 2008* is equal to one if an employee receives the highest rating in the first year. Other control variables include gender, CGPA training, language similarity, logical and verbal scores, and prior migration experience.

The estimated coefficients are presented in Table 8. Column 1 presents the baseline results without the interaction term. We already see that the coefficient on travel time is close to zero and not statistically significant, suggesting that distance from home in fact has no role in explaining employee attrition in our sample. Our key results with the interaction term are presented in Column 2. We do not find evidence that high-performing distant employees are more likely to leave the company. Therefore, we conclude that the attrition-based mechanism is not satisfactory to make sense of the contrasting effects of travel time on employee productivity over time.

Table 8 about here

5.5. Additional Robustness Checks

After establishing that the attrition-based mechanism does not explain the disparity in our main results between the short and longer-term, we now turn to other robustness checks. First, we explore whether the estimates change with different functional form assumptions. Second, we check whether a few outliers in fact drive the entire results.

To test whether our results are sensitive to specific functional form assumptions, we re-run all specifications using OLS rather than using the ordered logit model estimated by the MLE. In the previous analyses, we prefer the ordered logit model, because it does not impose the assumption that the differences between cut points are substantially homogeneous. Estimating the main specification using OLS gives us substantially similar results (Table A1; Table A2).

To explore the possibility that a few outliers are driving the main results, we perform two sensitivity tests. First, we use the winsorization technique and replace the extreme distance from home values beyond the bottom and top 5 percentiles with less-extreme values at each percentile. We get very similar results when we re-run the analyses. Second, we drop observations from the three smallest production centers, Mangalore, Trivandrum, and Chandigarh, and re-run the analyses. As Table 1 shows, fewer than 20 employees in the sample are assigned to Mangalore and Trivandrum (and to recap, only one employee was assigned to Chandigarh), so making any comparisons within each center for these samples may not be reasonable. Our results are robust to dropping employees assigned to these three centers.

We also tested whether our final sample is similar or different from other 2007 employees without the first-year performance rating in 2008. If our sample is different in some aspects, then our findings might not be informative even to TECHCO because of generalizability issues. Reassuringly, we do not find such evidence. In Table A3, we compare observable characteristics of employees in our sample with other 2007 intakes without the first-year performance rating in 2008. To avoid confounding issues, we use three individual-level characteristics that are pre-determined before assignment to training batches: gender and logical and verbal scores in the recruitment test. The p-values indicate that we cannot reject the null hypothesis that our sample employees and the remaining 2007 intakes are from the same population.

6. Discussion and Conclusion

In this paper, we attempt to establish a causal relationship between distance from hometown and individual performance in the short and longer-term. We exploit a unique HR protocol within a large Indian technology firm in which entry-level employees hired from colleges across India are randomly assigned to eight production centers owned by the firm, also distributed across India. Our findings suggest that travel time to the hometown has opposite effects on individual performance in the short and longer-term. In the short-term, travel time to hometown positively affects short-term individual performance, i.e., the farther an employee is from his or her hometown, the more likely that his or her first-year performance rating is higher. However, that relationship reverses in the longer-term: employees with longer travel time tend to receive lower performance ratings three years after assignment. The negative relationship between travel time and longer-term performance is particularly salient for employees who need to travel longer than 23 hours to visit their hometown. We employ the mirror-image concepts of 'psychic income' from leisure and from spending time with friends and family (Becker, 1965) and 'psychic costs' of being away from friends and family (Sjaastad, 1962; Schwartz, 1973) to theorize about our results. We control for several alternative explanations such as attrition and burnout. We also provide evidence that the negative effect of distance on longer-term performance is attenuated if the individual is (a) of *ex ante* higher ability and (b) socially embedded in the production center cohort through regional language ties.

Our findings make a valuable contribution to the nascent literature on how personal preferences drive the geography of work for individuals. While that literature has provided empirical evidence that scientists, engineers, and entrepreneurs choose work that is close to home (Dahl and Sorenson, 2010a, 2010b, 2012, Kulchina 2016, Yonker 2017), to the best of our knowledge, no prior studies have established a causal relationship between distance from home and individual performance. Researchers such as Dahl and Sorenson (2010) also theorize about more than one underlying mechanism driving individuals to seek work closer to family and friends. It is possible that family and friends act as sources of information on job opportunities (the 'information' hypothesis); it is also possible that individuals value frequent interactions with family and friends through propinquity (the 'psychic cost or social attachment' hypothesis). Our setting allows us to control for the information hypothesis (individuals are not deciding to move to a location based on information they possess; the firm randomly assigns employees to production centers) and as a result, we are able to focus exclusively on the effect on productivity of being away from family and friends and losing out on frequent interactions with loved ones. We invoke an important and yet relatively less well-studied theoretical construct in the literature, i.e., the psychic costs of being away from family and friends, and provide empirical evidence suggesting that in the longer-term, employees working far from their hometown (especially those who are not socially embedded in their workplace and/or are of *ex ante* lower ability) experience psychic costs that affect individual productivity. An important caveat is that we do acknowledge that family and friends could act as sources of information and resources, but our empirical setting does not allow us to study these mechanisms. However, our setting does provide an opportunity to validate the psychic costs/social attachment hypothesis.

Our findings also contribute to two streams of the literature on strategic human capital – one focused on hiring and another on employee mobility. The literature in strategic human capital has long explored the topic of firms hiring external employees (Bidwell 2011; Dokko et al., 2009; Bidwell and Keller, 2014). While individuals might be located across the country, there are agglomeration economies with respect to firm location (Shaver and Flyer, 2000; Alcacer and Chung, 2014), and being hired by a different firm often involves geographic mobility on part of the employee (Song et al. 2003). Working for geographically distant firms could help individuals if such mobility enables them to leverage inter-organizational career ladders, i.e., working in certain kinds of organizations earlier in a career and in other kinds of organizations later, acquiring skills and developing human capital (Bidwell and Briscoe, 2010).

However, the literature also offers a set of conditions under which firms are less likely to hire externally – e.g., when job performance has a strong firm-specific component (Bidwell and Keller, 2014) or when promotions entail a great expansion in responsibilities, such as in promotions from lower- to upper-management positions (Bidwell and Mollick, 2015). Our insights add to this literature and suggest yet another consideration for firms evaluating external hires vis-à-vis promoting internal employees. In other words, when deciding which production center/office to assign employees to, firms should consider the twin constructs of psychic income/psychic costs of proximity to family and friends and their effect on individual performance.

