Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?

JOB MARKET PAPER

For latest version of paper, click here [*]

Devaki Ghose

February 17, 2021

Abstract

How do trade shocks affect welfare and inequality when human capital is endogenous? Using an external IT demand shock and detailed internal migration data from India, I first document that both IT employment and engineering enrollment responded to the rise in IT exports, with IT employment responding more when nearby regions have higher college age population. I then develop a quantitative spatial equilibrium model featuring two new channels: higher education choice and differential costs of migrating for college and work. Using the framework, I quantify the aggregate and distributional effects of the IT boom, and perform counterfactuals. Without endogenous education, estimated aggregate welfare gain from the export shock would have been half and regional inequality about a third higher. Reducing barriers to mobility for education, such as reducing in-state quotas for students at higher education institutes, would substantially reduce inequality in the gains from the IT boom across districts.

JEL Classification: F16, F63, I24, J24, R12
Keywords: trade, human capital, inequality, migration, gravity, education

1   Introduction

New economic opportunities that arise from globalization are often accompanied by a rising demand for different types of skills. Inequalities in local access to education and jobs, along with mobility frictions, make it costly for individuals in some regions to acquire education or pursue better job opportunities. These frictions could be particularly large in developing countries. To what extent do these frictions limit the gains from trade and exacerbate inequality? What policies can help reduce these inequalities? The main challenge in answering these questions is disentangling the different ways in which individuals respond to these opportunities, such as choosing the sector and the locations of work and education, and the interdependence between these decisions.

In this paper, I analyze the effects of trade on welfare and inequality when education choice is endogenous and when there are mobility frictions to access both education and work. Combining detailed spatial and migration data, I document that IT employment and engineering enrollment responded to the rise in Indian IT exports in the late 1990s, and this response was heterogeneous across regions. Consistent with these stylized facts, I develop and quantify a spatial equilibrium model that adds two new margins of response relative to the existing economic geography literature: first, agents can acquire new skills and second, they can migrate internally to acquire these skills. I find that without higher education choice, estimated aggregate welfare gains from the IT boom would be halved and estimated regional inequality would be a third higher. Restricting individuals to go to college only in their home districts (i.e., not allowing for mobility for education), reduces the estimated aggregate welfare gains marginally but increases regional inequality by 15%.

The paper begins by providing a set of stylized facts about the labor market consequences of the IT boom using spatially granular sectoral labor and education data that I compiled and a unique census dataset tracking migration flows between Indian districts, disaggregated by reason for migration. From 1998 to 2008, while Indian IT as a fraction of total service exports increased from 15% to 40%, engineering enrollment as a fraction of total enrollment more than doubled, and total college enrollment increased three-fold. I document two salient stylized facts: 1) IT employment and engineering enrollment positively respond to IT exports, with IT employment responding more when nearby regions have higher engineering enrollment and exports; and 2) distance affects migration, and individuals migrate more for work than for education. State borders restrict migration flows for education more than that for work, reflecting state-level barriers to mobility for education, such as in-state quotas for students at higher education institutes.

Consistent with these stylized facts, I develop a quantitative spatial equilibrium model that allows individuals to make education and work decisions in two stages. In the

first stage, they decide what and where to study, accounting for access to higher education and job opportunities. In the second stage, individuals choose the sector and location of work. The first and second stage decisions generate the education and employment responses respectively, documented by stylized fact 1. To my knowledge, this is the first paper to allow for and estimate differential mobility costs for work and education. The estimated mobility costs are much higher for education than for work, consistent with stylized fact 2.

In the model, sector specific trade shocks, such as the Indian IT boom, change the relative returns to occupations across locations depending on two factors: 1) the location's comparative advantage in that sector and 2) the location's connectivity to other locations. The changes in the relative returns to occupations affect an individual's incentives to invest in different skill types. Skill investments are constrained by the local availability of higher education and the costs of moving to regions with colleges. Thus, regions differ in how much skilled labor they can access and consequently, by how much they can expand IT production. To the extent that the external demand shock and historical regional differences in comparative advantage are not correlated with unobserved productivities that are determined from the supply side, I can leverage the IT boom to estimate the structural model parameters, such as the elasticity of software exports to software prices. Differences in local access to jobs and education, along with differential moving costs for work and education, generate regional inequalities in the welfare gains from the IT boom.

People face differential migration costs when they move for education or for work. The dependence of job opportunities on skill levels makes it challenging to separately estimate work mobility costs using migration data that do not track the skill level of the migrant. I use the two stage structure of the model to explicitly account for such dependence and use the unique census data that track why people move, to estimate these two costs separately.[1] I find that the mobility costs across districts, measured as the dis-utility from moving, for education are 7 percentage points higher than those for work. Estimated state border effects are large: being in the same state increases migration between neighboring districts by 269% for education and 59% for work. There are several reasons why the mobility cost of education could differ from that of work, such as policy-induced mobility barriers and language barriers that could have a differential effect depending on an individual's age. In India (as in many other countries like the U.S. and China), there are state quotas in higher education institutes for in-state students. This policy could result in higher costs of crossing state borders for education than for work.

---

[1] To my knowledge, this is only the second project to use this highly confidential data and the first that uses it to estimate migration costs. Kone et al. (2018) were the first to use this data to show how migration flows relate to geographic and cultural distances in India

Compared to a benchmark quantitative model with fixed skill types, I find significantly different aggregate and distributional consequences of trade across regions after incorporating the mechanism of endogenous education choice. Almost half of the gains in average welfare are driven by the ability to change skills: without endogenous education, the average individual would have benefited by 0.67% compared with an average gain of 1.12% in the endogenous education case. Even though inequality in the distribution of employment is reduced, the rise in inter-regional welfare inequality due to the IT boom, measured as the coefficient of variation in regional welfare at the origin district, is 37% more in the fixed-skills model than in the model with endogenous education choice. The key mechanisms leading to higher aggregate welfare and lower welfare inequality in the endogenous education model, compared with the fixed-skills model are the ability to acquire skills and to move across regions for education.

How important is this mobility cost for education that I introduce in a spatial equilibrium model? To quantify the importance of mobility costs for education, I restrict individuals to attend college in their home districts. This counterfactual increases regional inequality by 15%. The gap between the welfare gains of the worst off and the best district increases by 63%.

The question of how to reduce inequality across both regions and skill groups lies at the heart of many policy debates. This paper suggests policy interventions in the education market that can reduce trade-driven regional inequality, but not by moving jobs directly. The policy of reducing in-state quotas for students at colleges can reduce the migration costs for education relative to work. In the model, this is implemented by restricting the effect of state borders on migration for education to be exactly the same as that for work. I find that the rise in average welfare would have been 1% higher compared with the rise in the model without these restrictions on work and education costs, and regional inequality, measured by the coefficient of variation, would have been 27% lower. Reducing inter-state barriers to education can significantly increase access to education for out-of-state students. This can increase the opportunity for people from remote regions to gain access to education and migrate to areas with more high-skilled jobs. Although this policy does not reduce inequality in the distribution of employment, it reduces inequality in the distribution of welfare by increasing access to education. This underscores the importance of general equilibrium effects induced by the expansion of exports, which requires us to give more consideration to the interactions of trade, education, and labor markets.

This paper makes three contributions. First, I introduce human capital acquisition decisions in a general equilibrium economic geography model. The general equilibrium aspect is important, since human capital takes time to respond to employment opportunities, during which both people and goods can move. Second, to my knowledge, this is

the first paper to estimate the mobility costs for work and education separately and show that these costs are quantitatively different. The unique Census data tracking migration flows disaggregated by reason, obtained through an agreement with the Indian government, made this estimation possible. I show that access to both jobs and education are individually important for determining the spatial dispersion in the gains from trade. Third, the framework is well-suited for analyzing the effects of policy-induced spatial frictions to moving for higher education, such as in-state quotas at colleges. Reducing these barriers would increase aggregate welfare marginally but substantially decrease the impact of the export shock on regional inequality. The results underscore the potential for education policies to distribute the gains from globalization more equally.

The model builds on a large theoretical literature in the fields of international trade, economic geography, labor, and migration. Similarly to Caliendo et al. (2019), Fuchs (2018), and Kucheryavyy et al. (2016) the model features multiple sectors. Like Allen et al. (2018) and Tsivanidis (2018), the model features agents with heterogeneous skill types. However, unlike the above models, the theory developed here endogenizes the formation of skills across space.

Costly labor mobility relates this paper to the class of gravity migration models, such as those by Allen et al. (2018), Tombe and Zhu (2019), Fan (2019), and Bryan and Morten (2019) that feature multiple sectors or regions with costly mobility of goods and people. Kone et al. (2018) use the Indian migration data to provide evidence of how migrations flows relate to distance and cultural differences. Imbert and Papp (2020) provides evidence that the seasonal cost of migration from rural to urban India is very high such that persistent wage differences between the rural and urban sectors can persist. Differently from these papers, I provide separate estimates for mobility costs by reasons for migration.

A few structural trade models study endogenous human capital acquisition in trade. Khanna and Morales (2017) studies how US immigration policy and the internet boom affected aggregate welfare in both the US and India in a dynamic setting with international migration. In contrast, this paper studies the regional distributional consequences of the IT boom, quantifying how costs of migration contributed to regional inequality induced by the IT boom. Compared to Ferriere et al. (2018), who build a dynamic multi-region model of international trade with heterogeneous households, incomplete credit markets, and costly endogenous skill acquisition, this paper, in a static setting, additionally features costly mobility for education. A few other theoretical works in this literature focus on quantifying the overall response of endogenous education to trade, without considering regional differences, such as Danziger (2017). Seminal works on the dynamic Heckscher-Ohlin (HO) model, which embed endogenous factor formation in response to trade in the classic HO framework, include Stiglitz (1970), Findlay and Kierzkowski (1983), and

Borsook (1987). Consistent with this literature, I demonstrate that trade can strengthen a country's initial comparative advantage by changing the incentives to acquire skills, and thereby reduce regional inequality in the gains from trade.

Endogenizing education relates my model to the class of human capital accumulation models prominent in the education and labor literature. In these models, forward-looking individuals make education decisions based on labor market returns and costs of tuitions (Jones and Kellogg (2014), Johnson (2013), and Lee (2005)). Compared to this class of models which requires keeping track of a large number of state spaces, I use a simpler two-stage model that allows me to tractably incorporate many regions and bilateral migration flows between these regions.

Given the emphasis in the trade literature on the effect of exports and trade liberalization on skill premium, there has been relatively little research on the effect of trade on skill acquisition. A number of empirical studies such as Atkin (2016), Blanchard and Olney (2017), Edmonds et al. (2010), Greenland and Lopresti (2016), Shastry (2012), Liu (2017), and Oster and Steinberg (2013) focus on the impact of trade on primary and secondary education. Exceptions to these are Li (2018) and Khanna and Morales (2017) which study the response of college enrollment to high-skill export shocks. More evidence has emerged recently (Li (2019), Hou and Karayalcin (2019), Ma et al. (2019)). Complementing this literature, I provide reduced form evidence about the response of tertiary enrollment to shocks in the high-tech sector in a large developing country, and I document regional heterogeneity in this response.

2 Data

A major constraint in studying the effects of IT export growth on human capital acquisition in the presence of costly migration is the lack of employment and education data, disaggregated at the sector of work and field of education level, combined with the absence of detailed migration data. To this end, I use three sources to collect data on India's IT sector and access confidential Indian Census data to obtain district-to-district migration flows by reasons for migration.

2.1 Data on the Indian IT sector

I use three rounds of Economic Census data (1998, 2005, and 2013) to obtain data on total IT employment across all districts of India. While the advantage of the Census data is that it covers the entirety of all Indian firms and hence reports total employment, the data are not disaggregated by level of education. To supplement this information,

I use data from the National Sample Survey (NSS) rounds 50, 55, 60, 61, 62, 64, 66, and 68. These surveys record information on the sector and location of occupation as well as the field of study. The drawback of the NSS data is that it represents only a small sample, and hence does not contain a lot of important sector-level information. However, it does report multipliers on each unit of observation, which, in the NSS, is an individual. This allows me to obtain the unbiased ratios of engineers, non-engineers, and both college-educated and non-college-educated individuals in each sector of employment. By multiplying these ratios with total employment from the Economic Census, one can recover the distribution of the population by field of study and sector of employment in each district of India. Data on wages by sector of occupation and field of education are also obtained from the NSS, supplemented with data from the Economic Census. For more details on combining NSS with Economic Census, see appendix A.3 for further details. Tables VII and VIII report the daily average district-level raw wages in Rs. and the average district-level employment respectively. From these tables, observe that both the wages and employment of college-educated and non college-educated workers increased more in high-skill intensive industries between the pre and post boom periods compared to those in manufacturing.

As an additional source, for the reduced form analysis, I supplement the IT employment and wage data with data on IT exports from NASSCOM (the leading trade association of the software industry in India) directories 1992, 1995, 1998, 1999, 2002, and 2003. The strength of the NASSCOM dataset is that it contains data on "95% of all registered IT firms in India" [2]. NASSCOM also contains data on IT employment, and this information is divided according to whether employees are technical employees (that is, associated directly with the provision and deliverance of IT services) or non-technical employees (all other employees). Several papers have used the NASSCOM data, which is the most comprehensive source of data on Indian IT firms; among these, Tharakan et al. (2005) and Shastry (2012) are notable.[3]

Throughout the paper, 1995-1999 is referred to as the pre-boom period and 2001-2011 is referred to as the post-boom period. The relatively longer choice for the post boom period is based on the fact that it takes at-least 2 or 3 years to prepare for college and at-least 4 years to complete a college degree. Thus, the effect of the IT boom on enrollment and graduation will be observed with a lag.

---

[2]Source: NASSCOM

[3]Tharakan et al. (2005) cross-checked the quality of NASSCOM software exports data by comparing yearly aggregate software exports from India from International Data Corporation with the figures obtained from NASSCOM data and claim that they are of comparable magnitude.

2.2 Data on internal migration

The National Census of India for 2001 is the main data source for internal migration in India. An individual is a migrant, according to the Census,"if the place in which he is enumerated during the census is other than his place of immediate last residence" (Census, 2001). The Census includes additional questions based on the last residence criteria. These questions include reason for migration, such as marriage, education, or employment; the urban/rural status of the last residence's location; and the duration of stay in the current residence since migration. This level of disaggregation is crucial for separately estimating the costs of migration due to education and work. Publicly available Census data only report the destination district and whether the migrant's origin is in the same state or out-side the state, aggregated over all reasons for migration. I obtain the more disaggregated data through a special agreement with the Census of India. More information on this data can be found in Section 4 and descriptive evidence about the proportion of people migrating for work and education can be found in Table II.

2.3 Other data

Data on the linguistic distance of each Indian district from Hindi was obtained from Gauri Kartini Shastry who constructed the linguistic distance measures for Shastry (2012). Construction of the index, which is key to my empirical strategy, is detailed in Shastry's paper and in appendix **??** of this paper. The linguistic distance I use is calculated by ethno-linguistics based on the similarity of grammar and cognates. For example, daughter in English is "dokhtar" in Perisan and "nuer" in Mandarin Chinese. While Persian and English are both part of the Indo-European language family, Chinese is derived from the Sino-Tibetan language family. Linguistic distance between Persian and English is therefore lower than between Chinese and English or Chinese and Persian. In India, languages differ across regions. The 1961 Census of India documented speakers of 1652 languages from five language families. There can be wide linguistic diversity between districts, and most people adopt a second language that is a widely accepted speaking medium across districts. Of all multilingual people who were not native speakers, 60 percent chose to learn Hindi and 56 percent chose English (Shastry (2012)).

Shastry (2012) proxies English-learning costs as linguistic distance from Hindi relative to English. She shows that since a necessary condition for employment in the IT industry is fluency in English, IT firms locate more in districts that have a higher proportion of English speakers, as proxied by linguistic distance of that district to English relative to Hindi.

Data on the college-age population, college enrollment, and literacy are collected

from the decadal Census data of 2001 and 2011. Summary statistics for enrollment are reported in Table IX in the appendix. The most notable is the rise in engineering enrollment. Between the pre and the post boom period, the proportion of engineers in total college enrollment more than doubled from 5% to 11%. During this time, the total number of people enrolled in college also increased by three-fold. Thus, the total number of students studying engineering also increased in absolute numbers.

3   Background of India's IT growth

While the last two decades have witnessed a world-wide expansion of IT and consequent increase in demand for computing skills, this expansion has been disproportionately larger for India than for any other country in the world (International Trade Center (2017)). Panel A of Figure IV plots the growth in IT exports over time, where the value of IT exports in 1993 has been normalized to one. This figure shows that IT exports from India have been steadily increasing since 1993, but a large jump occurred in the late 1990s and early 2000, when normalized software exports increased by more than 76% in one year. Figure IV shows that during this period, IT employment as a fraction of total employment was also rising. From 1998 to 2000, IT employment as a fraction of total employment almost doubled. Engineering as a fraction of total enrollment was also generally increasing, but the largest jump occurred after 2000.

While many factors are responsible for the growth of IT in India, the lack of domestic demand for IT means that the sector's growth is constrained by the growth in world demand for Indian IT. This constraint was eased during the late 1990s and early 2000s, when several major events suddenly escalated demand for Indian IT. The Y2K phenomenon dominated from 1998 to 2000, along with the earlier dot-com boom and, later on, the dot-com bust. In order to solve Y2K-related computer problems, commonly known as the "Y2K bugs", IT firms started offshoring large parts of their work to developing countries such as India.[4] The dot-com boom was a historic economic bubble and period of excessive speculation that occurred from roughly 1995 to 2000; it was marked by extreme growth in the use and adaptation of the Internet. The dot-com bust caused many firms in the US (two-thirds of India's IT market) and elsewhere to slash their IT budgets, prompting even more outsourcing to India. (Economist (2003))

Most notably, technological progress in the worldwide Internet which had been un-

---

[4]Before 2000, all computers stored dates using only the last two digits of a year. The Y2K problem refers to the problem that can occur in computer systems as the year 1900 becomes indistinguishable from 2000. The majority of programs with Y2K problems were business applications written in a 40-year-old language called COBOL (UC Berkeley (1999)). While COBOL programming was already obsolete in US universities, in India it was still a part of the regular course curriculum. (Mathur (2006)).

derway for some time, was responsible for bringing world outsourcing demand to Indian firms. As Khanna and Morales (2017) notes:

> The absence of world-wide Internet during the 1980s meant that on-site work ("body-shopping") dominated, because otherwise software had to be transported on tapes that faced heavy import duties. But in 1992, satellite links were set up in Software Technology Parks (STP), negating the need for some kinds of on-site work, and this boosted the offshoring of work to India. In 1993, the shift from B-1 to H-1 visas in the US further lowered the incentives to hire Indian engineers for on-site work, as they were to be paid the prevailing market wage.

While world-wide events such as the Y2K shock, the dot-com boom and bust, and changes in US H-1B visa policies provided considerable external demand stimuli for the growth of the Indian IT sector, certain factors inherent to India are responsible for this expansion of Indian IT exports. It is generally agreed that the availability of low-cost, high-skill human resources has given India a comparative advantage in the IT sector over its competitor nations (Kapur (2002)). Moreover, much of the population (over 60%) is under 25, and India has one of the largest pools of technical graduates in the world. India also has a large English-speaking population due to its British legacy, and this fact is considered one of the key ingredients in the success of IT. As Shastry (2012) has shown, IT firms in India are located mostly in regions with a larger English-speaking population. A natural advantage of India is its time difference with the US, which is one of India's biggest customers for IT services; this enables India to offer overnight services to the US, effectively creating round-the-clock working hours for outsourcing firms (Carmel and Tjia (2005)).