Our insights are also relevant to the literature on employee mobility. Firms could benefit from hiring distant employees from a learning-by-hiring and knowledge-flows perspective (Rosenkopf and Almeida 2003, Song et al. 2003). As Rosenkopf and Almeida (2003) have articulated, external hires can serve as bridges to distant contexts. Song et al. (2003) argue that external hiring can extend the geographical boundaries of interfirm knowledge transfer and offer evidence that hiring distant domestic or even international employees is similarly conducive to learning-by-hiring. Our research adds to this literature by highlighting the effects of distance from home of mobile employees on their productivity. A recent study in this literature (Choudhury, 2017) highlights the importance of 'temporary mobility,' i.e., intra-firm mobility, on assignments that last for a few weeks. Our results indicate that for distant employees, mobility could have very different effects on employee productivity based on whether the mobility is temporary or permanent.

Our findings are also relevant to the migration literature. While the idea of psychic costs was at the forefront of this literature in the 1960s-1970s (Sjaastad, 1962; Bowles, 1970; Schwartz, 1973; Greenwood, 1975; Ritchey, 1976) and was also discussed in migration studies of the 1990s (e.g. Zhao, 1999), to the best of our knowledge, there has been no empirical study on how psychic costs could affect the individual productivity of migrants in the longer-term. In fact, Borjas' seminal 1994 study on migration talks about the 'costs of migration' but does not discuss psychic costs: "Migration costs C will differ among workers. For instance, newly arrived immigrants may be unemployed while they look for employment, suggesting that high-wage migrants might have higher migration costs. High-wage migrants, however, are more likely to have prior job connections and better information about job opportunities, suggesting a negative correlation between migration costs C and wages. The immigrant also incurs transportation costs" (Borjas, 1994; page 1688). Our results indicate that the underlying model of self-selection in the context of migration (i.e., Roy, 1951) should take into account variable migration costs and acknowledge the psychic costs of migration. In fact, Borjas (1994) does urge the field to consider an extension of the Roy (1951) model by incorporating variable migration costs.

Our study has several limitations. Given that we focus on a single firm (in the spirit of the "insider-econometrics" approach), there could be concern around the external validity or generalizability of our results. First, our findings might not be applicable to smaller countries, or countries whose transportation is more developed than India's. Recall that the mean value of our travel time is about 16 hours, and the maximum value is 49 hours. It would be interesting to determine whether one could still identify a relationship between distance from home and employee productivity in smaller countries where air travel might be more economical and feasible, as well as in international settings where employees are assigned to workplace contexts outside of their home countries. Second, given that the psychic costs of being far away from family and friends might affect employees more strongly early in their careers than it would later, a follow-up question for research is whether this finding could change when employees acquire families of their own.[12] A third limitation of our study is the three-year time frame. It is plausible that the longer-term negative effect of distance on individual performance is reversed when the employee gets married and begins a family of her own. This is related to the U-curve adjustment theory in the field of cross-cultural adjustment (Lysgaard 1955; Adler 1986), which refers to four phases in cultural adjustment for migrants: the (i) honeymoon, (ii) culture shock, (iii) adjustment, and (iv) mastery. It is plausible that our short-term and longer-term results relate to the honeymoon and culture-shock phases,

[12] Here it is important to note that *none* of the employees in our sample were married or had kids during the period of our study. We confirmed this insight during our field interviews.

respectively. Future work should explore whether psychic costs undergo inversion over longer periods of time.

Our insights open up several other avenues for future research. It would be interesting to study other workplace substitutes and complements to family and friends. It would also be interesting to study other interventions that firms could implement to mitigate the psychic costs incurred by employees hired from far away. Finally, it would be interesting to study whether there is heterogeneity in the effects of distance from home on employee performance across countries and career stages.

In conclusion, our study provides important causal evidence on how distance from home affects individual performance in the short-term and the longer-term and advances the nascent literature of the geographic preferences of knowledge workers, entrepreneurs, and CEOs. We invoke the mirror-image constructs of psychic income from leisure (by being close to family and friends) and psychic costs of being far away from family and friends, concepts introduced by economists in the 1960s but sparsely used in the recent literature, and exploit a randomized employee assignment protocol to provide causal results. Our results speak to the literatures in hiring, employee productivity, and migration and have several managerial implications. Several recent articles in the popular press indicate that individuals increasingly prefer to live close to their hometown. In one such study, 61% percent of respondents said the probability of relocation for work was not very likely—with 41% saying it wasn't likely at all.[13] It is plausible that individuals are less likely to migrate far from home due to the psychic costs of being away from family and friends. If future research corroborates this fact, this would indicate that managers would be well served to hire locally and/or mitigate the psychic costs incurred by distant employees by creating a "home away from home" in their distant workplace. Our results also suggest that firms might maximize employee performance by assigning them to geographically distant locations for short periods of time, and by bringing them back to locations closer to their hometowns before the enhanced performance begins to attenuate.

---

[13] Source: https://www.theatlantic.com/business/archive/2015/03/staying-close-to-home-no-matter-what/387736/. Website accessed on February 8, 2017.

References

Adler, Nancy J. "Cross-cultural management research: The ostrich and the trend." Academy of management Review 8.2 (1983): 226-232.

Alcacer, Juan, and Wilbur Chung. "Location strategies for agglomeration economies." Strategic Management Journal 35.12 (2014): 1749-1761.

Arrow, Kenneth. "Economic welfare and the allocation of resources for invention." In *The rate and direction of inventive activity: Economic and social factors*, pp. 609-626. Princeton University Press, 1962.

Becker, Gary S. "A Theory of the Allocation of Time." *The Economic Journal* 75.299 (1965): 493-517.

Bidwell, Matthew. "Paying more to get less: The effects of external hiring versus internal mobility." *Administrative Science Quarterly* 56.3 (2011): 369-407.

Bidwell, Matthew, and Forrest Briscoe. "The dynamics of interorganizational careers." *Organization Science* 21.5 (2010): 1034-1053.

Bidwell, Matthew, and J. R. Keller. "Within or without? How firms combine internal and external labor markets to fill jobs." *Academy of Management Journal* 57.4 (2014): 1035-1055.

Bidwell, Matthew, and Ethan Mollick. "Shifts and ladders: Comparing the role of internal and external mobility in managerial careers." *Organization Science*26, no. 6 (2015): 1629-1645.

Borjas, George J. *"The Economics of Immigration." Journal of Economic Literature, 32.4 (1994): 1667-1717.*

Bowles, Samuel. "Migration as investment: Empirical tests of the human investment approach to geographical mobility." *The Review of Economics and Statistics* (1970): 356-362.