The growth of Indian IT is the result of much more than a single transitory demand shock that temporarily catapulted the sector upward. With the expansion in Indian IT exports, Indian IT employment continued to increase. Wages peaked during the sudden expansion of the late 1990s and early 2000s. Arguably, in response to rising IT employment opportunities, engineering enrollment started to respond after 2000, as shown in Figure ??.

4   Reduced-form facts

In this section, I present four facts about internal migration, the relationships between IT exports, regional employment, and enrollment over the short-run and the long-run in India. I use the expansion of IT during 1998-2002, largely driven by external demand

shocks as described in Section 1.3, to study the labor market effects in the long-run, that is, between 2005-2011. The choice of this time frame is dictated by the fact that an engineering degree takes at-least four years to complete and thus any effect on the labor market related to skill acquisition will occur after 2004-2005.

Fact 1: *IT employment and engineering enrollment positively respond to exports.*

To understand how IT employment and engineering enrollment changed across regions after the IT boom, I estimate the following event study specification:

$$Y_{dt} = \alpha_t + \gamma_d + \chi_d * t + \beta_t Exports_{d,1995} + \epsilon_{dt} \tag{1}$$

where Y_{dt} is standardized IT employment or standardized engineering enrollment in district d at time t. $Exports_{d,1995}$ is the proportion of software exports from district d in the year 1995 out of total Indian IT exports in 1995. α_t are time fixed effects that capture any factors that are common to all districts at time t. γ_d are district fixed effects that capture any factors that are fixed over-time in district d. $\chi_d * t$ is a district-level time-trend capturing any linear trend in the outcome variable at the district level. Standard errors are clustered at the state-year level, with alternative clustering assumptions at the state and district levels explored in Appendix B, Table XI.

The idea is that districts which initially had higher connections with the rest of the world, as measured by the proportion of software exports in 1995, will gain more from the expansion in world demand for Indian IT than districts that had little or no connection with the rest of the world. In alternative specifications reported in Online Appendix B, following Shastry (2012), I instrument the initial software exports with the historical linguistic distance of a district from English.

In panel A of Figure II, I plot the estimated coefficients along with the confidence intervals for the years 1995-2013. From this figure, we can see that post-1998, IT employment increased more in districts that had a higher level of software exports in 1995. This effect is significant in all years available in the data from 1999-2013. The insignificant coefficients for 1995 and 1998, the pre-boom years, indicates the absence of pre-trends.

In panel B of Figure II, I plot the response of engineering enrollment at ten year intervals, as the available census data allows. As the graph shows, engineering enrollment has also been rising since 2001. Since the Census data is available at decadal intervals, I cannot show the pre-trend estimates for enrollment.

Fact 2: *The effects are heterogeneous. Employment responds more when nearby regions have higher engineering enrollment and higher IT exports. The heterogeneous effects are stronger in the long run.*

In equation 2 below, I add an interaction term between the number of students enrolled in engineering in 1991 and the proportion of software exports from district d in 1995. Estimated coefficient δ_t is plotted in Figure III. δ_t measures the differential response of IT employment between the pre and post boom periods depending on the historical level of engineering college enrollment in 1991, in districts that already had prior software exports in 1995.

$$Y_{dt} = \alpha_t + \gamma_d + \beta_t * Exports_{d,1995} + \chi_t Enrollment_{d,1991} + \\ \delta_t Exports_{d,1995} * Enrollment_{d,1991} + \epsilon_{dt} \tag{2}$$

Figure III shows that, conditional on the level of software exports, post 1998, IT employment responds more in districts that, in 1991, had more enrolled engineering students in same-state, nearby districts. The intuition, formalized in the model, is that in these districts, it is easier to expand future IT production due to having access to more college-educated, engineering program graduates in close proximity. Regression results are reported in appendix B in Table XIII.

Both IT employment and engineering enrollment thus respond more in districts that had prior IT exports compared to districts that did not. Figure IV shows the spatial distribution of IT employment as a proportion of total employment and engineering enrollment as a proportion of total enrollment in 2011. The graph shows that there is a positive correlation between the percentage of people employed in the IT sector and the percentage of people enrolled in engineering at the district-level. While the contemporaneous correlation in 2011 is 0.38, the corresponding correlation between the proportion of IT employment in 2005 and the proportion of engineering enrollment in 2011 is 0.43. Districts in the south have a higher proportion of both, districts in the north and northeast are relatively deprived of both.

I next establish a set of facts related to the costs of migration over distance for both work and for education.

Fact 3: *Migration reduces over distance. In addition, state borders negatively affect migration and this effect is significantly larger when people migrate for education than when they migrate for work or for any other reason.*

Using the Poisson pseudo maximum likelihood procedure (PPML), I estimate (3), similar to Kone et al. (2018). [5] PPML is a non-linear estimation procedure which performs better than a log-log estimation in the presence of zeros and has been traditionally used in the estimation of migration gravity equations (Santos Silva and Tenreyro (2006)).

---

[5]Kone et al. (2018) ran this specification for 585* 584 districts, excluding the own district. Following the literature on gravity estimation, for e.g., see Bryan and Morten (2019)), I estimate it on a 585*585 sample, including own district.

$$l_{oj} = C + f_j + f_o + \beta_1 ln(Dist_{oj}) + \beta_2 lang_{oj} +$$
$$\gamma_1 Diff_{oj}^{diff-NBR} + \gamma_2 D_{oj}^{same-NBR} + \gamma_3 D_{oj}^{same-notNBR} + \epsilon_{oj} \qquad (3)$$

where l_{oj} is the stock of migrants migrating from district o to district j for education (column 1 in Table I), for work (column 2) or for other reasons (column 3). $Dist_{oj}$ is a measure of geographic distance between two districts.[6] For bilateral distance between any two districts, I use the geodesic (flight) distance between the geographic centers of districts i and j. All these variables included in the gravity specification are obtained from the calculations by Kone et al. (2018). [7]

$lang_{oj}$ denotes the likelihood of any two individuals from districts i and j being able to communicate in a common language. This is given by:

$$CommonLanguage = \sum_l s_i^l . s_j^l$$

where s_i^l is the share of people from district i having mother tongue l.

There are three contiguity variables: $diff - NBR_{ij}$ is a dummy variable that takes the value 1 if districts i and j are in different states but are neighbors; $same - NBR_{ij}$ is a dummy variable that is equal to 1 if the districts i and j are in the same state and are neighbors; $same - notNBR_{ij}$ is a dummy variable that is equal to 1 if the districts i and j are in the same state but are not neighbors. The base group is 'not in the same state and not neighbors'. The difference between γ_1 and γ_2 gauges the role of the state borders.

Table I shows that the coefficient for same-state-neighbor dummy is larger than the different-state-neighbor coefficient in every column, and this difference is statistically significant. For two districts that are neighbors, being in the same state iis associated with an increase in migration by 217% when people migrate for education.[8] This effect is about 30% when people migrate for work and for other reasons.

This shows that the effect of state borders differ substantially depending on the reason for migration. One reason why the state border dummy is so important when people migrate for education is the policy of reserving a large proportion of seats in public as well as private colleges for in-state students. Most state colleges have home state quotas of 50 % with such limits being as high as 85% for some states.[9] While such

---

[6]Other reasons include marriage, business and other unclassified reasons

[7]Geodesic distance is the length of the shortest curve between two points along the surface of a mathematical model of the earth—between the districts' geographical centers, denoted as distance centroids.

[8]This effect is calculated by $(e^{3.577-2.422} - 1) * 100$

[9]Support for the 85% reservation policy started in Maharashtra from the year 2011 with the backing

quotas also exist for jobs and thus create significant hurdles for moving across states, the employment quotas are more specific and less ubiquitous than the in-state education quotas.

Fact 4: *Individuals migrate more for work than for education and the distributions of flows for work and for migration across districts differ accordingly.*

Figure V shows the histogram of migration flows by reason for migration. The x-axis plots the percentage of people who migrated for work and for education out of the total number of migrants at the destination district. The y-axis plots the number of destination districts with the corresponding percentages. As is clear from the plots of these very different and almost non-overlapping distributions, out of the total migrant population in most destination districts, a much higher percentage had migrated for work compared to that for education.

Facts 3 and 4 are also borne out by Table II below. Reading off column 3, out of all migrants who migrated out of their district of past residence in the last 10 years, 48% migrated for work, but only 3% migrated for education. Column 6 shows that out of all individuals who migrated for education, only 31% crossed state borders, while over half of those migrating for work did so.

Informed by these four facts, the next section presents a general equilibrium model featuring many locations, costs of movement of people and goods between locations, and costly human capital acquisition decisions.

5   A Quantitative spatial equilibrium model with endogenous education choice

There are discrete locations $d \in D$ where D includes the many regions within a country, in this case, districts within India. There is also the Rest of the World (RoW) that these many regions trade with. The small open economy assumption holds. The regions differ from each other in their distances to other regions and the RoW and in the distribution of population eligible to attend college, that is, individuals who have already completed high school. There are individuals in each region who make decisions in two stages. In the first stage, they decide whether or not to go to college, and if they go to college, what field to study and in what location. There are F fields individuals can choose to study, such as engineering. In the second stage, given their education decisions, individuals decide where and in which sector to work. There is a representative firm in each sector in each location, and within each sector the firm in each location produces a different

_____

of nationalist state parties

variety which is costly to trade across locations, as in an Armington set up. Each worker is endowed with an unit of labor which they supply inelastically. There are S sectors in the economy.

Individuals

Utility of an individual i who attained college education in field f from region o_2 and then works in sector S in region d depends on wages, amenities, migration costs, price indices, and idiosyncratic productivity shocks, is given by: (suppressing individual subscript i from utility for expositional clarity)

$$V_{o_2 f, dS} = \left(\frac{w_{f, dS}}{P_d} \right) \cdot u_{f, dS} \cdot \eta_i \cdot \mu_{o_2 d}^2 \qquad (4)$$

where $w_{f, dS}$ is the wage of a worker in region d with a degree in field f who is working in sector S, $u_{f, dS}$ is the amenity of living in region d for a worker with degree f working in S, henceforth referred to as type (f, S) worker. P_d is the cost of living in region d, which is endogenously determined as described in Section 5. $(1 - \mu_{o_2 d}^2)$ is the utility cost of migrating from o to work in d.

The idiosyncratic productivity shocks for each individual i, $\eta_{i o_2 f, dS}$ are drawn from a Frechet distribution where

$$F(\eta_{i o_2 f, dS}) = exp(-\eta_{i o_2 f, dS}{}^{-\theta})$$

θ determines the dispersion of the Frechet productivity shocks.

Utility cost of education

The above formation of utility ignores the utility cost of education. To add workers' education choice, I introduce an utility cost of education. Let $a_{o_2 f}$ denote the amenity of studying f in o_2, which includes the unobserved preferences for studying f in o_2 and the time and money cost of education. In other words, it is the fraction of utility lost in order to study field f in region o_2. People who choose not to go to school earn income $w_{u, dS}$ and people who go to school earn a normalized stipend 1. Let ζ_{iof} denote the idiosyncratic preference shock of individual i for his field of choice f in location o_2, where

$$G(\zeta_{i o_2 f}) = exp(-\zeta_{i o_2 f}^{-\gamma})$$

γ again determines the dispersion of amenities of studying f in o_2. There is also a migration cost incurred due to moving from one's location of birth o_1 to one's location of study o_2 denoted by $(1 - \mu_{o_1 o_2}^1)$. Thus utility of an individual i born in o_1 who chooses to study field f in location o_2 and then decides to work in sector S in region d is given

by: (suppressing individual subscript i in utility for expositional clarity)

$$U_{o_1 o_2 f, dS} = \left(\mu_{o_1 o_2}^1 \cdot \frac{a_{o_2 f} w_{u,o_2 S}}{P_{o_2}} \cdot \zeta_i \right)^{IU} \left(\left(\frac{w_{f,dS}}{P_d} \right) \cdot u_{f,dS} \cdot \eta_i \cdot \mu_{o_2 d}^2 \right)^{(1-IU)} \tag{5}$$

where IU is the weight placed on period 1 utility.[10] w_{u,o_2} is the wage earned by unskilled workers in stage 1 in region o_2. w_{u,o_2} is 1 if the person is not employed in stage 1. In other words, people who are not working in stage 1 earn a normalized stipend of just one. Derivation of this utility is given in Online Appendix C.1.

Migration decisions for education and work

When choosing the location and field of education in stage 1, the individual takes into account his expected utility from stage 2. She does not know the exact utility in stage 2 since the idiosyncratic productivity shock is not yet observed. We thus solve the individual's problem backwards. In stage 2, given the choice of location and field of education (sector of work for an unskilled person), the individual makes his choice of sector of occupation (S) and location (d), given by:

$$argMax_{d,S} \left(\left(\frac{w_{f,dS}}{P_d} \right) \cdot u_{f,dS} \cdot \eta_i \cdot \mu_{o_2 d}^2 \right) | o_1, o_2, f$$

Given the Frechet distribution of the idiosyncratic productivity shock, the proportion of people with degree in f from region o_2 who goes to region d to work in sector S is given by:

$$m_{o_2 f, dS} = \frac{\left(\left(\frac{w_{f,dS}}{P_d} \right) \cdot u_{f,dS} \cdot \mu_{o_2 d}^2 \right)^{(\theta)}}{\Phi_{o_2 f}} \tag{6}$$

where $\quad \Phi_{o_2 f} = \sum_{d''S''} \left(\left(\frac{w_{fd''S''}}{P_d''} \right) \cdot u_{fd''S''} \cdot \mu_{o_2 d''}^2 \right)^{(\theta)}$

$\Phi_{o_2 f}$ is a measure of access to jobs for an individual from o_2 with degree f. It summarizes the expected value of all the job opportunities available to a person from o_2 with a degree in f, taking into account costs of migration and the distribution of job opportunities.

In stage 1, the individual maximizes $E(U_{io_1 o_2 f, dS})$ by choosing (o_2, f).

Proposition 1: If $\eta \sim Frechet(\theta)$, then $\eta^\alpha \sim Frechet(\frac{\theta}{\alpha})$

Proof: See Online Appendix C.5

---

[10]Here, being born in o_1 is equivalent to completing non-tertiary education in o_1.

Proposition 2: If $\eta_i \sim Frechet(\theta)$, then $E(max_i(a_i \times \eta_i)) = (\sum_i a_i^{\theta})^{\frac{1}{\theta}} \Gamma(1 - \frac{1}{\theta})$

Proof: See Online Appendix C.5

Using propositions 1 and 2, the results of the maximization problem of an individual in stage 1, described by the left-hand side, is given by:

$$Max_{o_2,f}((\frac{a_{o_2f}w_{u,o_2}}{P_{o_2}} \cdot \mu_{o_1o_2}^1 \cdot \zeta_{io_2f})^{IU} E\Big(max_{d,S}\Big((\frac{w_{f,dS}}{P_d}) \cdot u_{f,dS} \cdot \eta_i \cdot \mu_{o_2d}^2\Big)^{(1-IU)}|o_1, o_2, f\Big)$$

$$= Max_{o_2,f}(\frac{a_{o_2f}}{P_{o_2}} \cdot \mu_{o_1o_2}^1 \cdot \zeta_{io_2f})^{IU} \Phi_{o_2f}^{\frac{(1-IU)}{\theta}}$$

where $E\Big(max_{d,S}\Big((\frac{w_{f,dS}}{P_d}) \cdot u_{f,dS} \cdot \eta_i \cdot \mu_{o_2d}^2\Big)^{(1-IU)}|o_1, o_2, f\Big) = \Phi_{o_2f}^{\frac{(1-IU)}{\theta}} \Gamma(1 - \frac{\theta}{(1-IU)})$ is the expected income prior to drawing match productivities for workers trained in field f at location o_2.

The proportion of people living in o_1 who studies f in region o_2 is then given by:

$$l_{o_1o_2f} = \frac{\Big((\frac{a_{o_2f}w_{u,o_2}\mu_{o_1o_2}^1}{P_{o_2}})^{IU} \Phi_{o_2f}^{\frac{(1-IU)}{\theta}}\Big)^{\frac{\gamma}{IU}}}{\Phi_{o_1}} \tag{7}$$

where $\Phi_{o_1} = \sum_{o_2',f'} \Big((a_{o_2'f'}\mu_{o_1o_2'}^1)^{IU}\Phi_{o_2'f'}^{\frac{(1-IU)}{\theta}}\Big)^{\frac{\gamma}{IU}}$.

Φ_{o_1} is a measure of access to education for an individual from o_1. It summarizes the expected returns from all the different types of education an individual from o_1 can access, taking into account the costs of migration and the distribution of work and education amenities.

Firms

There is perfect competition in the production of each variety. The representative firm of sector S in location d produce a variety of the sector S good using both high-skilled L_{hdS} and low-skilled labor L_{ldS}, combined in a nested CES constant returns to scale production function:

$$Q_{dS} = (Q_{hdS}^{\frac{\rho_S-1}{\rho_S}} + Q_{ldS}^{\frac{\rho_S-1}{\rho_S}})^{\frac{\rho_S}{\rho_S-1}} \tag{8}$$

where

$$Q_{hdS} = (\sum_{f \in college} A_{f,dS}(\tilde{L}_{f,dS})^{\frac{\rho_{hS}-1}{\rho_{hS}}})^{\frac{\rho_{hS}}{\rho_{hS}-1}} \tag{9}$$

and

$$Q_{ldS} = \Big(\sum_{f \in nocollege} A_{f,dS} (\tilde{L}_{f,dS})^{\frac{\rho_{lS}-1}{\rho_{lS}}} \Big)^{\frac{\rho_{lS}}{\rho_{lS}-1}}$$

where $\tilde{L}_{f,dS} = \overline{\eta}_{f,dS} L_{f,dS} = \Gamma(1 - \frac{1}{\theta}) L_{f,dS}$ is the effective labor supply.

The Armington structure of the model delivers a cost of living index P_d for each region, where $P_d = \Pi_S P_{dS}^{\alpha_S}$ and $(P_{dS})^{1-\sigma_S} = \sum_j (\tau_{jdS} p_{dS})^{1-\sigma_S}$.

External trade

The country exports a tradeable good to the RoW where each region of the country produces a variety of the tradeable good, and, in turn, imports an importable good from the RoW. The country is a price-taker in the world market so the price of the importable good is given. The income of the RoW is also exogenously given. [11] Gravity determines the level of trade between each region of the country and the RoW. People can move within the country but not outside the country. The demand for IT exports from region d $(E_{d,IT})$ is given by:

$$E_{d,IT} = \underbrace{\Big[\frac{\tau_{d,IT} p_{d,IT}}{\sum_{d'} (\tau_{d',IT} p_{d',IT})^{1-\sigma_{IT}}} \Big]^{(1-\sigma_{IT})}}_{\text{gravity}} E_{IT} \tag{10}$$

where $p_{d,IT}$ is the price of IT variety from region d, $\tau_{d,IT}$ are the costs of exporting IT to the RoW, mostly consisting of communication and management costs, E_{IT} is the RoW's income spent on the IT sector. Using equation 10, we can solve for IT prices in each district:

$$p_{d,IT} = \Big(\frac{E_{d,IT}}{E_{IT}} \Big)^{\frac{1}{1-\sigma_{IT}}} \frac{\big(\sum_{d'} (\tau_{d',IT} P_{d',IT})^{1-\sigma_{IT}} \big)^{\frac{1}{1-\sigma_{IT}}}}{\tau_{d,IT}} \tag{11}$$

Internal trade

The sectors other than IT and the importable goods sector are all internally traded. The gravity equations determining the flows of these internally traded sectors are given by:

$$Y_{dS} = \sum_j X_{djS} = \sum_j \tau_{dj}^{1-\sigma_S} p_{dS}^{1-\sigma_S} P_j^{\sigma_S-1} E_{jS} \tag{12}$$

$$E_{jS} = \sum_k X_{jkS} = \sum_k \tau_{jk}^{1-\sigma_S} p_{kS}^{1-\sigma_S} P_j^{\sigma_S-1} E_{jS} \tag{13}$$

---

[11] In theory, the income of the RoW consists of income from sales to itself, the domestic country, and all other countries. In this particular empirical setting, given that US exports to India consists of a negligible proportion of total US income, this assumption is tenable.