Carrothers, Gerald AP. "An historical bedew of the gravity and potential concepts of human interaction." *Journal of the American Institute of Planners* 22.2 (1956): 94-102.

Chattopadhyay, Shinjinee, and Prithwiraj Choudhury. "Sink or Swim: The Role of Workplace Context in Shaping Career Advancement and Human-Capital Development." Organization Science, 2017.
Choudhury, Prithwiraj. "Innovation Outcomes in a Distributed Organization: Intra-Firm Mobility and Access to Resources." *Organization Science*, 2017.

Dahl, Michael S., and Olav Sorenson. "The migration of technical workers." *Journal of Urban Economics* 67.1 (2010a): 33-45.

Dahl, Michael S., and Olav Sorenson. "The social attachment to place." *Social Forces* 89.2 (2010b): 633-658.

Dahl, Michael S., and Olav Sorenson. "Home sweet home: Entrepreneurs' location choices and the performance of their ventures." *Management science* 58.6 (2012): 1059-1071.

DeRue, D. Scott, and Ned Wellman. "Developing leaders via experience: the role of developmental challenge, learning orientation, and feedback availability." *Journal of Applied Psychology* 94, no. 4 (2009): 859.

Dokko, Gina, and Lori Rosenkopf. "Social capital for hire? Mobility of technical professionals and firm influence in wireless standards committees." *Organization Science* 21, no. 3 (2010): 677-695.

Dokko, Gina, Steffanie L. Wilk, and Nancy P. Rothbard. "Unpacking prior experience: How career history affects job performance." *Organization Science* 20.1 (2009): 51-68.

Dragoni, Lisa, Paul E. Tesluk, Joyce EA Russell, and In-Sue Oh. "Understanding managerial development: Integrating developmental assignments, learning orientation, and access to developmental opportunities in predicting managerial competencies." *Academy of Management Journal*52, no. 4 (2009): 731-743.

Eichinger, R., M. Lombardo, and D. Ulrich. "Best People Practices for Manager & HR." (2004).

Fabricant, Ruth A. "An expectational model of migration." *Journal of Regional Science* 10.1 (1970): 13-24.

Farber, I. E., and Kenneth W. Spence. "Complex learning and conditioning as a function of anxiety." *Journal of Experimental Psychology* 45, no. 2 (1953): 120.

Gelber, Alexander M., and Joshua W. Mitchell. "Taxes and time allocation: Evidence from single women and men." *The Review of Economic Studies* 79, no. 3 (2011): 863-897.

Gilfillan, S. C. "The Sociology of Innovation." (1935).

Gulati, Ranjay, and Maxim Sytch. "Does familiarity breed trust? Revisiting the antecedents of trust." *Managerial and Decision Economics* 29.2-3 (2008): 165-190.

Granovetter, Mark. "Economic action and social structure: The problem of embeddedness." American journal of sociology 91.3 (1985): 481-510.

Granovetter, M., 1992. Problems of explanation in economic sociology. Networks and organizations: Structure, form, and action, 25, p.56.

Graff Zivin, Joshua, and Matthew Neidell. "Temperature and the allocation of time: Implications for climate change." *Journal of Labor Economics* 32, no. 1 (2014): 1-26.

Greenwood, Michael J. "An analysis of the determinants of geographic labor mobility in the United States." *The review of Economics and Statistics* (1969): 189-194.

Gronau, Reuben. "The allocation of time of Israeli women." *Journal of Political Economy* 84.4, Part 2 (1976): S201-S220.

Hall, Douglas T. *Career development in organizations*. Jossey-Bass Inc Pub, 1986.

Heckman, James J., and Guilherme Sedlacek. "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." *Journal of Political Economy* 93, no. 6 (1985): 1077-1125.

Isard, Walter. "The scope and nature of regional science." *Papers in Regional Science* 6.1 (1960): 9-34.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics* 108, no. 3 (1993): 577-598.

Kalnins, A. and Lafontaine, F., 2013. Too far away? The effect of distance to headquarters on business establishment performance. American Economic Journal: Microeconomics, 5(3), pp.157-179.

Katzell, Raymond A., and Donna E. Thompson. "Work motivation: Theory and practice." *American psychologist* 45, no. 2 (1990): 144.

Kerr, Sari Pekkala, William R. Kerr, Çağlar Özden, and Christopher Parsons. "High-Skill Migration and Agglomeration." *Annual Review of Economics* 9, no. 1 (2017).

Kulchina, Elena. "Personal preferences, entrepreneurs' location choices, and firm performance." *Management Science* 62, no. 6 (2015): 1814-1829.

Lansing, J.B., Mueller, E. and Survey research center (Ann Arbor, Mich.), 1967. The geographic mobility of labor. Ann Arbor: Survey Research Center, University of Michigan.

Lee, Thomas W., Terence R. Mitchell, Chris J. Sablynski, James P. Burton, and Brooks C. Holtom. "The effects of job embeddedness on organizational citizenship, job performance, volitional absences, and voluntary turnover." *Academy of Management Journal* 47, no. 5 (2004): 711-722.

Lewer, Joshua J., and Hendrik Van den Berg. "A gravity model of immigration." *Economics letters* 99, no. 1 (2008): 164-167.

Marsden, Peter V. "Core discussion networks of Americans." *American sociological review* (1987): 122-131.

Marsden, Peter V. "Homogeneity in confiding relations." *Social networks* 10, no. 1 (1988): 57-76.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a feather: Homophily in social networks." *Annual review of sociology* 27, no. 1 (2001): 415-444.

Mitchell, Terence R., Brooks C. Holtom, Thomas W. Lee, Chris J. Sablynski, and Miriam Erez. "Why people stay: Using job embeddedness to predict voluntary turnover." *Academy of management journal* 44, no. 6 (2001): 1102-1121.

Nelson, Lowry. "An Italian Farm and Family: 1934 and 1955." *Land Economics* 35.4 (1959): 362-365.

Olsson, Gunnar. "Distance and human interaction. A migration study." *Geografiska Annaler. Series B, Human Geography* 47.1 (1965): 3-43.

Oyer, Paul, and Scott Schaefer. *Firm/employee matching: An industry study of American lawyers*. No. w18620. National Bureau of Economic Research, 2012

Ritchey, P. Neal. "Explanations of migration." *Annual review of sociology* 2, no. 1 (1976): 363-404.

Rosenkopf, Lori, and Paul Almeida. "Overcoming local search through alliances and mobility." *Management science* 49.6 (2003): 751-766.

Roy, A. D. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3.2 (1951): 135–146.