Equation 12 states that the income of sector S in region d equals the sum of exports from sector S in region d to all other districts. Equation 13 states that the expenditure of region j on sector S good must equal the sum of imports of good S from all other regions.

Equilibrium

For each region (in our analysis, district), equilibrium in the steady-state is defined as a set of sectoral employment according to field of study ($L_{f,dS}$), field-wise college enrollment (L_{o_2f}), wages ($w_{f,dS}$), prices (P_d), and quantities (Q_{dS}). For each district, the equilibrium takes as given population, amenities and bilateral migration costs of studying and working according to fields of education and sectors of employment, trade costs between domestic districts and between domestic districts and the RoW. It also takes as given the parameters governing the dispersion of productivity shocks (θ) and amenity shocks (γ), expenditure shares (α), the elasticity of substitution between high-skilled and low-skilled workers (ρ_S) and between different types of high-skilled workers (ρ_{hS}).

The steady-state equilibrium is governed by the following equations describing goods and labor market clearance:

1. Given productivities and the initial distribution of population, the quantity produced in each location is determined by the production functions.

2. Given quantities produced in each location and trade costs, exogenously given world income spent on IT, from equation 11 the price of the tradeable good is given by the market clearing for the tradeable good S' in each region d:

$$p_{S',d} = \frac{1}{Q_{S',d}} \frac{(p_{S',d}\tau_{S',d,RoW})^{1-\sigma_{S'}}}{\sum_{d'}(p_{S',d'}\tau_{S',d',RoW})^{1-\sigma_{S'}}}(Y_{RoW}) \tag{14}$$

3. Given quantities produced in each location and trade costs, price of the tradeable good, from equations 12 and 13 prices of the externally non-tradeable goods but internally tradeable goods S is given by market clearing of the non-tradeable goods:

$$p_{S,d} = \frac{1}{Q_{s,d}} \sum_j \tau_{dj}^{1-\sigma_S} p_{S,d}^{1-\sigma_S} P_j^{\sigma_S-1} \alpha(\sum_S p_{S,j}Q_{S,j}) \tag{15}$$

where the price index $P_d^{1-\sigma_S} == \sum_j \tau_{jd}^{1-\sigma_S} p_{S,j}^{1-\sigma_S}$, $(\sum_S p_{S,j}Q_{S,j})$ is the income of region j, and α is the proportion of income spent on the non-tradeable good. [12]

---

[12]Note that, conditions 2 and 3 automatically ensure that the trade balance condition is maintained. Summing over d in condition 2, one can easily see that the sales from IT in the domestic country is the same as the amount of income spent on IT goods in the foreign country. Now, for balanced trade, the

4. Given prices of both tradeable and non-tradeable goods, the wages of workers with field of education f working in industry S in region d are given by:

$$w_{f,dS} = p_S A_{f,dS} Q_{Sd}^{\frac{1}{\rho_S}} Q_{hSd}^{\frac{1}{\rho_{hS}} - \frac{1}{\rho_S}} (L_{f,dS}\Gamma(1 - \frac{1}{\theta}))^{\frac{1}{\rho_{hS}}} \tag{16}$$

5. Given wages and prices, migration flows for education determine the population distribution of skill at each location. The proportion of people from o migrating to j to seek education in field f is given by:

$$L_{o_2 f} = \sum_{o_1} l_{o_1, o_2 f} L_{o_1} \tag{17}$$

where L_{o_1} is the college eligible population in o.

6. Given wages, prices, and the distribution of skill in each region, the distribution of people with skill f working in industry S in region d is given by:

$$L_{f,dS} = \sum_{o_2} m_{o_2 f, dS} L_{o_2 f}$$

7. In the steady-state, the initial distribution of population working in different industries with different skill levels is equal to the final distribution.

This completes the description of equilibrium in this model. In appendix C.4 , I show that a competitive general equilibrium exists.

Summary of the mechanics of the model

This section describes how a rise in the demand for the externally traded good, in this case IT, affects employment, education, and ultimately welfare of individuals in different regions within the country. The rise in IT export demand translates into differential changes in IT real wages across regions, depending on the region's geographic location that determines how difficult it is to migrate there and the regions comparative advantage in IT, as measured by the historical regional software exports and the region's linguistic distance from English. People start moving into regions where the real wages rise faster. This is the place where the mobility costs for work matter. This part of the model is like a specific factors model in that engineers are more required in the IT sector. This is a standard spatial model with no changes in skills. But then the rise in real wages changes

amount of income spent on IT in the foreign country also has to be equal to the amount of income spent on imports by the domestic country. (3) uses the condition that the income spent on Non-IT goods by each region is α proportion of its income, that is, $\alpha(p_{IT} qIT + p_{NonIT} q_{NonIT})$. This implies that $(1 - \alpha)(p_{IT} qIT + p_{NonIT} q_{NonIT})$ is spent on imports. Since condition 2 ensures that sum of imports is equal to value of sales from IT, trade balance is maintained.

the incentives for higher education, specially engineering. Individuals who are closer to skilled jobs or who are closer to good education facilities are more likely to get educated. Thus, enrollment rises. This is where the mobility costs for education matter and this is the new component that I add to existing spatial models. This generates a Heckscher-Ohlin (HO) type response to changes in skilled wages.

6 Identification and estimation

In this section, I estimate the structural parameters that determine the migration and IT trade costs and measure the expenditure shares on goods using available expenditure data. I then use the estimated parameters and the measured quantities, along with the available data on employment, wages, migration and enrollment, to back out the unknown amenities and productivities consistent with the model. Adapting the model for estimation, I assume $F=3$, where $f \in F$. f can be college degree in engineering, college degree in any other field, henceforth referred to as non-engineering, or no college degree at all. There are two types of high-skilled workers: those who complete a college degree in engineering and those who complete a college degree but not in engineering. There is only one type of low-skilled worker, those who do not go to college. There are 7 sectors in the economy ($S=7$), where these sectors are: agriculture and allied activities, manufacturing, wholesale and retail trade, low-skill services, skilled services except IT, the IT sector and an importable sector.

IT is only consumed by the RoW. There is an importable sector: goods in this sector are not produced domestically but are consumed domestically. Goods in the other sectors are all traded internally between districts.

6.1 Estimation of migration costs

6.1.1 Estimation of migration costs due to education

In this section, the migration costs of people moving to acquire education are estimated. Taking the logarithm on both sides of equation 49, the ideal gravity equation of flows of workers from o_1 to o_2 who move to study field f would be estimated by:

$$\ln(l_{o_1 o_2 f}) = \gamma \ln(\mu^1_{o_1 o_2}) + \gamma \frac{(1-IU)}{IU\theta} \ln(\Phi_{o_2 f}) - \ln(\Phi_{o_2}) + a'_{o_1 o_2 f} \qquad (18)$$

where f is engineering, non-engineering, or no college. The equation states that the the proportion of people who move from o_1 to o_2 to study field f depends on

i) The expected return from studying f in o_2 $(\ln(\Phi_{o_2 f}))$

ii) The bilateral migration costs of moving from o_1 to o_2, given by $\mu^1_{o_1 o_2}$

iii) The geographic advantage of the origin district, determined by its proximity to regions with good job and education opportunities (Φ_{o_1})

iv) $a'_{o_1 o_2 f}$ is an error term which captures any field of education f, origin and destination specific factors such as the time and money cost of education and unobserved preferences for education.

However, flows of people disaggregated by reason for migration, origin, destination, and field of study are not available. Instead, the data informs us of the number of people migrating for education from every origin district o_1 to every destination district o_2, aggregated across all fields of education f that they chose to study. Aggregating equation 18 across all fields of education and taking the logarithm on both sides, one gets the following equation:

$$\ln\left(l_{o_1 o_2}\right) = \gamma \ln(\mu^1_{o_1 o_2}) - \ln(\Phi_{o_1}) + \gamma(\sum_f (a_{o_2 f})^{IU}(\Phi_{o_2 f}^{\frac{(1-IU)}{\theta}})) \tag{19}$$

Following the migration gravity literature, I parameterize the costs of migration in equation 18 where the migration costs depend on geographic and cultural distances:

$$\ln\left(\mu^1_{o_1 o_2}\right) = \lambda'_1 \ln\left(DistCentroid_{o_1 o_2}\right) + \lambda'_2 lang_{o_1 o_2} + \lambda'_3 Diff_{o_1 o_2}^{diff-NBR} + \\ \lambda'_4 D_{o_1 o_2}^{same-NBR} + \lambda'_5 D_{o_1 o_2}^{same-notNBR} \tag{20}$$

where $DistCentroid_{o_1 o_2}$ measures the distance between district-centroids and $(lang_{o_1 o_2})$ measures the proportion of people speaking a common language in districts o_1 and o_2. If two districts belong to different states but share the same border, $diff-NBR$=1. If two districts belong to the same state and also share a border, $same-NBR$=1. If two districts belong to the same state and are not neighbors, $same-notNBR$=1.

The estimating equation becomes:

$$log(l_{o_1 o_2}) = f_{o_1} + f_{o_2} - \lambda lndist_{o_1 o_2} + \epsilon_{o_1 o_2} \tag{21}$$

, where $f_{o_2} = \gamma log(\sum_f (a_{o_2 f})^{IU}(\Phi_{o_2 f}^{\frac{(1-IU)}{\theta}}))$, $f_{o_1} = log\Phi_{o_1}$, $\lambda' = (\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4, \lambda'_5)$ and $\lambda = \gamma\lambda'$ is a combination of elasticity of migration flows to migration costs, and the elasticity of migration costs to distance. $\epsilon_{o_1 o_2}$ includes any measurement error or random factors not correlated with the distance measures. $\mu^1_{o_1 o_2}{}^\gamma$ can be estimated using equations

20 and 21 in the usual gravity estimation framework using origin, destination fixed effects and bilateral measures of distances.

Given bilateral migration data on the number of people moving from district o_1 to district o_2 to acquire education, bilateral geographic and cultural distances, the composite parameter λ is identified in the cross-section by the elasticity of migration flows to distances. The key assumption required for the identification of λ is that the unobserved error term $\epsilon_{o_1 o_2}$ which is not derived from the model and does not represent any structural object, is random measurement error and is uncorrelated with bilateral district to district cultural and geographic distances.

Regression 21 thus gives an estimate of

$$\mu_{o_1 o_2}^1{}^\gamma = exp(\gamma\lambda_1' ln(DistCentroid_{o_1 o_2}) + \gamma\lambda_2' lang_{o_1 o_2} + \gamma\lambda_3' Diff_{o_1 o_2}^{diff-NBR} + \gamma\lambda_4' D_{o_1 o_2}^{same-NBR} + \gamma\lambda_5' D_{o_1 o_2}^{same-notNBR})$$

The results of the estimation are given in Table III.

6.1.2   Joint estimation of migration due to work and amenities

To estimate the migration costs for work, the ideal regression would be estimating the log of 6, the migration flow equation for work:

$$log(m_{o_2 f,dS}) = \theta log(w_{f,dS}) + \theta log\mu_{o_2 d}^2 - \theta log(P_d) + \theta log u_{f,dS} - \theta log\Phi_{o_2 f} + \epsilon_{o_2 f,dS} \quad (22)$$

This relates the proportion of people from o_2 with degree f who move to d to work in sector S, $(m_{o_2 f,dS})$, to the wages of workers with degree f working in sector S in location d $(w_{f,dS})$, bilateral migration costs $(\mu_{o_2 d}^2)$, destination-specific prices (P_d), and the option value of a degree from location o_2 in field f, $(\Phi_{o_2 f})$. The option value summarizes the job opportunities available to a person who has completed degree f in location o_2. The error term, $\epsilon_{o_2 f dS}$, represents random measurement errors.

However, the available data does not inform us of bilateral migration flows disaggregated according to degree of education and sector of work. The data give information on the number of people who moved from district o_2 to district d for work, aggregated across all fields of education and all sectors of work. Summing across all fields f and sectors S, the estimable regression equation is given by:

$$ln(m_{o_2 d}) = \theta log(\sum_{f,S} \frac{w_{f,dS}}{\Phi_{o_2 f}}) + \theta log(\mu_{o_2 d}^2) - \theta log P_d + \theta(log\sum_{f,S} u_{f,dS}) + \epsilon_{o_2 d} \quad (23)$$

This relates the flow of people who move from location o_2 to location d for work to the average wage in location d in field f weighted by the option value of studying f

($\Phi_{o_2 f}$). Since the option value of education varies by origin (o_2) and field of education (f), the relative attractiveness of a destination is no longer separable in just the origin and destination fixed effects, as in traditional gravity models. The problem is that the option value of education contains the unobserved migration costs $\mu^2_{o_2 d}$ and amenities ($u_{f,dS}$), and so this relative attractiveness is not known. If we treat this relative attractiveness of a destination as unknown, the existence of unobserved migration costs in the error will bias the estimate of θ.

First, for a relatively remote district o_2, a rise in bilateral migration cost to d will reduce migration to d but by not as much compared to a district that is relatively well-connected to regions with employment opportunities, since people from the remote district have fewer options to choose from. Even this effect will differ according to an individual's skill level depending on how valuable destination d is for that skill group. Thus, the existence of this unaccounted for and unknown remoteness measure in the error term will bias the estimate of the elasticity of migration costs downward.

On the other hand, for people in well-connected locations, if the migration cost to a particular district falls, they can more easily turn to other districts compared to their more remote counterparts and this effect varies according to their field of training f. For districts in well-connected locations, the elasticity of migration to migration costs are thus over-estimated.

On the aggregate, it remains an empirical question as to which effect dominates.

The costs of migration depend on distance:

$$
\begin{aligned}
log\mu^2_{o_2 d} = \zeta'_1 log(DistCentroid_{o_2 d}) + \zeta'_2 loglang_{o_2 d} + \zeta'_3 D^{diff-NBR}_{o_2 d} + \\
\zeta'_4 D^{same-NBR}_{o_2 d} + \zeta'_5 D^{same-notNBR}_{jd}
\end{aligned}
\tag{24}
$$

Rewriting the estimating equation by inserting the migration cost in terms of distance,

$$
ln(m_{o_2 d}) = \theta log(\sum_{f,S} \frac{w_{f,dS}}{\Phi_{o_2 f}}) - \zeta logdist_{o_2 d} - \theta logP_d + \theta(log \sum_{f,S} u_{f,dS}) + \epsilon_{o_2 d}
\tag{25}
$$

where $logdist_{o_2 d}$ is the vector of distances mentioned above, and $\zeta = -\theta\zeta'$ and $\zeta' = (\zeta'_1, \zeta'_2, \zeta'_3, \zeta'_4, \zeta'_5)$

I use a nested nonlinear least squares approach to estimate ζ. The idea is to explicitly account for the effect of the unobserved option value of education by location and degree, thereby correcting the source of the bias in traditional gravity estimation.

After accounting for the unobserved option value of education as I will describe below, the moment condition that identifies ζ is:

$$E(\epsilon_{o_2d}|lndist_{o_2d}) = 0$$

The assumption is that after accounting for the unobserved attractiveness of the destination region relative to the origin region, the remaining unobserved term is white noise, uncorrelated with migration costs.

Since the option value of education contains the unobserved amenities, in practice, for estimation, I use a nested non-linear least squares procedure where I make a guess of migration costs and use labor market data on workers across fields, sectors of occupation, and locations to back out the unknown amenities as shown in equation 26. The unique district and field of education amenities are given by: (See equations 44 and 45 in appendix for detailed derivations)

$$
\begin{aligned}
L_{f,dS} &= \sum_{o_2} \frac{\mu_{o_2d}^2{}^{\theta}\left(\frac{W_{f,dS}u_{f,dS}}{P_d}\right)_f^{\theta}}{\sum_{o_2''S''}\mu_{o_2d''}^2{}^{\theta}\left(\frac{W_{f,d''S''}u_{f,d''S''}}{P_d''}\right)^{\theta}}L_{fo_2j} \\
&= u_{f,dS}^{\theta}\sum_k\sum_{o_2}\frac{\mu_{o_2dS}^2{}^{\theta}\left(\frac{W_{f,dS}}{P_d}\right)_s^{\theta}}{\sum_{d''S''}\mu_{o_2d''}^2{}^{\theta}\left(\frac{W_{f,d''S''}u_{f,d''S''}}{P_d''}\right)^{\theta}}L_{fo_2} \\
u_{f,dS} &= \frac{(L_{f,dS}/(\sum_k\sum_{o_2}\frac{\mu_{o_2d}^2{}^{\theta}\left(\frac{W_{f,dS}}{P_d}\right)^{\theta}}{\sum_{d''S''}\mu_{o_2d''}^2{}^{\theta}\left(\frac{W_{f,d''S''}u_{f,d''S''}}{P_d''}\right)^{\theta}}L_{fo_2}))^{\frac{1}{\theta}}}{\frac{W_{f,dS}}{P_d}}
\end{aligned}
\tag{26}
$$

In the outer loop, I choose migration cost parameters to minimize the distance between bilateral migration flows predicted by the model and observed in the data in 2001, the only year for which such detailed migration data is available. Given the assumption that migration costs do not change during the period under study, I use the estimated migration costs and the distribution of employment post-2004 to recover unknown amenities. As unknown amenities are recovered in the last step from the distribution of population in each location, I update the amenities and re-estimate equation 25 until the migration costs converge. Note that the estimation of unknown amenities requires an estimate of θ, which is described in section 6.1.3.

In the same way, given estimated migration costs for education and the option value of education, one can use population with and without college degrees in each location to solve for unknown quantities a_{o_2f}, which includes the time and money cost of education as well as unobserved preferences for education.

The results of the estimation procedure are given in Table III. Columns 1 and 2

report the estimation results of the traditional PPML gravity regression where the reasons for migration are education and work, respectively. In the third column, I report the estimates for work using the non-linear least squares method.

Reading off columns 1 and 3 of Table III, a 1% increase in distance reduces migration by 0.60% when the reason for migration is education and by 0.88% when the reason for migration is work. These estimates are comparable with findings in the traditional gravity literature. For example, Bryan and Morten (2019) finds that a 1% increase in distance leads to a 0.7% reduction in the proportion migrating. Note that migration cost estimates from the literature, including Bryan and Morten (2019), are not able to estimate migration costs separately by reasons for migration, and their estimates could be a weighted average of the elasticity of migration flows to distances for work and for education. [13]

With these estimates, the average iceberg cost of migration when people migrate within state, either across or within a district, turns out to be 0.88 on average. This means that migrants have to be compensated about 88% in terms of lost utility when they migrate, and this loss includes many different factors such as cultural differences, loss of home network, and transportation costs.