Schein, Edgar H. *Career dynamics: Matching individual and organizational needs*. Vol. 6834. Addison Wesley Publishing Company, 1978.

Schwartz, Aba. "Interpreting the effect of distance on migration." *Journal of Political Economy* 81.5 (1973): 1153-1169.

Sengupta, Debapriya, and Goutam Saha. "Study on similarity among Indian languages using language verification framework." Advances in Artificial Intelligence 2015 (2015): 2.

Shaver, J. Myles, and Fredrick Flyer. "Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States." Strategic Management Journal (2000): 1175-1193.

Singh, Jasjit, and Ajay Agrawal. "Recruiting for ideas: How firms exploit the prior inventions of new hires." *Management Science* 57, no. 1 (2011): 129-150.

Sjaastad, L. A. "The costs and returns of human migration." *Journal of Political Economy* 70.5 Pt 2 (1962): 80-93.

Song, Jaeyong, Paul Almeida, and Geraldine Wu. "Learning–by–Hiring: When is mobility more likely to facilitate interfirm knowledge transfer?." *Management science* 49.4 (2003): 351-365.

Spence, Kenneth W., and I. E. Farber. "Conditioning and extinction as a function of anxiety." *Journal of Experimental Psychology* 45, no. 2 (1953): 116.

Stanton, Christopher, and Catherine Thomas. "Landing the first job: The value of intermediaries in online hiring." *The Review of Economic Studies* (2015): rdv042.

Taylor, Janet A., and Kenneth W. Spence. "The relationship of anxiety level to performance in serial learning." *Journal of Experimental Psychology* 44, no. 2 (1952): 61.

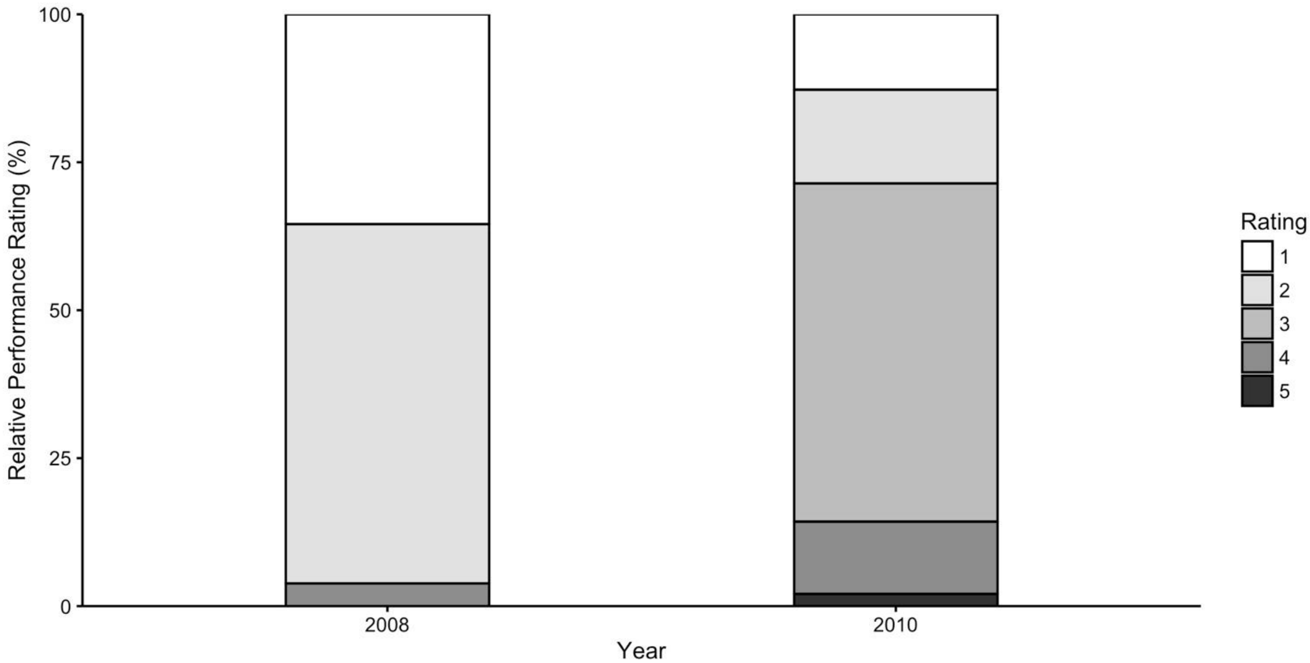
Vissa, Balagopal. "A matching theory of entrepreneurs' tie formation intentions and initiation of economic exchange." Academy of Management Journal 54.1 (2011): 137-158.

Vroom, V. "Motivation and work." (1964).

Yonker, Scott E. "Geography and the market for CEOs." *Management Science* (2016).

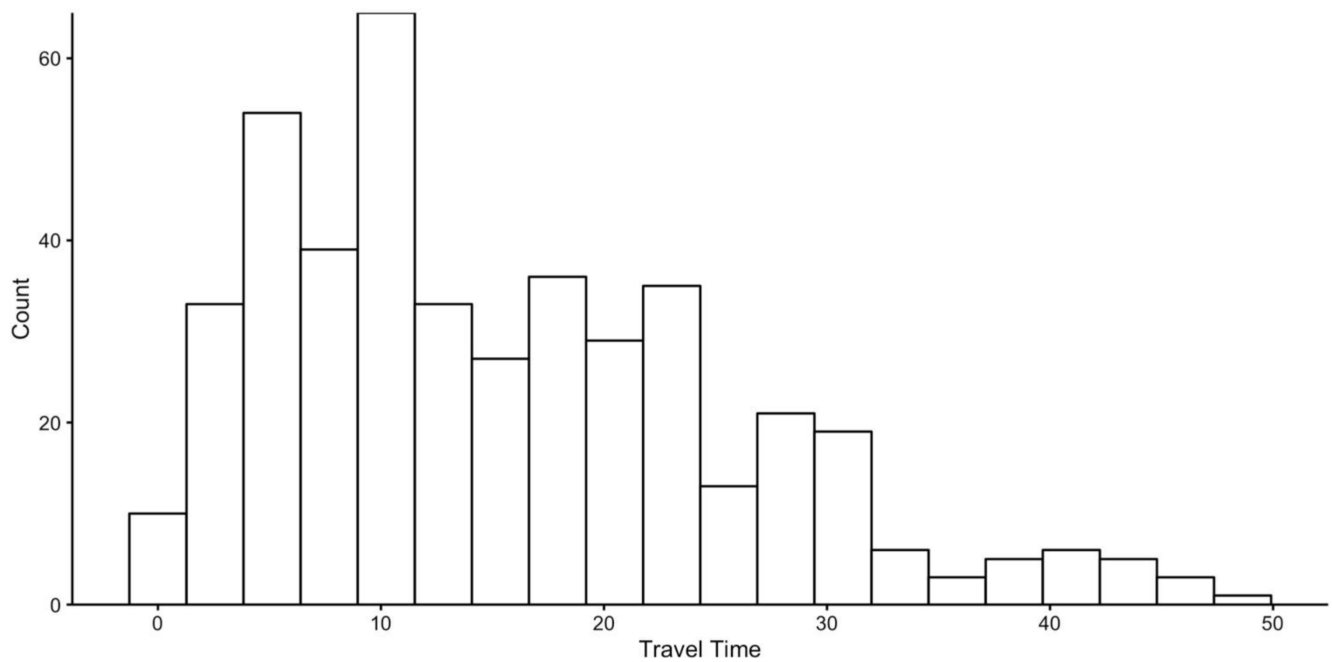
Zhao, Yaohui. "Leaving the countryside: rural-to-urban migration decisions in China." *The American Economic Review* 89, no. 2 (1999): 281-286.