The notable point from Table III is that the effect of state borders is large, and it is especially larger for education: Being in the same state increases migration between neighboring districts by about 109% when the reason for migration is education and by 49% when the reason for migration is work.[14] Many state-level policies, such as quotas for in-state students, create barriers to mobility that differ according to the reason for migration. These estimates suggest that there could be scope for policy interventions in reducing the mobility costs across state borders.

In the literature, only combined estimates of migration costs aggregated across reasons for migration are available, and these estimates vary across countries. For example, Tombe and Zhu (2019) finds that migrants, on average, have to be compensated 82% more than non-migrants in China in 2000, and this cost is almost 1.75 times larger when workers migrate across provinces than when they stay within their own province. Bryan and Morten (2019) finds that in Indonesia, on average, migrant workers have to be paid 38% more if they were to receive the same wage as non-migrant workers. They estimate a much lower cost, only 15%, for the US. It would be interesting to compare the finding of substantially larger migration cost for education relative to that for work for India with

---

[13]They do not separately account for cultural differences. As long as geographic distances are positively correlated with cultural distances, their estimates would be over-stating the effect of physical distance

[14]The state border effects are calculated as $exp(4.8030 - 3.7635)^{(\frac{1}{2.61})}$ for work and $exp(3.646 - 2.339)^{(\frac{1}{1.77})}$ for education. The estimated elasticities of migration costs with respect to work and education are 2.61 and 1.77 respectively, as show in section 6.1.3.

other countries if such data are available for other countries.

6.1.3 Estimation of elasticity of migration flows to migration Costs

The elasticity of migration flows to migration costs is the dispersion parameter θ that governs the variance of the idiosyncratic component of workers' productivity draws. The higher the value of θ, the lower is the variance in productivity, and thus workers are more identical. This means that workers tend to respond more similarly to changes in migration costs compared to when they are more heterogeneous in their productivities. Thus, for a given rise in migration cost, the higher is θ, the larger is the fall in migration.

Following Fan (2019), I use the variance in the wage distribution of stayers, that is, the wage distribution of people who do not migrate for work, to identify θ. Using the properties of the Frechet distribution, it can be shown that the productivity distribution of stayers also follows a Frechet distribution where the mean varies by field of education, sector of work, and location of degree. For any (f, d, S) combination, the wage observed in the data is the effective wage $(\tilde{w}_{f,dS})$, where

$$\tilde{w}_{if,dS} = w_{f,dS}\eta_{if,dS}$$

Taking logs on both sides,

$$\ln(\tilde{w}_{if,dS}) = F_{f,dS} + \ln(\eta_{if,dS})$$

where $F_{f,dS}$ is a sector of job, field of education, and district fixed effect which is a combination of average wage per effective unit of labor and the average productivity of stayers. The variance of exponentiated residuals $(\eta_{if,dS})$ identifies θ, which turns out to be 2.61. This is very similar to the estimate of Fan (2019), who used the same method to estimate these elasticities to be within the tight range of 2.50 to 2.73. The assumption is that after controlling for field of education, sector, and location of work, the remaining variation in individual wages for those who stay back in the same location is due to variation in the idiosyncratic component, which can include factors such as ability, talent, and family background.

Given the estimate of θ (elasticity of migration flows to migration costs for work) and γ (elasticity of migration flows to distance for work) from section 6.1.2, it is possible to separately identify the elasticity of migration flows to migration costs for education (ζ). The assumption required for this identification is the following: the elasticity of migration costs to geographic distance is the same irrespective of the reason for migration, once institutional boundaries such as state borders and neighboring districts dummies have been accounted for.

Note that this assumption does not require the elasticity of migration flows to distance to be the same. In fact, these elasticities are very different, as we estimated before. It only requires the costs of migration to respond to geographic distances in exactly the same way, once we have accounted for state-specific institutional barriers such as differential quotas for work and for education. An example of a violation of this assumption would be any factor that increases or decreases the migration costs for education relative to work over the same geographic distance. For example, one such factor would be the provision of special transportation for students.

By assumption, $\zeta_1' = \lambda_1'$ Thus,

$$\frac{\lambda}{\gamma} = \frac{\zeta}{\theta}$$

Given $\theta = 2.61$, $\zeta = .88$, $\lambda = .602$, the above identity yields $\gamma = 1.77$.

This completes the description of my estimation strategy for migration costs.

6.2 Trade costs

6.2.1 Trade costs in the IT sector:

In this model, IT is the only good traded with the RoW and it is not consumed domestically. Taking the logarithm on both sides of equation 11, the gravity equation expressing IT trade flows as a function of IT prices and comparative advantage, and getting rid of IT in the notation, I get the following estimating equation:

$$\ln(\frac{E_{d,t}}{E,t}) = C + (1 - \sigma_{IT})\ln(\tau_d) + (1 - \sigma_{IT})\ln(p_{d,t}) \tag{27}$$

where C$=-(1 - \sigma_{IT})\ln(\sum_{d'}(\tau_{d'}p_{d'})^{1-\sigma_{IT}})$ is a quantity that is constant across districts. Following Shastry (2012) and Banerjee and Duflo (2000), I parameterize the costs of exporting IT as a function of the linguistic distance of d from English and the prior software exports in 1995.

$$(1 - \sigma_{IT})\ln(\tau_d) = \kappa_{IT}\ln(dist_d) \tag{28}$$

The historical comparative advantage of a district in this sector depends on the prior links of a district to the RoW, measured by the proportion of software exports historically exported from that district. Prior connections, through building reputation, play an important role in determining the volume of transactions in this sector (Banerjee and Duflo (2000)). Shastry (2012) showed that linguistic distance of each regional language

spoken in a district from English determines the cost of learning English for individuals in that district. Since English proficiency is a necessary skill in this industry, the comparative advantage of a district also depends on the linguistic distance of the district from English. Let $\vec{dist}_{d,IT}$ be the vector denoting the linguistic distance of each district from English and the proportion of historical software exports from d, measured using 1995 export data.

Note that price is unobserved since it includes the unobserved productivities. Using the structure of the production function, price can be log-linearly decomposed into its known and unknown components. Using marginal cost pricing,

$$
\begin{aligned}
(p_{d,IT})^{1-\rho_{IT}} &= ((A_{e,d,IT}^{\rho_{h,IT}} w_{e,d,IT}^{(1-\rho_{h,IT})} + A_{ne,d,IT}^{\rho_{h,IT}} w_{ne,d,IT}^{(1-\rho_{h,IT})}))^{\frac{1-\rho_{IT}}{1-\rho_{h,IT}}} + (A_{l,d,IT}^{-1} w_{l,d,IT})^{(1-\rho_{IT})} \\
&= A_{ne,d,IT}^{\frac{-\rho_{h,IT}(\rho_{IT}-1)}{1-\rho_{h,IT}}} (\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}} + \tilde{x}_{l,d,IT}^{(-\rho_{IT})}(w_{l,d,IT})^{1-\rho_{IT}})
\end{aligned}
$$

(29)

where $\tilde{x}_{l,d,IT} = \dfrac{A_{ne,d,IT}^{(\frac{\rho_{h,IT}}{\rho_{h,IT}-1})(\frac{1}{\rho_{IT}}-1)}}{A_{l,d,IT}^{\frac{1}{\rho_{IT}}-1}} = \dfrac{W_{l,d,IT}(L_{l,d,IT})^{\frac{1}{\rho_{IT}}}}{\tilde{p}_{S,IT}^{h}(\tilde{Q}_{h,d,IT})^{\frac{1}{\rho_{IT}}}}$

and $\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}} = ((\frac{A_{e,d,IT}}{A_{ne,d,IT}})^{\rho_{h,IT}} w_{e,d,IT}^{1-\rho_{h,IT}} + w_{ne,d,IT}^{1-\rho_{h,IT}})^{\frac{1}{1-\rho_{h,IT}}}$

Due to the firm's first order conditions, the ratio of productivities is a function of known wages and employment, and therefore, both $\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}}$ and $\tilde{x}_{l,d,IT}$ are also functions of observables. For derivation see Appendix C.2.

Substituting this, the estimating equation becomes:

$$
\ln(\frac{E_{d,t}}{E_t}) = C + \kappa_{IT} \ln(dist_{d,RoW,IT}) + (1-\sigma_{IT})ln((\tilde{p}_{d,t}^{h}{}^{1-\rho_{IT}} + \tilde{x}_{l,d,t}^{(-\rho_{IT})}(w_{l,d,t})^{1-\rho_{IT}})^{\frac{1}{(1-\rho_{IT})}}) -
$$
$$
\frac{(\sigma_{IT}-1)}{(\rho_{IT}-1)} \frac{\rho_{h,IT}(\rho_{IT}-1)}{1-\rho_{h,IT}} \ln(A_{ne,d,t})
$$

Taking first differences,

$$
\Delta \ln(\frac{E_{d,t}}{E_{RoW,t}}) = (1-\sigma_{IT})\Delta OC_{dt} + \Delta \ln(\tilde{A}_{ne,d,t})
$$

(30)

where $ln((\tilde{p}_{d,t}^{h}{}^{1-\rho_{IT}} + \tilde{x}_{l,d,t}^{(-\rho_{IT})}(w_{l,d,t})^{1-\rho_{IT}})^{\frac{1}{(1-\rho_{IT})}})$ is the observable part of MC, referred to as OC_{dt}

$$
\Delta \ln(\tilde{A}_{ne,d,t}) = \Delta \frac{(1-\sigma_{IT})}{(1-\rho_{IT})} \frac{\rho_{h,IT}(\rho_{IT}-1)}{1-\rho_{h,IT}} \ln A_{ne,d,t}
$$

Intuitively, in equilibrium, how responsive IT exports are to changes in price depends on the elasticity of substitution between different varieties of IT products (σ_{IT}), where each variety corresponds to a region. The lower the elasticity of substitution, the more difficult it is to switch to a different variety as the price of a particular variety rises, and the less responsive is IT demand to IT prices..

Since in general equilibrium the unobserved district specific productivities determine the marginal cost of production, σ_{IT} cannot be recovered through a linear regression of IT exports on the observed part of marginal cost. I construct an instrument by leveraging the IT boom of the late 1990s and early 2000. As demand for IT increased, the prices of IT increased in all regions that produce IT. However, the capacities of IT production differ across regions. In particular, regions that are better connected geographically to other populous regions could expand supply more because people can migrate more easily into these regions and thus the supply of labor in these regions is more elastic. Also, regions with a historical comparative advantage in IT production attracted more IT demand from abroad. To formalize this intuition, I develop an instrument by interacting a measure of labor supply for each region (defined below) with the historical software exports of a region. A measure of labor supply access for each region is summarized by

$$LMA_d = \sum_o L_o(\frac{1}{distance_{o,d}})^{-1}$$

The instrument, referred to as I_d is formally defined below:

$$I_d = LMA_d * HistoricalSoftwareExport$$

Regions that are better connected historically and where the potential labor supply are high will see lower increases in marginal costs and hence lower increases in prices. On the other hand, regions that have potentially high labor supply but are not historically connected will have relatively larger increases in MC.

This estimation requires the assumption that changes in the productivity of non-engineers in the IT sector, in the pre and post-2000 boom periods are uncorrelated with the pre-period exports and the remoteness of a region during the period of the IT boom. The unobserved productivities, by model construction, do not depend on historical software exports, and the historical distribution of college educated workers. These productivities are the residual quantities that explain the deviation of predicted output from actual output, after these known quantities are taken into account. These historical factors, in turn, are not affected by future changes in productivities. However, to account for the fact that in reality district level productivities in the IT sector can be affected by these factors, I additionally run the following specifications:

First, I include controls for the geographic remoteness of a district, as measured by the average log distance of a district to other districts. Second, I include state fixed effects to account for any differential growth in productivities across states.

In Table (IV) column 1, I report the results from the OLS estimation. Columns 2,3, and 4 report the results from the IV estimation. The regression result in column 3 controls for the remoteness of districts and the regression result in column 4 additionally controls for state fixed effects, which, in a first difference equation implies controlling for a linear state-level time trend. Export demand responds negatively to changes in observable prices/ MCs. In the most demanding specification, reported in column 4, I find that a 1% increase in prices leads to a .45% fall in demand, which translates into an elasticity of substitution of 1.45.

The first stage is reported in Table (V). The instruments are strong across all specifications (F-stats of 210.37, 116.20, and 47.66). The IV has the right sign: marginal cost is lower in districts that have a larger access to college educated workers and are more connected abroad.

The estimated value of elasticity of substitution, at 1.45, is pretty low compared to elasticities of substitution between different varieties that the literature has estimated (between 3 and 5). There are a couple of caveats when comparing σ_{IT} to the estimates from the literature. First, the literature has mostly estimated the elasticities of substitution between varieties traded internationally. Second, my estimate is specifically for the IT industry, for which we do not have any known estimate in the literature. The low value of elasticity of substitution between the IT varieties of different regions could be justified on the ground that these regions specialize in very different types of tasks, such as, data processing, software development, multimedia graphics, as is reported in the NASSCOM software data.

6.2.2 Trade costs in the non-IT sector

The iceberg transport cost is taken to be, $\tau_{od} = distance_{od}^1$, calibrating the distance elasticity to the canonical value of -1 (Head and Mayer (2014)).[15]

6.3 Quantifying sector-specific productivities

We use the equality of marginal costs to prices to back out the unobserved amenities, after calculating the observable part of MC that depends on known wages and employment,

---

[15]The only estimate for India is from Donaldson (2018) who estimate it to be -1.69 from Colonial India. Since connectivity has much improved since then, taking the classic estimate seems more appropriate in this case.

following the long tradition in urban economics (see for example Allen and Arkolakis (2014), Allen et al. (2018)).

Equation (31) determines prices:

$$(p_{d,IT})^{1-\rho_{IT}} = A_{ne,d,IT}^{\frac{-\rho_{h,IT}(\rho_{IT}-1)}{1-\rho_{h,IT}}} (\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}} + \tilde{x}_{l,d,IT}^{(-\rho_{IT})}(w_{l,d,IT})^{1-\rho_S}) \qquad (31)$$

Since $(p_{d,IT})^{1-\rho_{IT}}$, $\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}}$, $\tilde{x}_{l,d,IT}^{(-\rho_{IT})}$, $(w_{l,d,IT})^{1-\rho_{IT}})$, $\rho_{h,IT}$ are all known, we can recover $A_{ne,d,IT}$ using

$$A_{ne,d,IT} = \Big(\frac{(\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}} + \tilde{x}_{l,d,IT}^{(-\rho_{IT})}(w_{l,d,IT})^{1-\rho_{IT}})^{\frac{1}{1-\rho_{IT}}}}{p_{d,IT}} \Big)^{\frac{\rho_{h,IT}-1}{\rho_{h,IT}}} \qquad (32)$$

Intuitively, how the magnitude of estimated prices differ from that of the observed components of marginal cost consisting of the information on wages and employment, helps determine productivities. Note that $(p_{d,IT})$ is known by recovering it from equation 11, given estimated trade costs σ_{IT} and exports. $\tilde{p}_{d,IT}^{h}{}^{1-\rho_{IT}}$ and $\tilde{x}_{l,d,IT}^{(-\rho_{IT})}$ are known as they are functions of observables. Finally we recover the productivity of low-skilled workers in the IT sector $A_{l,d,IT}$ in all locations by using the firm's first order condition below:

$$\tilde{x}_{l,d,IT} = \frac{A_{ne,d,IT}^{(\frac{\rho_{h,IT}}{\rho_{h,IT}-1})(\frac{1}{\rho_{IT}}-1)}}{A_{l,d,IT}^{\frac{1}{\rho_{IT}}-1}} = \frac{W_{l,d,IT}(L_{l,d,IT})^{\frac{1}{\rho_{IT}}}}{\tilde{p}_{d,IT}^{h}(\tilde{Q}_{h,d,IT})^{\frac{1}{\rho_{IT}}}}$$

and then $A_{e,d,IT}$ by using the firm's first order conditions.

To recover prices in the internally traded sectors, use equations 12 and 13, the identities that state that the income of sector S in district d equals the sum of exports from sector S in district d to all other districts, and the expenditure of district d on sector S good must equal the sum of imports of good S from all other districts, respectively. Combining these two equations, prices can be expressed as:

$$p_{d,S}^{1-\sigma_S} = \sum_j (\tau_{dj})^{1-\sigma_S} (\sum_k \tau_{kj}^{1-\sigma_S} p_{kS}^{1-\sigma_S})^{-1} \frac{E_{jS}}{Y_{dS}} \qquad (33)$$

where S is any sector other than IT and the importable goods sector.

Income of each region Y_{dS} is obtained by summing wage bill and employment. Expenditure of each region on sector S goods E_{dS} is calculated given share of GDP spent on sector S good. Internal trade costs τ_{jdS} are calculated given distances between districts and σ_S from the literature. Productivities in the internally traded sector can be recovered in exactly the same way as in the IT sector described above.

6.4 Calibration from the literature

The elasticity of substitution between engineers and non-engineers is calibrated to 2 across all sectors Ryoo and Rosen (2004). The elasticity of substitution between high and low skilled labor (college and non-college graduates) is taken to be 1.7 from Khanna and Morales (2017) which apply Card and Lemieux (2001) methodology to Indian data and find the estimate to be consistent with the literature (such as in Katz and Murphy (1992), Card and Lemieux (2001) and Goldin and Katz (2007)). The elasticity of substitution σ between different types of goods traded internally within India is taken to be 5 following Simonovska and Waugh (2014). Several other papers estimate elasticities of substitution that are close. For example, Van Leemput (2016) estimate an elasticity of substitution between different types of agricultural goods in India as 5.6. The weight on current period utility (IU) is taken as 0.53, which corresponds to an intertemporal elasticity of substitution of 0.9, which is standard in the literature. Share of consumption expenditure on agriculture $\alpha_2 = 0.38$, share of consumption expenditure on manufacturing $\alpha_3 = 0.16$, share of consumption expenditure on non-traded goods $\alpha_s = 0.37$, share of consumption expenditure on high-skill services $\alpha_1 = 0.07$, share of consumption expenditure on imports $(1 - \alpha_1 - \alpha_2 - \alpha_3 - \alpha_s) = 0.02$ are all obtained from official government reports. The price of US imports is normalized to be 1. Table VI summarizes the parameter values.

6.5 Model validity

Given model parameters and the exogenous amenities and productivities, the model makes predictions about the equilibrium wages, employment, and enrollment across districts, all of which are observable quantities in the data. In this section, I validate the model by first showing that the model generated data can replicate the reduced form facts established in section 5.

6.6 Replication of reduced form facts

Reduced form fact 1: IT employment and engineering enrollment positively respond to software exports

Using the model generated data, I repeat the reduced form regression and plot the coefficients β_t in figure VII

$$Y_{dt} = \alpha_t + \gamma_d + \chi_d * t + \beta_t Exports_{d,1995} + \epsilon_{dt} \tag{34}$$

where Y_{dt} is IT employment or engineering enrollment in district d at time t. $Exports_{d,1995}$

is the proportion of software exports from district d in the year 1995 out of total Indian IT exports in 1995. α_t are time fixed effects that capture any factors that are common to all districts at time t. γ_d are district fixed effects that capture any factors that are fixed over-time in district d. $\chi_d * t$ is a district-level time-trend capturing any linear trend in the outcome variable at the district level.