Figure 1
Distribution of Performance Ratings in 2008 and 2010



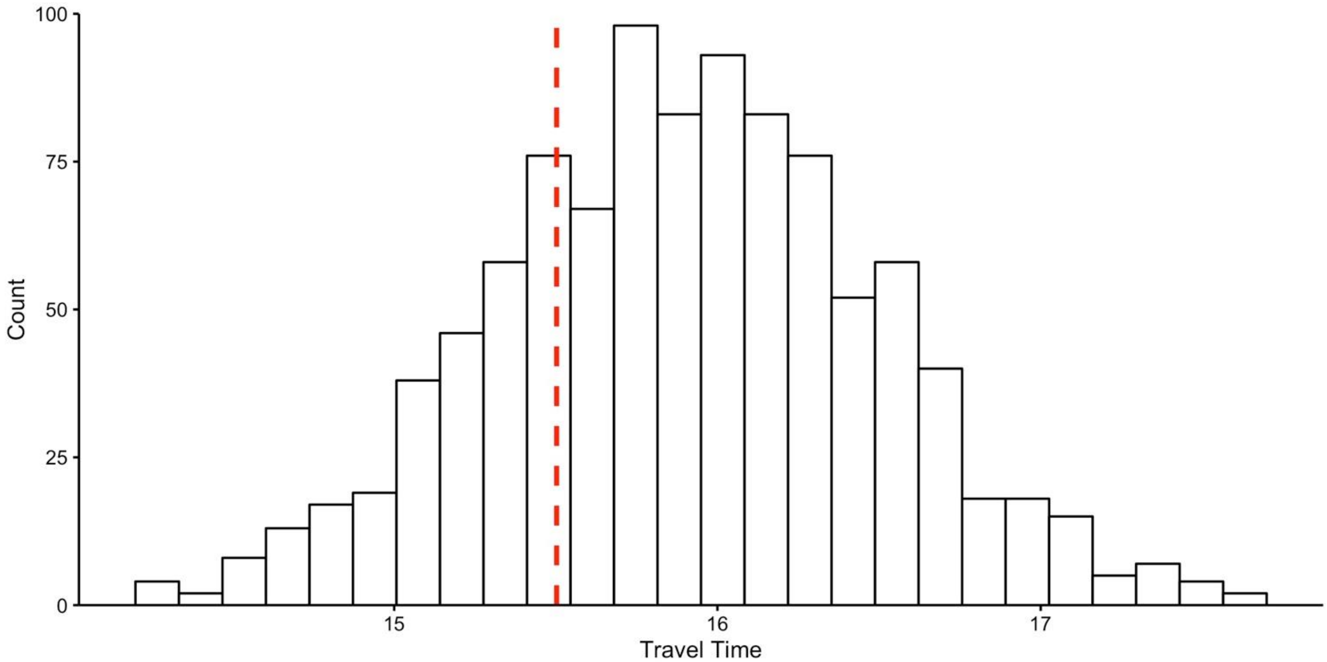
Note: This 100% stacked bar graph presents the distribution of performance ratings in 2008 (left) and 2010 (right). To plot this figure, we use original performance ratings in which lower score represents higher relative performance. In the regression analyses, we multiply the original ratings by -1 and use the transformed ratings as the dependent variables. With this transformation, we are able to interpret a positive regression coefficient as positive association between a focal independent variable and performance.

Figure 2
Histogram of Travel Time



Note: This histogram displays the distribution of travel time for 433 employees in the sample. "Travel Time" is the shortest travel time (in hours) from production center to hometown (one way) via train.

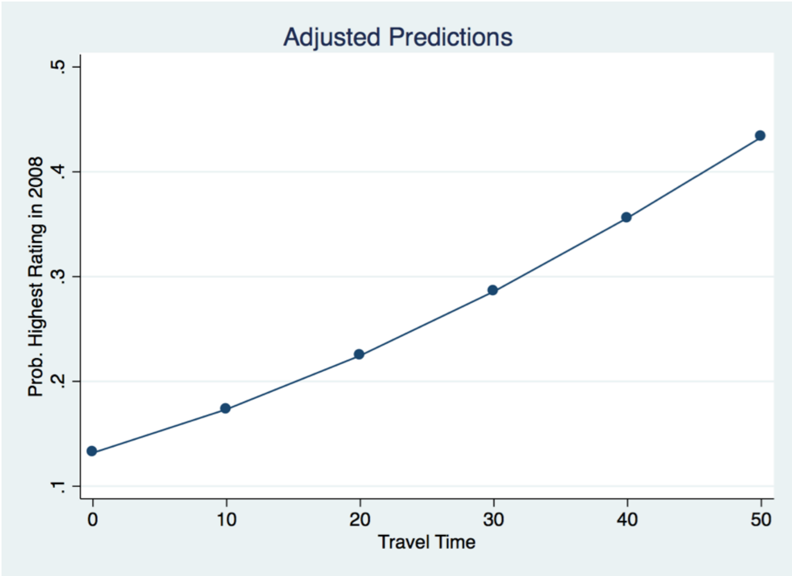
Figure 3
Simulated vs. Realized Value of Travel Time