From this figure, just like in the data, we can see that post-1998, IT employment increased more in districts that had a higher level of software exports in 1995.

In figure X, I plot the response of engineering enrollment and here also the reduced form results are replicated: post 2000, engineering enrollment increased in districts with higher level of software exports.

Fact 2: The effects are heterogeneous. Employment responds more when nearby regions have higher engineering enrollment and higher IT exports. The heterogeneous effects are stronger in the long run.

In equation 2 below, I add an interaction term between the number of students enrolled in engineering in 1991 and the proportion of software exports from district d in 1995. Estimated coefficient δ_t is plotted in figure VI.

$$
\begin{aligned}
Y_{dt} = \alpha_t + \gamma_d + \beta_t * Exports_{d,1995} + \chi_t Enrollment_{d,1991} + \\
\delta_t Exports_{d,1995} * Enrollment_{d,1991} + \epsilon_{dt}
\end{aligned}
\tag{35}
$$

As in the reduced form counterpart, figure 13 shows that, conditional on the level of software exports, post 1998, IT employment responds more in districts that, in 1991, had more enrolled engineering students in same-state, nearby districts. The intuition, formalized in the model, is that in these districts, it is easier to expand future IT production due to having more college-educated, engineering program graduates in close proximity.

6.7 Non-targeted moments

In figures XI and XII below, I plot the percentage changes in log engineering enrollment over the periods 1991-2001 and 1991-2011, respectively, and compare with the data. Both figures show a positive relationship with a correlation coefficient of 0.28.

7 Quantification and counterfactuals

I first use the model to quantify the effect of the IT boom on the Indian economy. To do that, I use data on the changes in IT exports between 1995 and 2001, holding the estimated parameters and the model fundamentals (i.e., the exogenous amenities and productivities) constant between the pre- and post-boom periods. This is the period during which major international demand shocks led IT exports to expand by more than 50% annually, whereas secular growth in the post boom period averaged about 26%. Between 1995 and 2001, the share of IT in total GDP rises from 1.2% to 3.2% in the data (in the model, the same targeted moment increases from 1.2% to 3.2%). The IT boom also partly explains the reorientation of production in the Indian economy, away from manufacturing and towards the service sector. According to the model, the Indian IT boom explains 49% of the decline in manufacturing from about 18% to 16% between 1995 to 2001. The IT boom also explains 67% of the rise services as a fraction of GDP from about 37% in 1995 to 44% in 2001. These number takes into account the general equilibrium effects of the IT boom on other sectors of the economy.

The short-run is defined as the period during which people cannot change their skills. Since it takes at-least four years of college and two years of pre- college to complete an engineering degree, the long-run changes in skill composition will not be visible in the labor market until 2001. The long-run is defined as the period from 2001 to 2007 [16]. In the long-run, welfare is defined as:

$$\Gamma(1 - \frac{1}{\gamma})(\Phi_{o_1})^{\frac{1}{\gamma}}$$

where Φ_{o_1} measures the access to higher education for college-eligible individuals from o_1.
$\Phi_{o_1} = \sum_{o'_1, f'}((a_{o'_2 f'}\mu^1_{o_1 o'_2})^{IU}\Phi_{o'_2 f'}^{\frac{(1-IU)}{\theta}})^{\frac{\gamma}{IU}}$, where access to education, in turn, depends on education amenities, connectivity of region, and the job opportunities available from that region.

When skill levels are fixed, regional welfare depends on the access to jobs for each skill-group, weighted by the distribution of skills. In this case, welfare is defined as:

$$\Gamma(1 - \frac{1}{\theta})(\Phi_{Eng}^{\frac{1}{\theta}}propEng + \Phi_{NonEng}^{\frac{1}{\theta}}propNonEng + \Phi_{Un}^{\frac{1}{\theta}}propUn)$$

where Φ_s measures access to jobs for skill-group s, s = engineers, non-engineers, and unskilled.

---

[16]The model fundamentals, ie, the amenities and productivities, are estimated using the post period data since the post period data is better available across regions

With fixed skill-level, welfare increases on average by 0.61%, with the regional gains ranging from 0.17% to 2%.

Using the full general equilibrium model with endogenous skill acquisition, costs of human mobility for both education and work, and costs of moving goods internally, I find that the IT boom increased the average welfare of an individual by 1.12%. The average masks substantial variation across districts, with individuals born in districts with good access to jobs and education gaining as much as 2.63% while their counterparts in remote districts experienced gains as low as 0.67%.

Figure XIV below plots the histogram of regional welfare gains from the IT boom in the short and long-run. The long-run distribution of regional welfare gains lies to the right of the short-run distribution and the welfare gains are pareto improving.

The gains are positively correlated with the amenities for education, productivities in the IT sector, and the geographical connectivity of a district d, roughly approximated by the sum of inverse distances of d from all other districts. Districts in large states with good facilities for engineering education and connectivity to regions with jobs experienced the largest gains. This underscores the importance of accounting for endogenous skill acquisition in a general equilibrium setting.

To quantify the importance of mobility for education, I run a counterfactual where the option to move for education is shut off. I find that regional inequality increases by 15% and the gap between the welfare gains of the worst off and the best district increases by 63%. The gains in aggregate welfare would have been 1.79% lower if there were no education mobility. These numbers, especially the changes in regional welfare, are large despite the fact that the estimated migration costs for education are quite high. A high migration cost for education makes it more difficult to migrate across districts for education, undermining the importance of the endogenous education channel compared to a situation with no mobility frictions for education. However, since zero mobility frictions are not possible in reality, I conduct a counterfactual experiment where I reduce the costs of migrating for education across states.

Counterfactual policy: Reducing state quotas for education:

In the particular case of India, the widespread prevalence of in-state student quotas for higher educational institutions, reflected in the significantly higher costs of crossing state borders for education relative to that for work, increases the potential for districts in larger states with good educational facilities to gain more from the IT boom. Given that migration costs in India are one of the highest compared to available migration costs for other countries, the geographical connectivity of the district also plays an important

role in determining the welfare gains of the district. [17] There seems to be an obvious policy intervention in the education market – the reduction of state quotas for education – that is easier to implement than labor market policies that aim to move jobs. In the counterfactual, this is achieved by reducing the effect of state borders on migration costs for education to the same level as that for work.

The existing magnitude of quotas in higher education institutes in India is huge: most state colleges have home state quotas of 50%, with some being as high as 85 % . [18] The size of the state quota varies by state and by whether the university in question is public or private, but in general, it is a substantial proportion of the total class size (Kone et al. (2018)). [19] These domicile quotas, legally defined as quotas pertaining to the "place of living" or permanent residence, create huge costs of migrating to a different state. While such quotas also exist for jobs and thus create significant hurdles for moving across states, the employment quotas are more specific and less ubiquitous than the in-state education quotas. Such quotas are in no way unique to India, and exist in many other countries, including the US and China.

One way of looking at the effect of reduction in state quotas for education on the aggregate and distributional consequences of the IT boom is to reduce the effect of state border on migration for education to be the same as that for work. The reduction of migration costs has the effect of increasing aggregate welfare due to the IT boom by 1% and reducing regional inequality by 27%, compared to the equilibrium change with full migration costs. An interesting point to note is that the reduction in state quotas increases aggregate welfare by very little since not all districts gain from such a measure. A little less than a third of districts actually gain less from the IT boom in this case compared to the case with the current levels of higher education quotas. In fact, the gains are negatively correlated with the initial education amenities in a district, implying that districts that gained the most from this policy are those that did not initially have good education facilities. Figure XIII shows that the histogram of welfare gains with reduced education quotas has a lower spread than that with education quotas.

It is also clear from the histograms that reducing in-state quotas is not a Pareto improving measure and is likely to meet with political resistance from districts which benefit from the quota policy. In 2016, when the Chinese government announced a policy of reducing provincial quotas to increase opportunities for students from poorer provinces

---

[17]For example, many districts in Uttar Pradesh, the largest state of India in terms of land area and also the number of colleges, gained more than the average district

[18]Support for the 85% reservation policy started in Maharashtra from the year 2011 with the backing of nationalist state parties

[19]Reservation policy in India is a contentious issue. The magnitude of reservation at private institutions varies hugely from state to state and is still a matter of legal debate. For example, some private universities reserve seats following the state laws under which they were established. For example, in Haryana, private universities also have to reserve 25% of it's seats for students domiciled in Haryana.

to study in elite colleges, mostly located in the more prosperous provinces, there were wide spread protests in Beijing and Shanghai, fueled by the fear that this will hurt local students. [20] In the mid- 2000s, the Haryana state government invested land and money to build a hub for higher learning and a center for research, at the same time implementing a policy that reserves 25% seats for in-state students in all colleges across the state. [21]

8 Conclusion

This paper assesses the aggregate and distributional consequences of human capital response to trade for the spatial distribution of welfare. In answering this question, the paper makes three contributions: first, it introduces human capital acquisition decisions in a general equilibrium model with multiple locations. It shows that studying the effects of trade on the labor market without taking into account endogenous skill acquisition can underestimate the aggregate welfare gains from trade. Second, a key innovation of this paper compared to the existing literature on migration is that people can move either for work or for education. Using confidential and unique district-to-district Indian migration data disaggregated by reasons for migration, this paper provides the first separate estimates of mobility costs by reasons for migration. I show that quantifying both of these costs separately is important as these costs can significantly alter the welfare gains from trade depending on their relative magnitudes. Third, as a result of studying the interaction of education and labor market choices in the presence of changes in export-driven employment opportunities, this paper is able to suggest new forms of policy intervention to reduce inequality in regional welfare gains from trade.

Despite a lot of interest surrounding the IT boom and its effect on geographic inequality in India, the lack of disaggregated data made it challenging to quantify its effect on overall economic growth. This paper also takes the first step in collecting district-level data and building a general equilibrium model to quantify the effect of IT boom on skill acquisition and the regional distribution of welfare gains in India. Using the model, it finds that between 1995 and 2005, the IT boom in India increased the average individual welfare by 1.116%, with individuals born in districts with good access to jobs and education gaining as much as 2.36% while those in remote districts experienced gains as low as .67%. These gains are attenuated by high costs of mobility for education and for work across Indian districts, leaving scope for policy interventions in both the education and labor markets that have the potential to reduce regional inequality as well as increase aggregate welfare. There is scope for future work to further the research agenda presented in this paper by studying the regional welfare implications of endogenous edu-

---

[20]Source: South China Morning Post, 15th May 2016

[21]Source: Scroll in, 27th July, 2019

cation choice with trade in a dynamic framework, which can trace how welfare changes during the transition period from short to long run. The challenge will be to devise a way to tackle the large number of state spaces as people migrate across regions and over time for work and for education.

References

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*.

Allen, T., C. d. C. Dobbin, and M. Morten (2018, nov). Border Walls. Technical report, National Bureau of Economic Research, Cambridge, MA.

Atkin, D. (2016, aug). Endogenous skill acquisition and export manufacturing in Mexico. *American Economic Review 106*(8), 2046–2085.

Banerjee, A. V. and E. Duflo (2000). Reputation effects and the limits of contracting: A study of the Indian software industry. *Quarterly Journal of Economics 115*(3), 989–1017.

Blanchard, E. J. and W. W. Olney (2017, may). Globalization and human capital investment: Export composition drives educational attainment. *Journal of International Economics 106*, 165–183.

Borsook, I. (1987). Earnings, ability and international trade. *Journal of International Economics 22*(3-4), 281–295.

Bryan, G. and M. Morten (2019, jul). The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy*, 000–000.

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

Card, D. and T. Lemieux (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *Quarterly Journal of Economics 116*(2), 705–746.

Carmel, E. and P. Tjia (2005). *Offshoring Information Sourcing and Outsourcing to a Global Workforce*. Cambridge University Press.

Danziger, E. (2017, jul). Skill acquisition and the dynamics of trade-induced inequality. *Journal of International Economics 107*, 60–74.

Donaldson, D. (2018, apr). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

Economist (2003). America's pain, India's gain.

Edmonds, E. V., N. Pavcnik, and P. Topalova (2010, oct). Trade adjustment and human capital investments: Evidence from indian tariff reform. *American Economic Journal: Applied Economics 2*(4), 42–75.

Fan, J. (2019, jul). Internal geography, labor mobility, and the distributional impacts of trade. *American Economic Journal: Macroeconomics 11*(3), 252–288.

Ferriere, A., G. Navarro, and R. Reyes-Heroles (2018). Escaping the Losses from Trade: The Impact of Heterogeneity on Skill Acquisition *. Technical report.

Findlay, R. and H. Kierzkowski (1983, dec). International Trade and Human Capital: A Simple General Equilibrium Model. *Journal of Political Economy 91*(6), 957–978.

Fuchs, S. (2018). The Spoils of War: Trade Shocks during WWI and Spain's Regional Development Job Market Paper *. Technical report.

Goldin, C. and L. Katz (2007, nov). Long-Run Changes in the U.S. Wage Structure: Narrowing, Widening, Polarizing. Technical report, National Bureau of Economic Research, Cambridge, MA.

Greenland, A. and J. Lopresti (2016, may). Import exposure and human capital adjustment: Evidence from the U.S. *Journal of International Economics 100*, 50–60.

Head, K. and T. Mayer (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. In *Handbook of International Economics*, Volume 4, pp. 131–195. Elsevier B.V.

Hou, Y. and C. Karayalcin (2019, aug). Exports of primary goods and human capital accumulation. *Review of International Economics*.

Imbert, C. and J. Papp (2020). Costs and benefits of rural-urban migration: Evidence from india. *Journal of Development Economics 146*, 102473.

International Trade Center (2017). TRADE IMPACT FOR GOOD BRICS COUNTRIES: EMERGING PLAYERS IN GLOBAL SERVICES TRADE. Technical report.

Johnson, M. T. (2013, oct). Borrowing constraints, college enrollment, and delayed entry. *Journal of Labor Economics 31*(4), 669–725.

Jones, B. F. and . Kellogg (2014). The Human Capital Stock: A Generalized Approach †. *American Economic Review 104*(11), 3752–3777.

Kapur, D. (2002). India Review The causes and consequences of India's IT boom.

Katz, L. F. and K. M. Murphy (1992, feb). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Khanna, G. and N. Morales (2017, sep). The it Boom and Other Unintended Consequences of Chasing the American Dream. *SSRN Electronic Journal*.

Kone, Z. L., M. Y. Liu, A. Mattoo, C. Ozden, and S. Sharma (2018, jul). Internal borders and migration in India. *Journal of Economic Geography 18*(4), 729–759.

Kucheryavyy, K., G. Lyn, and A. Rodríguez-Clare (2016, aug). Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale. Technical report, National Bureau of Economic Research, Cambridge, MA.

Lee, C. (2005). Labor Market Status of Older Males in the United States, 1880-1940. *Social Science History 29*(1), 77–105.

Li, B. (2018, sep). Export expansion, skill acquisition and industry specialization: evidence from china. *Journal of International Economics 114*, 346–361.

Li, L. (2019). Skill-biased Imports, Human Capital Accumulation, and the Allocation of Talent *. Technical report.

Liu, M. Y. (2017). How does globalization affect educational attainment? Evidence from China *. Technical report.

Ma, S., Y. Liu, and M. Zhou (2019, aug). Trade, educational costs, and skill acquisition. *Review of International Economics*.

Mathur, S. K. (2006). Indian Information Technology Industry: Past, Present and Future& A Tool for National Development. Technical Report 2.

Oster, E. and B. M. Steinberg (2013, sep). Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics 104*, 123–135.

Ryoo, J. and S. Rosen (2004, feb). The engineering labor market. *Journal of Political Economy 112*(1).

Santos Silva, J. M. and S. Tenreyro (2006, nov). The log of gravity. *Review of Economics and Statistics 88*(4), 641–658.

Shastry, G. K. (2012). Human capital response to globalization: Education and information technology in India. *Journal of Human Resources 47*(2), 287–330.

Simonovska, I. and M. E. Waugh (2014, jan). The elasticity of trade: Estimates and evidence. *Journal of International Economics 92*(1), 34–50.

Stiglitz, J. E. (1970, may). Factor Price Equalization in a Dynamic Economy. *Journal of Political Economy 78*(3), 456–488.

Tharakan, P. K., I. Van Beveren, and T. Van Ourti (2005, nov). Determinants of India's software exports and goods exports. *Review of Economics and Statistics 87*(4), 776–780.

Tombe, T. and X. Zhu (2019). Trade, migration, and productivity: A quantitative analysis of China. In *American Economic Review*, Volume 109, pp. 1843–1872. American Economic Association.

Tsivanidis, N. (2018). The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio . Technical report.

UC Berkeley (1999). 8.02.99 - UC Berkeley computer scientists attack Y2K bug with new program to find millennium glitches in C applications.

Van Leemput, E. (2016, dec). A Passage to India: Quantifying Internal and External Barriers to Trade. *International Finance Discussion Paper 2016*(1185), 1–65.

Figures

Figure I: **Growth in software exports, IT employment, and engineering enrollment**
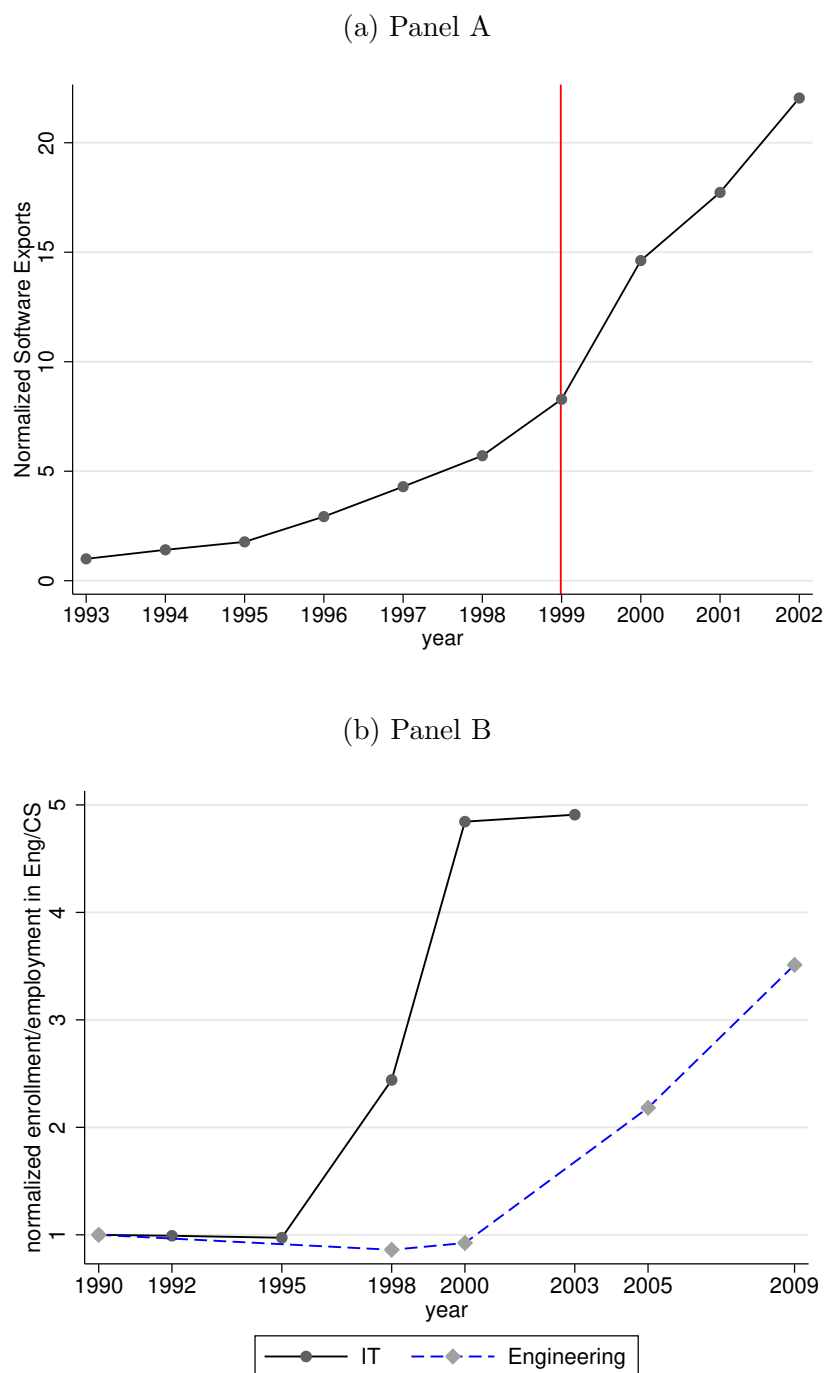
(a) Panel A

(b) Panel B

Notes: Panel A graph shows the growth in IT exports over time, with IT exports normalized to their 1993 levels. This was generated using data on software exports compiled by Richards Heeks. Panel B figure shows the growth in IT employment and engineering enrollment over time, with both IT employment and engineering enrollment normalized to their 1990-1991 levels. This was generated using IT employment data from NSSO, NASSCOM, Economic Census and engineering enrollment data from the Population Census.

Figure II: **Response of IT employment and engineering enrollment across districts with different levels of software exports**
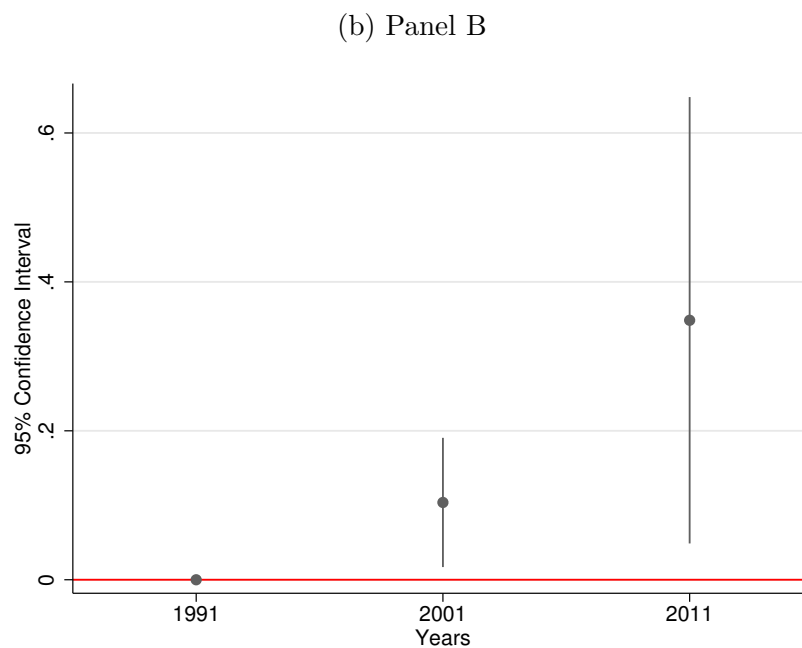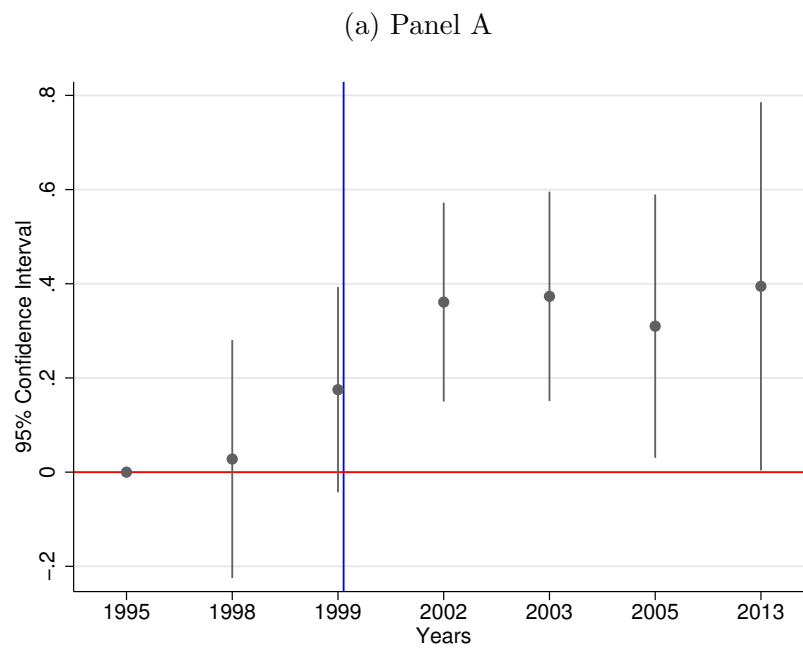
(a) Panel A

(b) Panel B

Notes: These figures plot the confidence intervals for the year by year response of standardized IT employment (panel A) and engineering enrollment (panel B) over the pre and the post boom periods in districts that had any IT exports in 1995 compared to districts that did not. Note that the engineering enrollment, gathered from the Population Census, is only available at decadal intervals.

Figure III: Heterogeneous Response of IT Employment

Notes: This graph plots the confidence intervals for the year by year heterogeneous response of IT employment to differences in historical college enrollment among districts that had any existing level of IT exports over the pre and the post boom periods. The unit is denoted as per '000 engineering students.

Figure IV: **Spatial Distribution of IT Employment and Engineering Enrollment in 2011**

(a) Panel A

■ (0.16,20.75]
■ (0.08,0.16]
■ (0.04,0.08]
■ [0.00,0.04]
□ No data

(b) Panel B

■ (2.12,13.44]
■ (0.96,2.12]
■ (0.51,0.96]
■ [0.09,0.51]
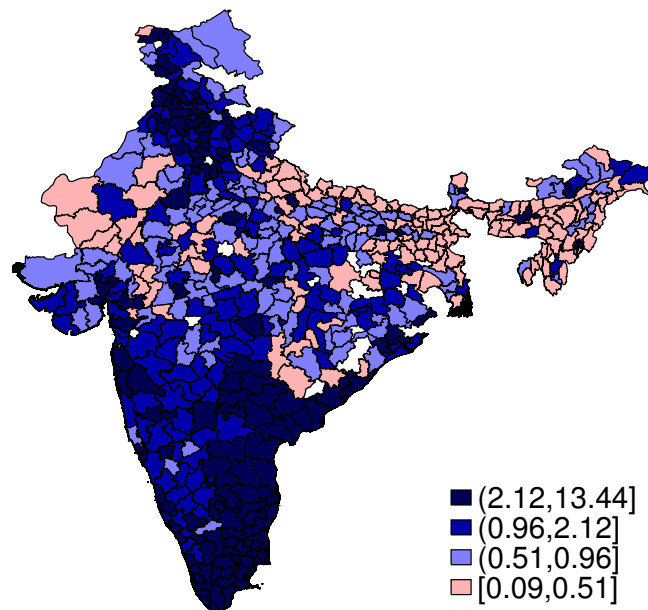
Notes: These figures show the percentage distribution of employment in IT sector out of total employment (panel A) and the percentage distribution of engineering enrollment out of total enrollment (panel B).

Figure V: Histogram of Work-flows by Reason for Migration.
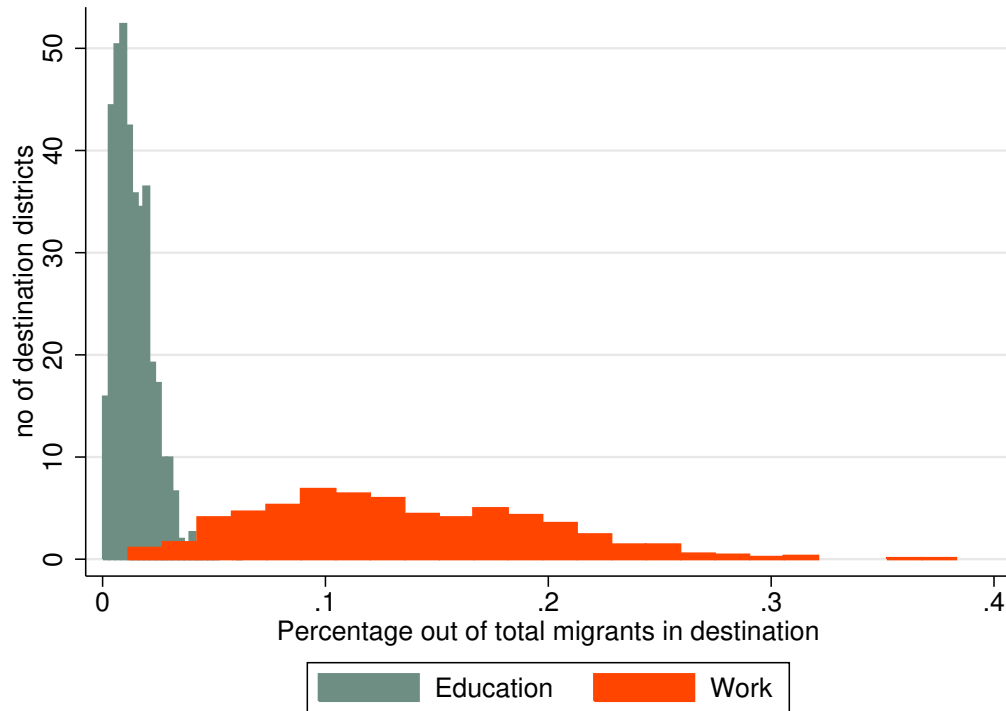
Notes: On the x-axis, this histogram plots the proportion of migrants in a district that migrated for work and education respectively as shown by the pink and blue colors. On the y-axis, the number of destination districts with the corresponding proportions of migrants for work and education are plotted. Data source is the 2001 Census migration data.

Figure VI: Heterogeneous Response of IT Employment. *Units: Per '000 engineering students in own and neighboring districts(own state)*
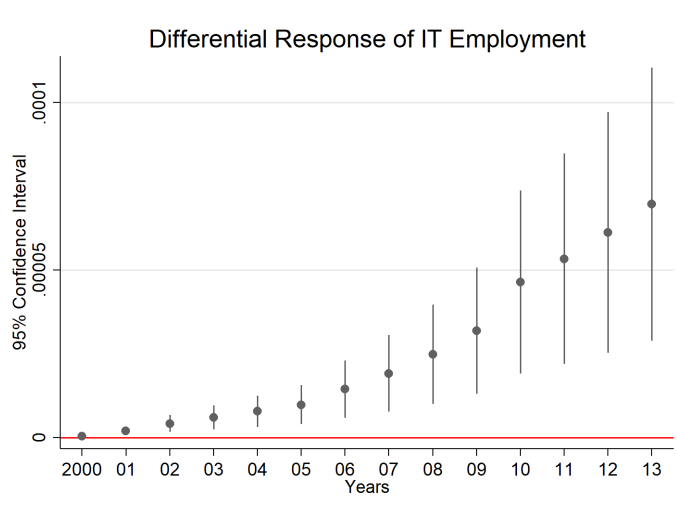
NOTE: This graph plots the confidence intervals for the year by year heterogeneous response of IT employment to differences in historical college enrollment among districts that had any existing level of IT exports over the pre and the post boom periods. The unit is denoted as per '000 engineering students.

Figure VII: Response of IT Employment Across Regions with Different Levels of Software Exports
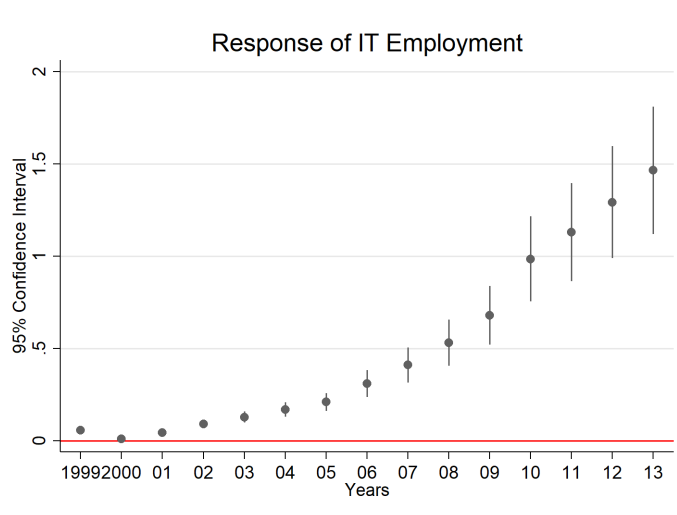
NOTE: This graph plots the confidence intervals for the year by year response of standardized IT employment over the pre and the post boom periods in districts that had any IT exports in 1995 compared to districts that did not using model generated data.

Figure VIII: Estimated migration costs for non-neighboring districts in different states
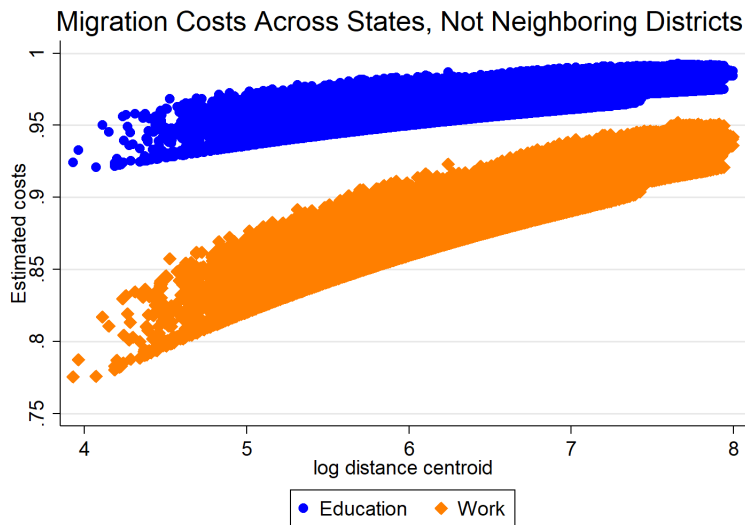
NOTE: This graph plots the log distance between district centers on the x-axis and the estimated costs by reason for migration on the y-axis for non-neighboring districts that fall in different states.

Figure IX: Estimated migration costs for non-neighboring districts in the same state
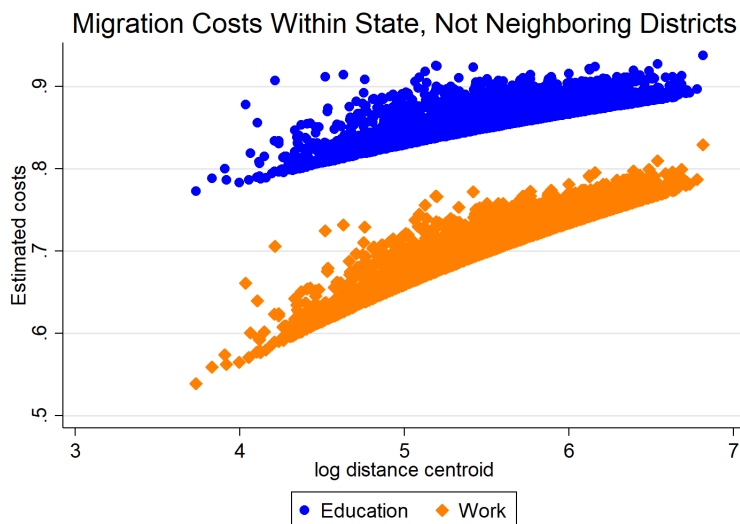
NOTE: This graph plots the log distance between district centers on the x-axis and the estimated costs by reason for migration on the y-axis for non-neighboring districts that fall in different states.

Figure X: Response of Engineering Enrollment Across Regions with Different Levels of Software Exports
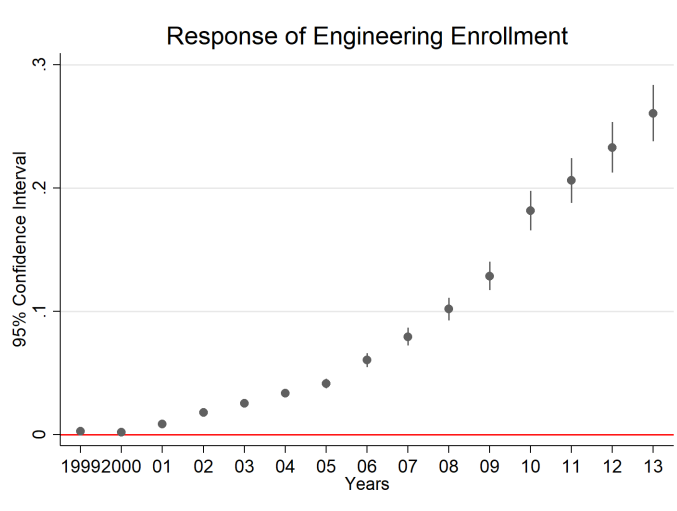
NOTE: This graph plots the confidence intervals for the year by year response of standardized engineering employment over the pre and the post boom periods in districts that had any IT exports in 1995 compared to districts that did not using model generated data

Figure XI: Percentage rise in log Engineering Enrollment from 1991 to 2001, standardized
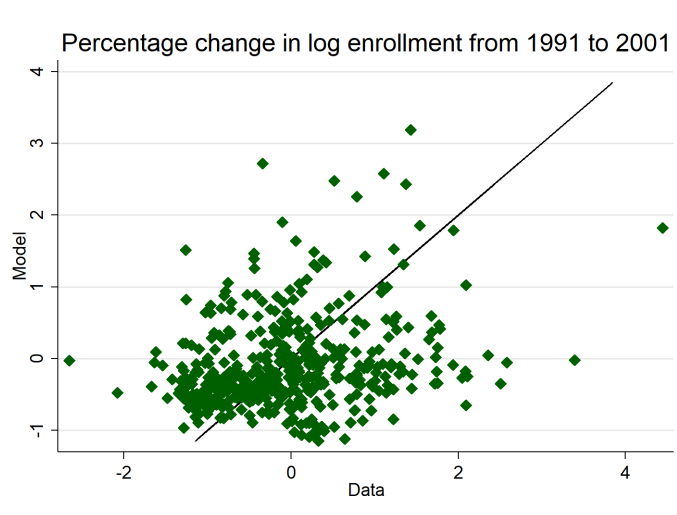
NOTE: This graph plots the percentage change in log engineering enrollment from 1991 to 2001 in the data on the x-axis and the percentage change in log engineering enrollment predicted by the model on the y-axis.