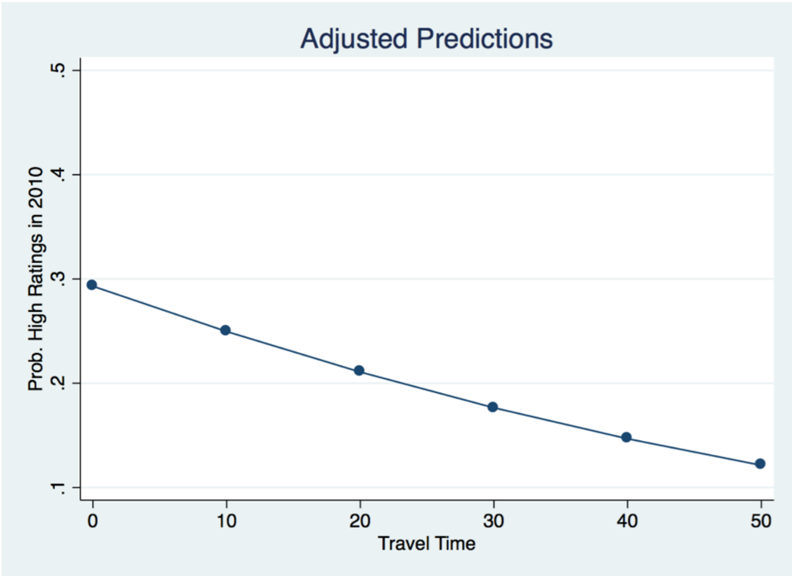
Note: This figure compares the distribution of travel time from Monte Carlo simulation to the realized mean value of travel time. For the simulation, we randomly draw (with replacement) from the entire employee sample the same number of employees actually assigned to one of the eight locations. We conduct 1,000 random draws and present the sampling distribution of mean travel time values in the histogram. The realized mean value of travel time is presented as a thick dotted line. The realized mean value of travel time is not statistically different from a hypothetical mean value of travel time when employee assignment is entirely random, providing additional quantitative evidence that the employee assignment process is random.

Figure 4
Effect of Travel Time on the Likelihood of Receiving High Ratings

(A) Short-Term Performance



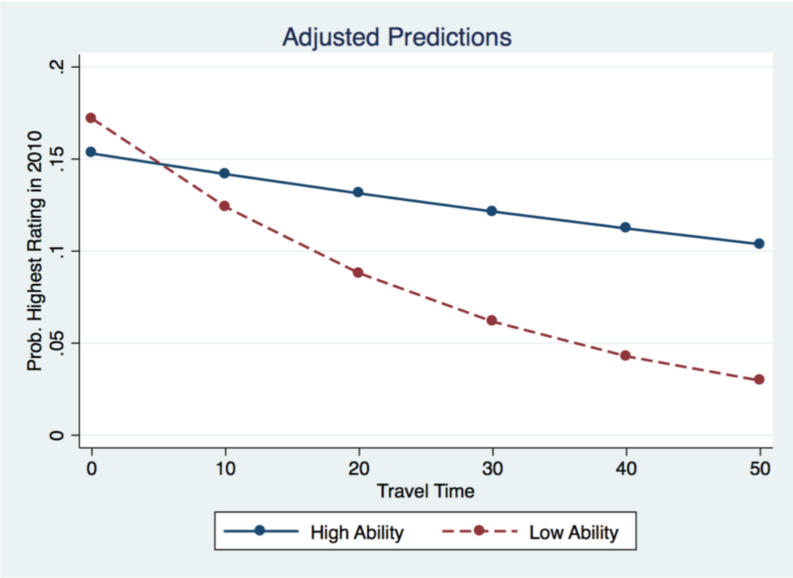
(B) Longer-Term Performance



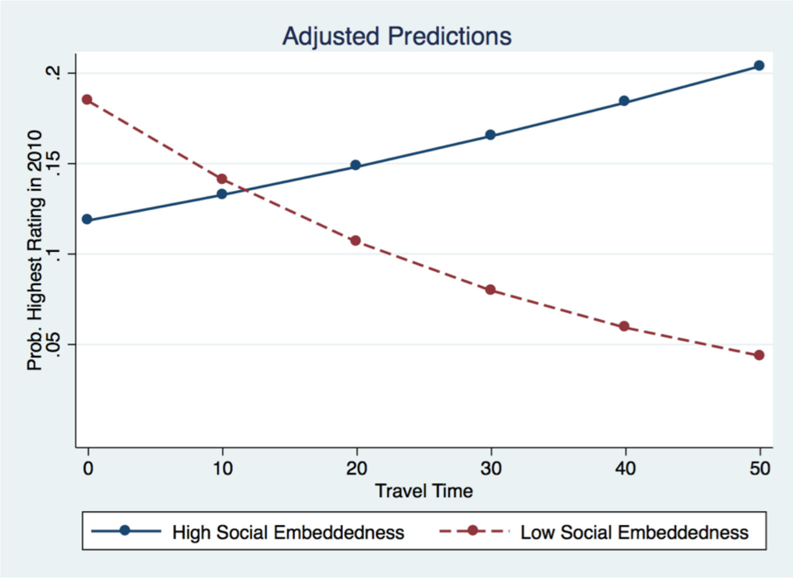
Note: These graphs present the relationship between travel time and the likelihood of receiving high performance ratings in the short and longer-term. We calculate the adjusted predicted values by plugging in different travel time values for an average employee. The top panel shows the likelihood of receiving the highest performance rating in 2008 and indicates a positive relation between travel time and the probability of receiving the highest performance rating. The bottom panel plots the likelihood of receiving the highest or second highest performance rating in 2010 and indicates a negative relation between travel time and the probability of receiving the highest performance rating. Our results remain robust to using the highest performance rating alone to produce a similar graph in 2010.

Figure 5
Heterogeneous Effect of Travel Time on Longer-Term Performance

(A) Effect of Ability



(B) Effect of Social Embeddedness



Note: These graphs present the heterogeneous effect of travel time on longer-term performance. We calculate the adjusted predicted values by plugging in different travel time values for an average employee. The top panel compares the effect of travel time between employees with low/high initial ability, measured using CGPA training scores. The bottom panel compares the effect between employees with low/high embeddedness. Low Social Embeddedness is measured as one if an employee's hometown language is not a majority language among cohorts in the same location, and zero otherwise.

Table 1
Distribution of Production Centers

Location	Freq.	(%)
Bangalore	105	23.7
Hyderabad	72	16.3
Pune	68	15.4
Chennai	67	15.1
Mysore	60	13.5
Bhubaneshwar	50	11.3
Mangalore	15	3.4
Trivandrum	6	1.4
Total	443	100

Note: This table presents the production center assignment results for TECHCO employees in our sample hired in 2007. There are originally nine production centers but we drop the production center of Chandigarh from our sample, as only one employee is assigned to it, making us unable to conduct comparisons among employees at the production-center level.