Figure XII: Percentage rise in log Engineering Enrollment from 1991 to 2011, standardized
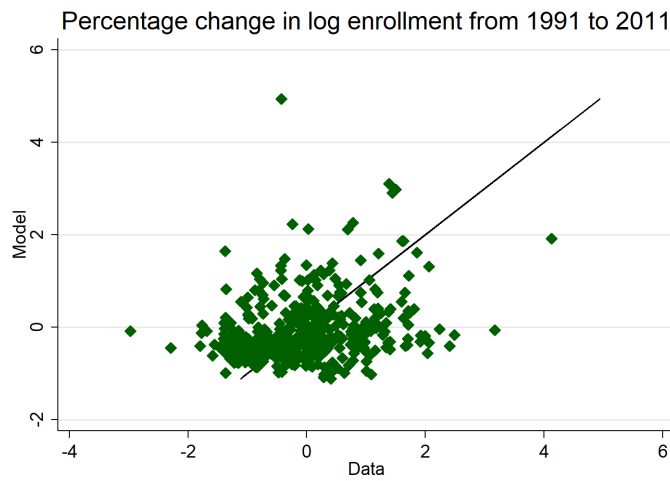
NOTE: This graph plots the percentage change in log engineering enrollment from 1991 to 2011 in the data on the x-axis and the percentage change in log engineering enrollment predicted by the model on the y-axis.

Figure XIII: Distribution of Welfare Gains from the IT Boom
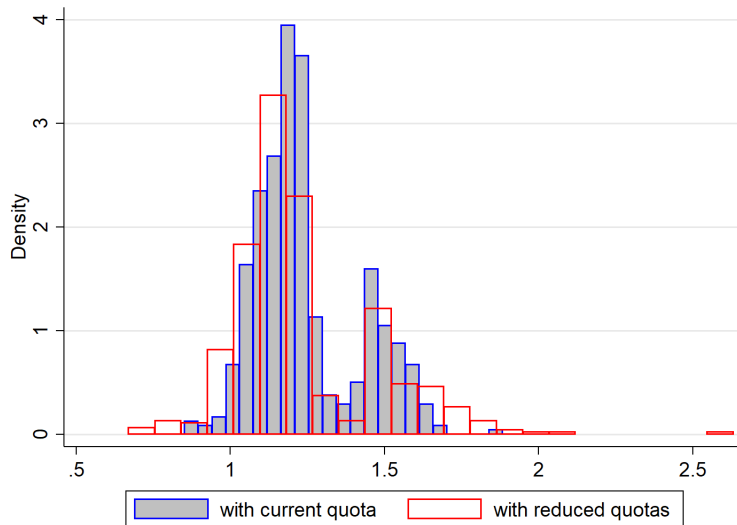
NOTE: The histograms show the distributions of welfare gains from the IT boom in the short and the long run. In the short run education choice is fixed and in the long run education choice is endogenous

Figure XIV: Distribution of Short-run and Long-run Welfare Gains from the IT Boom
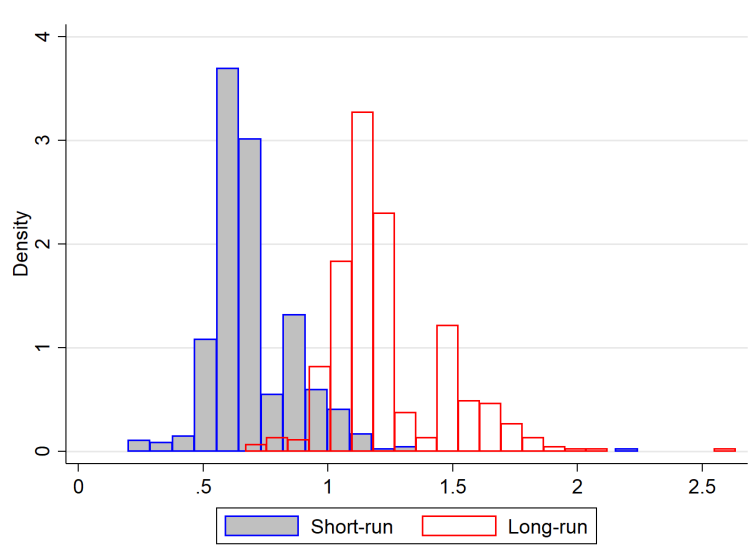
NOTE: The histograms show the distributions of welfare gains from the IT boom in the short and the long run. In the short run education is fixed and in the long run education choice is endogenous

Tables

Table I: **PPML gravity estimation on district to district migration by reason for migration**

	(1) Education	(2) Work	(3) Other reasons
log distance between district centers	-0.585***	-0.567***	-0.752***
	(-71.67)	(-48.87)	(-60.02)
common language	0.656***	0.478***	0.335***
	(8.05)	(5.11)	(4.52)
Same state; neighboring districts	3.577***	3.002***	3.126***
	(64.25)	(39.58)	(35.16)
Same state; not neighboring districts	2.559***	2.088***	1.935***
	(47.51)	(29.10)	(25.72)
Different state, neighboring districts	2.422***	2.737***	2.845***
	(32.31)	(37.56)	(33.66)
N	342225	342225	342225

NOTE: The table shows the PPML estimation results, differentiated by reason for migration. t-statistics are reported in parenthesis.

Table II: Migration Flows by Reason for Migration.

Reason for Migration	No. of Migrants	Percentage	Out of State	Percentage
Work	18,901,992	48	9,771,841	52
Education	11,507,98	3	3,59,029	31
Other	19,746,588	49	72,00,884	36
Total	39,799,378	100	17,331,754	44

NOTE: Column 1 lists the reason for migration. Column 2 lists the number of people migrating out of their district of previous residence by reason for migration in 2001. Column 3 shows the percentage distribution of migrants by reason for migration. Column 4 shows the number of people who migrated out of their own state of birth by reason for migration. Column 5 shows the percentage of people migrating out of their state of birth among the total number of migrants by reason for migration. Data source is the 2001 Census migration data.

Table III: **Estimation of response of district to district migration flows to distances by reason for migration**

	(1)	(2)	(3)
	Education	Work: Traditional	Work:NLS
log distance between district centers	-.602***	-.554***	-.885***
	(-62.04)	(-55.55)	(-34.27)
common language	.307 ***	.393 ***	.145***
	(17.77)	(3.00)	(1.90)
Same state; neighbors	3.646***	3.158***	4.803***
	(45.48)	(36.13)	(14.68)
Same state; not neighbors	2.379***	2.125***	2.776***
	(27.49)	(23.39)	(6.890)
Different state, neighbors	2.339***	2.737***	3.763***
	(17.16)	(35.17)	(7.515)
N	280900	280900	280900

NOTE: The table shows the PPML estimation results, differentiated by reason for migration. t-statistics are reported in parenthesis. Columns 1 and 2 reports the traditional PPML estimation results. Column 3 reports the results for work migration estimation using non linear least squares.

Table IV: **Trade Cost Estimation**

	(1)	(2)	(3)	(4)
	Exports	Exports	Exports	Exports
	OLS	IV	IV	State-level trend
Observable MC	-0.44 ***	-0.38***	-.36***	-.45***
	(-10.96)	(-7.88)	(-9.53)	(-6.66)
State-time Trend	Yes	No	No	Yes
Remoteness as control	No	No	Yes	Yes
IV	No	Yes	Yes	Yes
First Stage F-stat:		210.37***	116.20***	47.66***
N	523	523	523	523

Robust standard errors are used. t statistics are reported in parenthesis.

Table V: **First Stage**

	(1)	(2)	(3)	(4)
	Observable MC	Observable MC	Observable MC	Observable MC
	OLS	IV	IV	State-level trend
IV	-	-2.40e-08***	-2.31e-08***	-2.46e-08***
	-	(-13.12)	(-9.53)	(-6.44)
State-time Trend	Yes	No	No	Yes
Remoteness as control	No	No	Yes	Yes
IV	No	Yes	Yes	Yes
First Stage F-stat:		210.37***	116.20***	47.66***
N	523	523	523	523

Robust standard errors are used. t statistics are reported in parenthesis.

Table VI: **Summary of estimated parameter values**

Parameter	Value	Source
Productivity dispersion (θ)	2.61	Estimated
Education amenity dispersion (ζ)	2.42	Estimated
IT trade elasticity (σ_{IT})	1.27	Estimated
ρ_S[*]	1.41	Katz and Murphy (1992)
ρ_{hS}[*]	2	Ryoo and Rosen (2004)
Internal trade elasticity (σ_S)	5	Simonovska and Waugh (2014)
Agriculture share	.38	Ministry of Statistics, Govt of India
Manufacturing share	.16	-
High-skill services	.07	-
Other services	.37	-

Part I

Appendix for online publication only

Table of Contents

A   Data Appendix

A.1   Education and labor market data

Table VII: Summary statistics of daily district-wise wages in Rupees

	Pre Boom		Post Boom	
Variable	Mean	Std. Dev.	Mean	Std. Dev
Labor market: Wages				
High-skill services				
College-educated	363.64	336.98	467.18	243.15
Non college-educated	171.02	80.28	275.17	235.45
Manufacturing				
College-educated	324.25	570.00	307.96	209.08
Non college-educated	115.08	111.37	152.21	104.28

Table VIII: Summary statistics of annual district-wise employment

	Pre Boom		Post Boom	
Variable	Mean	Std. Dev.	Mean	Std. Dev
Labor market: Employment				
High-skill services				
College-educated	7256	16526	9664	16125
Non college-educated	7574	13553	11750	204375
Manufacturing				
College-educated	25276	43491	30799	49503
Non college-educated	34231	61375	44380	66592

Table IX: Summary statistics of annual district-wise enrollment

Education	Pre Boom	Post Boom
Engineers/college enrollment	5.10%	11.45%
College enrollment	12,404	32,632

A.2   Linguistic distance

I summarize the way Shastry (2012) described the construction of this index in her paper. The 1961 Census of India documented speakers of 1652 languages from five language

families. Much linguistic diversity is between districts. A district's primary language is native to 83 percent of residents on average, ranging from 22 percent to 100 percent. Most people thus adopt a second language that is a widely accepted speaking medium across districts. Of all multi-linguals who were not native speakers, 60 percent chose to learn Hindi and 56 percent chose English. Think of an individual who is not a native Hindi speaker. Given everything else, whether this individual learns Hindi or English as a second language depends on the relative costs of learning that language, which in turn depends on her mother tongue. Someone whose mother tongue is similar (not similar) to Hindi will find Hindi easier (more difficult) to learn relative to English. To quantify what is the relative cost of learning Hindi or English , Shastry constructed three measures of linguistic distance of each native language from Hindi.The first measure classifies languages into five "degrees" of linguistic distance from Hindi based on cognates, grammar, and syntax (see Table 2). The second measure is the percent of words from a core list that are cognates of Hindi words. The third measure is based on language family trees from the Ethnologue database. These measures are highly correlated: 0.935 between degrees and percent cognates and 0.903 between degrees and nodes.

Table X: Measures of linguistic distance to hindi

Sample Language	Degrees	% Cognates	Nodes	% Native Speakers
0 Degrees	0	100	0	.456
Hindi, Urdu				
1 Degree	1	67.1	5	0.084
Gujarati, Punjabi, Rajasthani				
2 Degrees	2	56.4	6.5	.076
Konkani, Marathi				
3 Degrees	3	64.1	7	.133
Assamese, Bengali, Bihari, Oriya				
4 Degrees	4	53.3	7.3	0.005
Kashmiri, Sindhi, Sinhalese				
5 Degrees	5	5	12.5	.244
All non-Indo European Languages				

Source: Gauri Kartini Shashtry, 2012

From the 1991 census of India, Shastry calculates a district's linguistic distance from Hindi in two ways-1) the population weighted average distance of all native languages from Hindi and 2) the population share of languages at least 3 degrees away from Hindi. All my analysis that follows is conducted with measure 2) but the analysis are robust to using measure 1) instead. Shastry proxies English-learning costs as linguistic distance from Hindi. One may think the natural proxy is linguistic distance from English, but it is the relative costs of learning Hindi and English that should determine which language one learns. A native Hindi speaker can choose to learn English as a second language at a much lower cost than a non-native speaker whose language is close to Hindi. So

there is a non-monotonicity in the relationship- native Hindi speakers are more likely to learn English but speakers of languages close to Hindi learn Hindi rather than English. Then as distance to Hindi rises, the probability of learning English as a second language rises except for at distance 0. Shastry (2012) shows that such a relation holds. From now on, I would use linguistically distant to Hindi and linguistically closer to English interchangeably.

A.3   Missing value imputation

The NSS is a sample as opposed to the Census which is a complete enumeration. In the NSS, individuals are included in the sample so that it is representative. However, there are many observations in the NSS where no individual working in sector S, having a degree in s in district d has been interviewed, even though according to the Census there are individuals working in sector S, having a degree in s in district d. I use a weighted knn where the weights can be uniform, ie, only the reciprocal of the distance or the weights could be Gaussian Kernel, Epanechnikov, Cosine etc (isotropic Kernels) to impute missing values. This machine learning technique involves using a training data to choose the value of "k" that minimizes the sum of squared distances between the actual and predicted values, where the predicted values are obtained by taking a weighted average of the variables values of the "k" nearest neighbors. I have used the uniform Kernel as weight here.

B   Reduced Form Facts

B.1   Stylized facts 1 and 2

The table below reports the regression results for total IT employment graph reported in the reduced form facts section.

	(1)
	Standardized Employment
year=1995 × Standardized Historical Export	0.000
	(.)
year=1998 × Standardized Historical Export	0.028
	(0.22)
year=1999 × Standardized Historical Export	0.175
	(1.58)
year=2002 × Standardized Historical Export	0.361***
	(3.37)
year=2003 × Standardized Historical Export	0.373***
	(3.31)
year=2005 × Standardized Historical Export	0.310**
	(2.18)
year=2013 × Standardized Historical Export	0.395**
	(1.99)
Constant	-0.071***
	(-3.57)
N	3880.000

Table XIV: **Response of IT employment, with historical software exports predicted by linguistic distances**

	(1)
	Standardized Employment
year=1995 × Standardized Historical Export	0.000
	(.)
year=1998 × Standardized Historical Export	0.033
	(0.08)
year=1999 × Standardized Historical Export	0.318
	(0.97)
year=2002 × Standardized Historical Export	0.645**
	(1.91)
year=2003 × Standardized Historical Export	0.667*
	(1.88)
year=2005 × Standardized Historical Export	0.665**
	(1.92)
year=2013 × Standardized Historical Export	0.955
	(1.58)
Constant	-0.08***
	(-3.47)
N	3880.000

Table XI: **Response of the IT sector**

	(1)	(2)	(3)	(4)
	total employment	total employment	total employment	total employment
1998 × Software Exports 1995	0.028			0.028***
	(0.22)			(5.73)
1999 × Software Exports 1995	0.175			0.175***
	(1.58)			(4.28)
2002 × Software Exports 1995	0.361***			0.361***
	(3.37)			(3.10)
2003 × Software Exports 1995	0.373***			0.373***
	(3.31)			(2.79)
2005 × Software Exports 1995	0.310**			0.310
	(2.18)			(1.54)
2013 × Software Exports 1995	0.395**			0.395
	(1.99)			(1.63)
1998 × AnyExports		0.129**	0.129**	
		(2.40)	(2.14)	
year=1999 × AnyExports		0.932***	0.932***	
		(3.11)	(3.25)	
year=2002 × AnyExports		1.858***	1.858***	
		(2.81)	(2.91)	
year=2003 × AnyExports		1.823***	1.823**	
		(2.61)	(2.65)	
year=2005 × AnyExports		1.773**	1.773**	
		(2.31)	(2.59)	
year=2013 × AnyExports		3.849**	3.849**	
		(2.19)	(2.42)	
Clustering	District	District	State	State
Observations	3880	3880	3880	3880

Table XII: **Response of the Education Sector**

	(1)	(2)	(3)	(4)	(5)	(6)
	Engineers	Engineers	Engineers	Proportion of Engineers	Proportion	Proportion
AnyExports	0.000	0.000	0.000			
	(.)	(.)	(.)			
2001 × Any Exports	1.449	1.449**	1.449**			
	(1.40)	(2.43)	(2.23)			
2011 × Any Exports	4.998***	4.998**	4.998**			
	(3.08)	(2.51)	(2.18)			
Historical Software Exports				0.000	0.000	0.000
				(.)	(.)	(.)
2001 × Historical Software Exports				0.104	0.104**	0.104**
				(1.56)	(2.39)	(2.43)
2011 × Historical Software Exports				0.348***	0.348**	0.348**
				(3.33)	(2.23)	(2.36)
N	1726	1726	1726	1724	1724	1724

Table XIII: **Heterogeneous Response of IT Employment**

	(1)	(2)	(3)	(4)	(5)	(6)
	total employment	total employment	total employment	total employment	total employment	total employment
year=1998 × AnyExports × Engineers 1991	0.012*	0.012	0.012*			
	(1.85)	(1.51)	(1.84)			
year=1999 × AnyExports × Engineers 1991	0.018***	0.018***	0.018***			
	(3.00)	(2.95)	(2.97)			
year=2002 × AnyExports × Engineers 1991	0.041***	0.041**	0.041**			
	(2.76)	(2.47)	(2.50)			
year=2003 × AnyExports × Engineers 1991	0.045***	0.045***	0.045**			
	(2.75)	(2.60)	(2.65)			
year=2005 × AnyExports × Engineers 1991	0.045***	0.045**	0.045**			
	(2.76)	(2.41)	(2.51)			
year=2013 × AnyExports × Engineers 1991	0.052	0.052	0.052			
	(1.25)	(1.09)	(1.29)			
year=1998 × Historical Software Exports × Engineers 1991				0.000	0.000	0.000**
				(0.92)	(0.41)	(2.28)
year=1999 × Historical Software Exports × Engineers 1991				0.001**	0.001**	0.001**
				(2.26)	(2.53)	(2.24)
year=2002 × Historical Software Exports × Engineers 1991				0.004***	0.004***	0.004***
				(4.70)	(6.56)	(6.09)
year=2003 × Historical Software Exports × Engineers 1991				0.005***	0.005***	0.005***
				(5.53)	(7.79)	(7.39)
year=2005 × Historical Software Exports × Engineers 1991				0.005***	0.005***	0.005***
				(3.95)	(2.95)	(3.68)
year=2013 × Historical Software Exports × Engineers 1991				-0.002	-0.002	-0.002
				(-0.93)	(-0.62)	(-0.77)
Clustering	State time	District	State	State time	District	State
N	3838	3838	3838	3838	3838	3838

Table XV: **Predicting historical software exports**

	(1)
	Standardized Employment
Hindi speakers	0.115***
	(2.62)
English speakers	1020.29***
	(23.46)
linguistic distance	0.007
	(0.614)
Constant	-0.18***
	(-3.30)
N	2731
F(3,2727)	184.29
Adjusted R-squared	.167

Table XVI: **Gravity estimation by reason for migration, replicating table 8 from Kone et al**

	(1)	(2)	(3)	(4)
	Marriage	Work or Business	Move with Family	Education
log distance_centroid	-1.626***	-1.481***	-1.454***	-1.207***
	(-48.78)	(-50.19)	(-44.60)	(-47.62)
common	1.077***	0.499***	0.941***	1.338***
	(15.65)	(7.01)	(15.40)	(19.14)
Same state,; neighbors	2.265***	1.508***	1.681***	2.359***
	(40.98)	(24.20)	(24.41)	(43.37)
Same state; not neighbors	0.881***	1.226***	1.148***	1.777***
	(19.09)	(22.52)	(20.72)	(38.25)
Different state; neighbors	2.167***	1.054***	1.368***	1.036***
	(34.24)	(14.27)	(18.39)	(12.86)
N	341640	341640	341640	341640

t-stats reported in parenthesis

C   Theoretical Derivations

C.1   Worker's problem

Problem of worker i educated in o_2 in field f who goes to work in d in sector S is given by: Max $\Pi_S C_S^{\alpha_S}$ where $C_S = (\sum_k c_{kdS}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$

s.t $\sum_k \sum_{S'} p_{kdS'} c_{kdS'} = W_{f,dS} \eta_{io_2 f,dS} \mu_{o_2 d}$

This yields : Consumption of variety k of good S' for an individual who got his degree in o_2 and moved to d to work in occupation S is given by:

$$c_{ikf,dS'} = (p_{kdS'})^{-\sigma} P_{dS'}^{\sigma-1} (\alpha_{S'} W_{f,dS'} \eta_{ikf,dS'} \mu_{o_2 d})$$