Table 2
Summary Table

	N	Mean	Std. dev.	Min	Max
Travel Time (in hours, one-way)	443	15.503	10.376	0.333	48.967
Male	443	0.661	0.474	0.000	1.000
CGPA Training	443	4.554	0.332	2.890	5.000
Logical Score	413	5.068	3.287	-4.000	9.000
Verbal Score	413	4.322	3.777	-8.000	15.000
Similar Language	443	0.404	0.491	0.000	1.000
Migration Experience	443	0.632	0.483	0.000	1.000
Low Ability	443	0.503	0.501	0.000	1.000
Low Social Embeddedness	443	0.372	0.484	0.000	1.000
Left the Firm	443	0.282	0.451	0.000	1.000
Performance Rating in 2008	443	-1.722	0.657	-4.000	-1.000
Performance Rating in 2010	385	-2.751	0.901	-5.000	-1.000

Note: We present the transformed performance ratings here, in which the original ratings are multiplied by -1. Therefore, numerically smaller values (e.g., -1) represent higher performance.

Table 3
Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Travel Time	1.000											
(2) Male	0.039	1.000										
(3) CGPA Training	0.043	0.044	1.000									
(4) Logical Score	-0.081	-0.052	0.089	1.000								
(5) Verbal Score	-0.012	-0.004	0.080	0.362***	1.000							
(6) Similar Language	-0.458***	0.031	0.140**	0.108*	-0.068	1.000						
(7) Migration Experience	0.057	-0.127*	0.006	-0.110*	-0.125*	0.092	1.000					
(8) Low Ability	-0.046	-0.085	-0.807***	-0.113*	-0.056	-0.076	-0.032	1.000				
(9) Low Social Embeddedness	0.463***	-0.008	0.024	-0.059	0.090	-0.618***	-0.045	-0.037	1.000			
(10) Left the Firm	0.035	-0.028	0.073	-0.037	0.070	-0.028	-0.035	-0.072	0.064	1.000		
(11) Performance Rating in 2008	0.060	0.059	0.336***	0.168**	0.120*	0.094	-0.216***	-0.290***	0.057	0.028	1.000	
(12) Performance Rating in 2010	-0.075	0.164**	0.212***	0.068	0.114*	0.062	-0.125*	-0.197***	-0.080	-0.204***	0.223***	1.000

*p < .1; **p < .05; ***p < .01.

Table 4
Validity of the Random Assignment

	(1)	(2)	(3)	(4)
	Assigned to Bangalore			
Travel Time to Bangalore	-0.014	-0.011	-0.014	-0.011
	(0.016)	(0.016)	(0.016)	(0.016)
Male	0.139	0.137	0.111	0.107
	(0.240)	(0.246)	(0.241)	(0.247)
CGPA Training	0.030	-0.099	0.029	-0.089
	(0.375)	(0.370)	(0.376)	(0.372)
Logical Score		-0.044		-0.048
		(0.037)		(0.037)
Verbal Score		0.020		0.016
		(0.036)		(0.036)
Migration Experience			-0.323	-0.308
			(0.228)	(0.236)
Observations	443	413	443	413
Log-Likelihood	-241.84	-229.60	-240.85	-228.75

*p < .1; **p < .05; ***p < .01.

Note: Logit regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is a dummy indicating whether an employee is assigned to Bangalore, the largest and most important production center.

Table 5
Distance from Home and Short-Term Performance

	(1)	(2)	(3)	(4)	(5)
			Performance Rating in 2008		
Travel Time	0.017*	0.020*	0.024**	0.021**	0.032**
	(0.010)	(0.011)	(0.012)	(0.010)	(0.013)
Male	0.228	0.199	0.214	0.163	0.097
	(0.215)	(0.223)	(0.216)	(0.219)	(0.227)
CGPA Training	2.365***	2.085***	2.302***	2.454***	2.084***
	(0.376)	(0.379)	(0.384)	(0.396)	(0.407)
Logical Score		0.115***			0.100***
		(0.035)			(0.035)
Verbal Score		-0.000			-0.005
		(0.027)			(0.028)
Similar Language			0.280		0.406
			(0.257)		(0.272)
Migration Experience				-0.889***	-0.912***
				(0.228)	(0.237)
Location FE	Yes	Yes	Yes	Yes	Yes
Observations	443	413	443	443	413
Log-Likelihood	-320.11	-298.83	-319.51	-311.76	-290.73

*p < .1; **p < .05; ***p < .01.

Note: Ordered logit regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is transformed performance ratings in 2008, in which the higher rating represents higher performance. The main independent variable "Travel Time" is the shortest travel time from workplace to hometown via train. Control variables include gender, ability proxies measured during recruiting and training, language similarity between an employee's hometown and workplace location, and prior migration experience.

Table 6
Distance from Home and Longer-Term Performance

	(1)	(2)	(3)	(4)	(5)
			Performance Rating in 2010		
Travel Time	-0.030**	-0.030**	-0.039***	-0.028**	-0.035**
	(0.012)	(0.013)	(0.014)	(0.012)	(0.016)
Male	0.640**	0.603**	0.658**	0.579**	0.557**
	(0.255)	(0.262)	(0.256)	(0.261)	(0.272)
CGPA Training	1.587***	1.306***	1.680***	1.617***	1.410***
	(0.348)	(0.360)	(0.350)	(0.351)	(0.363)
Logical Score		0.002			-0.001
		(0.041)			(0.042)
Verbal Score		0.075**			0.070**
		(0.033)			(0.034)
Similar Language			-0.349		-0.234
			(0.295)		(0.317)
Migration Experience				-0.519**	-0.420
				(0.249)	(0.263)
Location FE	Yes	Yes	Yes	Yes	Yes
Observations	385	358	385	385	358
Log-Likelihood	-247.30	-226.84	-246.56	-245.05	-224.97

*p < .1; **p < .05; ***p < .01.

Note: Ordered logit regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is transformed performance ratings in 2010, in which the higher rating represents higher performance. The main independent variable "Travel Time" is the shortest travel time from workplace to hometown via train. Control variables include gender, ability proxies measured during recruiting and training, language similarity between an employee's hometown and workplace location, and prior migration experience.

Table 7

Heterogeneity of Longer-term Effect based on Employee Ability, Social Embeddedness

	(1)	(2)	(3)
	Performance Rating in 2010		
Travel Time (β_1)	-0.021**	-0.009	0.013
	(0.010)	(0.013)	(0.019)
Travel Time × Low Ability (β_3; H3a)		-0.029*	
		(0.017)	
Travel Time × Low Social Embeddedness (β_3; H3b)			-0.045**
			(0.023)
Male	0.750***	0.754***	0.746***
	(0.218)	(0.218)	(0.220)
CGPA Training	1.501***	1.100**	1.585***
	(0.320)	(0.488)	(0.309)
Low Ability (β_2; H3a)		0.140	
		(0.444)	
Low Social Embeddedness (β_2; H3b)			0.520
			(0.448)
Location FE	Yes	Yes	Yes
Observations	385	385	385
Log-Likelihood	-443.56	-441.94	-441.51

*p < .1; **p < .05; ***p < .01.

Note: Ordered logit regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is the transformed performance ratings measured in 2010, in which the higher rating represents higher performance. "Travel Time" is the shortest travel time (in hours) from workplace to hometown via train. "Low Ability" is measured as one if an employee's CGPA training score is below median, and zero otherwise. "Low Social Embeddedness" is measured as one if an employee's hometown language is not a majority language among cohorts in the same location, and zero otherwise.

Table 8
Extensive Margin: Employee Composition Changes based on Attrition

	(1)	(2)
	Has Left the Firm	
Travel Time	-0.005	-0.001
	(0.014)	(0.016)
Male	-0.308	-0.303
	(0.237)	(0.237)
CGPA Training	0.247	0.506
	(0.357)	(0.389)
Logical Score	-0.074**	-0.061*
	(0.036)	(0.037)
Verbal Score	0.059*	0.058*
	(0.031)	(0.032)
Similar Language	0.215	0.261
	(0.283)	(0.290)
Migration Experience	-0.282	-0.392
	(0.238)	(0.247)
Travel Time × High Performer in 2008		-0.004
		(0.022)
High Performer in 2008		0.528
		(0.425)
Location FE	Yes	Yes
Observations	413	413
Log-Likelihood	-238.29	-235.82

*p < .1; **p < .05; ***p < .01.

Note: Logit regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is a dummy equal to one if an employee leaves the company and zero otherwise.

Appendix

Table A1
Distance from Home and Short-Term Performance, Using OLS

	(1)	(2)	(3)	(4)	(5)
	Performance Rating in 2008				
Travel Time	0.005*	0.005*	0.007**	0.006**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Male	0.031	0.022	0.028	0.016	-0.001
	(0.062)	(0.065)	(0.062)	(0.062)	(0.066)
CGPA Training	0.583***	0.524***	0.561***	0.586***	0.505***
	(0.102)	(0.103)	(0.104)	(0.103)	(0.108)
Logical Score		0.028***			0.024**
		(0.011)			(0.010)
Verbal Score		0.000			-0.001
		(0.008)			(0.007)
Similar Language			0.091		0.108
			(0.077)		(0.082)
Migration Experience				-0.197***	-0.205***
				(0.065)	(0.068)
Location FE	Yes	Yes	Yes	Yes	Yes
Observations	443	413	443	443	413
R-squared	0.10	0.12	0.11	0.12	0.14

*p < .1; **p < .05; ***p < .01.

Note: OLS (ordinary least squares) regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is performance ratings measured in 2008, in which the higher rating represents higher performance. The main independent variable "Travel Time" is the shortest travel time from workplace to hometown via train. Control variables include gender, ability proxies measured during recruiting and training, language similarity between an employee's hometown and workplace location, and prior migration experience.

Table A2

Distance from Home and Longer-Term Performance, Using OLS

	(1)	(2)	(3)	(4)	(5)
			Performance Rating in 2010		
Travel Time	-0.009**	-0.010**	-0.011**	-0.008*	-0.009
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)
Male	0.316***	0.288***	0.319***	0.289***	0.255***
	(0.088)	(0.091)	(0.089)	(0.090)	(0.094)
CGPA Training	0.652***	0.565***	0.670***	0.658***	0.573***
	(0.131)	(0.132)	(0.132)	(0.130)	(0.132)
Logical Score		0.000			-0.003
		(0.015)			(0.016)
Verbal Score		0.023*			0.021
		(0.013)			(0.013)
Similar Language			-0.071		0.016
			(0.113)		(0.119)
Migration Experience				-0.228**	-0.222**
				(0.096)	(0.103)
Location FE	Yes	Yes	Yes	Yes	Yes
Observations	385	358	385	385	358
R-squared	0.11	0.11	0.11	0.13	0.12

*p < .1; **p < .05; ***p < .01.

Note: OLS (ordinary least squares) regression is used for estimation, and robust standard errors are presented in parentheses. The dependent variable is performance ratings measured in 2010, in which the higher rating represents higher performance. The main independent variable "Travel Time" is the shortest travel time from workplace to hometown via train. Control variables include gender, ability proxies measured during recruiting and training, language similarity between an employee's hometown and workplace location, and prior migration experience.

Table A3

Comparison among Employees with and without 2008 Rating

	With Rating (N = 443)	Without Rating (N = 633)	Diff.	(p-value)
Male	0.661	0.679	-0.019	0.522
Verbal Score	4.333	4.324	0.009	0.971
Logical Score	5.070	5.164	-0.094	0.654

Note: This table compares 2007 intakes with and without the first-year performance rating in 2008. We use three individual-level characteristics that are pre-determined before being assigned to training batches: gender, and logical and verbal scores in the recruitment test. The fourth column reports whether the differences among employees with and without performance rating in 2008 are significantly different from zero. The *p*-values indicate that we cannot reject the null hypothesis that our sample employees and the remaining 2007 intakes are from the same population. *, **, and *** denote significance at 10%, 5%, and 1%, respectively.

Table A4
Comparison among Employees with and without 2010 Performance Rating

	With Rating (N = 385)	Without Rating & Not left the firm (N = 7)	Diff.	(p-value)
Male	0.675	0.429	0.247	0.169
Verbal Score	4.271	5.500	-1.229	0.431
Logical Score	5.198	2.333	2.865	0.032
CGPA Training	4.556	4.617	-0.061	0.629
Performance Rating in 2008	-1.613	-1.857	0.244	0.203

Note: This table compares employees with performance rating in 2010 to employees without performance rating in 2010 but who do not leave the firm. We compare them across five individual-level characteristics: gender, and logical and verbal scores in the recruitment test, CGPA training, and performance rating in 2008. The fourth column reports whether the differences between the two groups are significantly different from zero. *, **, and *** denote significance at 10%, 5%, and 1%, respectively.