Assuming ice-berg transportation cost:

$$p_{kdS'} = \tau_{kdS'} p_{kS'}$$

Consumption of variety k of good S' for an individual who got his degree in o_2 and moved to d is given by:

$$c_{ikf,dS'} = (\tau_{kdS'} p_{kS'})^{-\sigma} P_{dS}^{\sigma-1} (\alpha_{S'} W_{f,dS} \eta_{if,dS} \mu_{o_2 d})$$

Using the above quantities, worker indirect utility in stage 2 is derived as:

$$U_{io_2 f,dS} = \frac{W_{f,dS} u_d \eta_{if,dS} \mu_{o_2 d}}{\Pi_S P_{dS}^{\alpha_S}} \tag{36}$$

We can derive the indirect utility for stage 1 very similarly and this gives a combined stage 1 and stage 2 utility of the following form:

$$U_{io_2 f,dS} = \left(\mu_{o_2 d} a_{o_2 s} \zeta_{io_2 s} \right) \left(\frac{W_{f,dS} u_d \eta_{if,dS} \mu_{o_2 d}}{\Pi_S P_{dS}^{\alpha_S}} \right) \tag{37}$$

where $P_{dS} = \sum_d \tau_{kdS} p_{kS}$

C.2 Firm's problem

The firm profit maximization condition for sector S is given by:

$$max_{L_{sdS} \forall d,s} P_{dS} Q_{dS} - \sum_s w_{sdS} \tilde{L}_{sdS} - w_{sdS} \tilde{L}_{sdS}$$

where $Q_{dS} = (Q_{hdS}^{\frac{\rho_S - 1}{\rho_S}} + Q_{ldS}^{\frac{\rho_S - 1}{\rho_S}})^{\frac{\rho_S}{\rho_S - 1}}$

$Q_{hdS} = (\sum_{s \in college, k} A_{sdS} (\tilde{L}_{sdS})^{\frac{\rho_{hS} - 1}{\rho_{hS}}})^{\frac{\rho_{hS}}{\rho_{hS} - 1}}$

and in theory we can have, $Q_{ldS} = (\sum_{s \in nocollege} A_{sdS} (\tilde{L}_{sdS})^{\frac{\rho_{lS} - 1}{\rho_{lS}}})^{\frac{\rho_{lS}}{\rho_{lS} - 1}}$

For this paper, I use only one type of unskilled labor. And thus, here $Q_{ldS} = (A_{ldS}\tilde{L}_{ldS})$

Differentiating with respect to $\tilde{L}_{s,S,k'}$ where s=e or ne,

$$P_{dS}\frac{\rho_S}{\rho_S-1}Q_{dS}^{\frac{1}{\rho_S}}\frac{\rho_S-1}{\rho}Q_{hdS}^{\frac{-1}{\rho_S}}\frac{\rho_{h,S}}{\rho_{h,S}-1}(\sum_{s\in college,k}A_{sdS}(\tilde{L}_{sdS})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}})^{\frac{\rho_{h,S}}{\rho_{h,S}-1}-1}A_{sdS}\frac{\rho_{h,S}-1}{\rho_{h,S}}(\tilde{L}_{sdS})^{\frac{-1}{\rho_h}}$$
$$= w_{k',s,S}$$

Simplifying ,

$$P_{dS}Q_{dS}^{\frac{1}{\rho_S}}Q_{hdS}^{\frac{-1}{\rho_S}}(\sum_{s\in college,k}A_{sdS}(\tilde{L}_{sdS})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}})^{\frac{1}{\rho_{h,S}-1}}A_{sdS}(\tilde{L}_{sdS})^{\frac{-1}{\rho_h}}$$
$$= w_{sdS}$$

Simplifying further, $P_{dS}Q_{dS}^{\frac{1}{\rho_S}}Q_{hdS}^{\frac{1}{\rho_{h,S}}-\frac{1}{\rho_S}}A_{sdS}(\tilde{L}_{sdS})^{\frac{-1}{\rho_h}} = w_{sdS}$ In the empirical model, we use engineers and non-engineers as two types of skilled labor. Denoting s=e and s=ne for engineers and non-engineers respectively, one can derive the following foc:

$$\frac{A_{e,d,S}L_{e,d,S}^{\frac{-1}{\rho_{h,S}}}}{A_{ne,d,S}L_{ne,d,S}^{\frac{-1}{\rho_{h,S}}}} = \frac{w_{e,d,S}}{w_{ne,d,S}} \tag{38}$$

Under the assumption that all productivities are drawn from the same Frechet distribution, and firms do not know worker productivities,the foc does not contain effective labor, only labor. We thus get the following estimating equation:

$$\frac{A_{e,d,S}L_{e,d,S}^{\frac{-1}{\rho_{h,S}}}}{A_{ne,d,S}L_{ne,d,S}^{\frac{-1}{\rho_{h,S}}}} = \frac{w_{e,d,S}}{w_{ne,d,S}}$$

Denote $(p_{dS}^h)^{1-\rho_S} = (p_{dS}^h)^{1-\rho_S} + (p_{dS}^l)^{1-\rho_S}$

where, $p_{dS}^l = A_{ldS}w_{ldS}$

where, $p_{dS}^{h\,(1-\rho_{hS})} = \sum_s A_{sdS}^{\rho_{h,S}}w_{sdS}^{(1-\rho_{h,S})}$

From firm first order condition for high-skilled labor, we can rewrite it as:

$$p_{dS}^{h}{}^{(1-\rho_{h,S})} = \sum_{s} A_{sdS}^{\rho_{h,S}} (P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{\frac{1}{\rho_{h,S}}-\frac{1}{\rho_S}} A_{sdS} (\tilde{L}_{sdS})^{\frac{-1}{\rho_h}})^{(1-\rho_{h,S})}$$

$$= (P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{\frac{1}{\rho_{h,S}}-\frac{1}{\rho_S}})^{(1-\rho_{h,S})} (\sum_{s} A_{sdS}^{\rho_{h,S}} (A_{sdS} (\tilde{L}_{sdS})^{\frac{-1}{\rho_{h,S}}})^{(1-\rho_{h,S})})$$

$$= (P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{\frac{1}{\rho_{h,S}}-\frac{1}{\rho_S}})^{(1-\rho_{h,S})} \sum_{s} (A_{sdS} \tilde{L}_{sdS}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}})$$

$$= (P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{\frac{1}{\rho_{h,S}}-\frac{1}{\rho_S}} Q_{hdS}^{\frac{-1}{\rho_{h,S}}})^{(1-\rho_{h,S})}$$

$$= (P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{-\frac{1}{\rho_S}})^{(1-\rho_{h,S})}$$

Thus we get the following equation for high-skilled,

$$p_{dS}^{h} = P_{dS}(Q_{dS}^{\frac{1}{\rho_S}} Q_{hdS}^{-\frac{1}{\rho_S}}) \tag{39}$$

I now solve the foc for low skilled workers.

$$\text{Max}_{L_{ldS}} \ P_{dS}(Q_{hdS}^{\frac{\rho_S-1}{\rho_S}} + (A_{k',S,l} L_{k',S,l})^{\frac{\rho_S-1}{\rho_S}})^{\frac{\rho_S}{\rho_S-1}} - \sum_s w_{s,S,k'} \tilde{L}_{s,S,k'} - w_{l,S,k'} \tilde{L}_{l,S,k'}$$

For low-skilled, taking the first order condition, we get,

$$w_{ldS} = P_{dS} \frac{\rho_S}{\rho_S-1}(Q_{hdS}^{\frac{\rho_S-1}{\rho_S}} + (A_{ldS} L_{ldS})^{\frac{\rho_S-1}{\rho_S}})^{\frac{1}{\rho_S-1}} \frac{\rho_S-1}{\rho_S}(A_{ldS} L_{ldS})^{\frac{-1}{\rho_S}} A_{ldS}$$

$$= P_{dS} Q_{dS}^{\frac{1}{\rho_S}} (A_{ldS} L_{ldS})^{\frac{-1}{\rho_S}} A_{ldS}$$

$$= P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{ldS}^{\frac{-1}{\rho_S}} A_{ldS}$$

Since $p_{dS}^{l} = \frac{w_{ldS}}{A_{ldS}}$

Thus, $p_{dS}^{l} = P_{dS} Q_{dS}^{\frac{1}{\rho_S}} Q_{ldS}^{\frac{-1}{\rho_S}}$ Combining the two, we get the following equation:

$$\frac{p_{dS}^{h}}{p_{dS}^{l}} = \frac{Q_{hdS}^{\frac{-1}{\rho_S}}}{Q_{ldS}^{\frac{-1}{\rho_S}}} \tag{40}$$

Thus,

$$ln(\frac{p^h_{dS}}{p^l_{dS}}) = \frac{-1}{\rho_S}ln(\frac{Q_{hdS}}{Q_{ldS}}) \tag{41}$$

Note however, that these are not observable quantities due to the presence of unobserved productivity.

$$ln(\frac{(\sum_s A^{\rho_{h,S}}_{sdS} w^{(1-\rho_h S)}_{sdS})^{\frac{1}{(1-\rho_{h,S})}}}{(A^{\rho_l}_{ldS}w_{ldS})}) = \frac{-1}{\rho_S}ln(\frac{(\sum_{s\in college,k} A_{sdS}(\tilde{L}_{sdS})^{\frac{\rho_{hS}-1}{\rho_{hS}}})^{\frac{\rho_{hS}}{\rho_{hS}-1}}}{A_{ldS}\tilde{L}_{ldS}}) \tag{42}$$

For ease of notation, I now use s=e and s=ne for engineers and non-engineers respectively.

$$ln(\frac{(A_{ne,d,S})^{\frac{-\rho_{h,S}}{\rho_{h,S}-1}}((\frac{A_{e,d,S}}{A_{ne,d,S}})^{\rho_{h,S}}(w_{e,d,S})^{1-\rho_{h,S}} + w^{1-\rho_{h,S}}_{ne,d,S})^{\frac{1}{1-\rho_{h,S}}}}{(A^{-1}_{ldS}w_{ldS})})$$

$$= ln(\frac{(A_{ne,d,S})^{\frac{\rho_{h,S}}{\rho_{h,S}-1}}((\frac{A_{e,d,S}}{A_{ne,d,S}})(\tilde{L}_{e,d,S})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}} + \tilde{L}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}}_{ne,d,S})^{\frac{\rho_{h,S}}{1-\rho_{h,S}}}}{(A_{l,d,S}\tilde{L}_{l,d,S})}) \tag{43}$$

This implies,

$$ln(\frac{((\frac{A_{e,d,S}}{A_{ne,d,S}})^{\rho_{h,S}}(w_{e,d,S})^{1-\rho_{h,S}} + w^{1-\rho_{h,S}}_{ne,d,S})^{\frac{1}{1-\rho_{h,S}}}}{(w_{l,s,k'})}) - \frac{\rho_{h,S}}{\rho_{h,S}-1}ln(A_{ne,S,k'}) - ln(A_{l,S,k'})$$

$$= (\frac{-1}{\rho_S})ln(\frac{((\frac{A_{e,d,S}}{A_{ne,d,S}})(\tilde{L}_{e,d,S})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}} + \tilde{L}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}}_{ne,d,S})^{\frac{\rho_{h,S}}{1-\rho_{h,S}}}}{\tilde{L}_{ldS}}) - \frac{1}{\rho_S}ln(A_{ldS}) - \frac{1}{\rho_S}(\frac{\rho_{h,S}}{\rho_{h,S}-1})ln(A_{ne,d,S})$$

Thus,

$$ln\left(\frac{\left(\left(\frac{A_{e,d,S}}{A_{ne,d,S}}\right)^{\rho_{h,S}}(w_{e,d,S})^{1-\rho_{h,S}} + w_{ne,d,S}^{1-\rho_{h,S}}\right)^{\frac{1}{1-\rho_{h,S}}}}{(w_{ldS})}\right)$$

$$= \left(\frac{-1}{\rho_S}\right)ln\left(\frac{\left(\left(\frac{A_{e,d,S}}{A_{ne,d,S}}\right)(\tilde{L}_{e,d,S})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}} + \tilde{L}_{ne,d,S}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}}\right)^{\frac{\rho_{h,S}}{1-\rho_{h,S}}}}{\tilde{L}_{ldS}}\right)$$

$$- \frac{1}{\rho_S}ln(A_{ldS}) - \frac{1}{\rho_S}\left(\frac{\rho_{h,S}}{\rho_{h,S}-1}\right)ln(A_{ne,d,S})$$

$$+ \frac{\rho_{h,S}}{\rho_{h,S}-1}ln(A_{ne,d,S}) - ln(A_{ldS})$$

$$= \left(\frac{-1}{\rho_S}\right)ln\left(\frac{\left(\left(\frac{A_{e,d,S}}{A_{ne,d,S}}\right)(\tilde{L}_{e,d,S})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}} + \tilde{L}_{ne,d,S}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}}\right)^{\frac{\rho_{h,S}}{1-\rho_{h,S}}}}{\tilde{L}_{ldS}}\right)$$

$$+ \left(\frac{\rho_{h,S}}{\rho_{h,S}-1}\right)\left(1 - \frac{1}{\rho_S}\right)ln(A_{ne,d,S}) + \left(1 - \frac{1}{\rho_S}\right)ln(A_{ldS})$$

$$= \left(\frac{-1}{\rho_S}\right)ln\left(\frac{\left(\left(\frac{A_{e,d,S}}{A_{ne,d,S}}\right)(\tilde{L}_{e,d,S})^{\frac{\rho_{h,S}-1}{\rho_{h,S}}} + \tilde{L}_{ne,d,S}^{\frac{\rho_{h,S}-1}{\rho_{h,S}}}\right)^{\frac{\rho_{h,S}}{1-\rho_{h,S}}}}{\tilde{L}_{ldS}}\right)$$

$$+ \left(1 - \frac{1}{\rho_S}\right)\frac{\frac{\rho_{hS}}{1-\rho_{hS}}A_{ne,d,S}}{A_{ldS}}$$

Note that, all the quantities in this equation are observable. If we plugin the first order condition 38, this is a regression of known quantities with the unobserved productivities as residuals.

Recover IT prices then non-IT Use the following equation for S=IT $(p_{S,k'})^{1-\rho_S} = (p_{S,k'}^h)^{1-\rho_S} + (p_{S,k'}^l)^{1-\rho_S}$

where $p_{S,k'}^{h}{}^{(1-\rho_h)} = \sum_s A_{k',s,S}^{\rho_{h,S}} w_{k',s,S}^{(1-\rho_{h,S})}$

and $p_{S,k'} = A_{u,l}w_{l,S,k'}$

Thus we can write price as:

$$(p_{S,k'})^{1-\rho_S} = ((A_{k',e,S}^{\rho_{h,S}} w_{k',e,S}^{(1-\rho_{h,S})} + A_{k',ne,S}^{\rho_{h,S}} w_{k',ne,S}^{(1-\rho_{h,S})}))^{\frac{1-\rho_S}{1-\rho_{h,S}}} + (A_{k',l,S}^{-1} w_{l,S,k'})^{(1-\rho_S)}$$

$$= A_{k',ne,S}^{\frac{\rho_{h,S}(1-\rho_S)}{1-\rho_{h,S}}} ((\frac{A_{k',e,S}^{\rho_{h,S}}}{A_{k',ne,S}^{\rho_{h,S}}} w_{k',e,S}^{(1-\rho_{h,S})} + w_{k',ne,S}^{(1-\rho_{h,S})})^{\frac{1-\rho_S}{1-\rho_{h,S}}} + (\frac{A_{k',ne,S}^{\frac{\rho_{h,S}(\rho_S-1)}{\rho_{h,S}-1}}}{A_{k',l,S}} w_{l,S,k'})^{1-\rho_S})$$

$$= A_{k',ne,S}^{\frac{-\rho_{h,S}(\rho_S-1)}{1-\rho_{h,S}}} (\tilde{p}_{S,k'}^{h}{}^{1-\rho_S} + (\frac{A_{k',ne,S}^{\frac{\rho_{h,S}}{\rho_{h,S}-1}}}{A_{k',l,S}} w_{l,S,k'})^{1-\rho_S})$$

$$= A_{k',ne,S}^{\frac{-\rho_{h,S}(\rho_S-1)}{1-\rho_{h,S}}} (\tilde{p}_{S,k'}^{h}{}^{1-\rho_S} + (\tilde{x}_{l,S,k'}^{(\frac{\rho_S}{\rho_S-1})} w_{l,S,k'})^{1-\rho_S})$$

$$= A_{k',ne,S}^{\frac{-\rho_{h,S}(\rho_S-1)}{1-\rho_{h,S}}} (\tilde{p}_{S,k'}^{h}{}^{1-\rho_S} + \tilde{x}_{l,S,k'}^{-\rho_S} (w_{l,S,k'})^{1-\rho_S})$$

The term in the bracket is a function of known quantities. How?

$\tilde{p}_{S,k'}^{h}{}^{1-\rho_S} = ((\frac{A_{e,S,k'}}{A_{ne,S,k'}})^{\rho_{h,S}} w_{e,S,k'}^{1-\rho_{h,S}} + w_{ne,S,k'}^{1-\rho_{h,S}})^{\frac{1}{1-\rho_{h,S}}}$ is known from 38.

and

$$\tilde{x}_{S,l,k'} = \frac{A_{k',ne,S}^{(\frac{\rho_{h,S}}{\rho_{h,S}-1})(\frac{1}{\rho_S}-1)}}{A_{l,S,k'}^{\frac{1}{\rho_S}-1}}$$

From 40, we get:

$$p_{S,k'}^{h} Q_{k',S,h}^{\frac{1}{\rho_S}} = p_{S,k'}^{l} Q_{k',S,l}^{\frac{1}{\rho_S}}$$

Or, substituting prices and quantities in terms of their observable components,

$$\tilde{p}_{S,k'}^{h} (\tilde{Q}_{k',S,h})^{\frac{1}{\rho_S}} A_{k',ne,S}^{(\frac{\rho_{h,S}}{\rho_{h,S}-1})(\frac{1}{\rho_S}-1)} = \frac{W_{l,S,k'}}{A_{l,S,k'}} (A_{l,S,k'} L_{l,S,k'})^{\frac{1}{\rho_S}}$$

Thus,

$$\frac{A_{k',ne,S}^{(\frac{\rho_{h,S}}{\rho_{h,S}-1})(\frac{1}{\rho_S}-1)}}{A_{l,S,k'}^{\frac{1}{\rho_S}-1}} = \frac{W_{l,S,k'} (L_{l,S,k'})^{\frac{1}{\rho_S}}}{\tilde{p}_{S,k'}^{h} (\tilde{Q}_{k',S,h})^{\frac{1}{\rho_S}}}$$

Denote

$$\tilde{x}_{S,l,k'} = \frac{A_{k',ne,S}^{(\frac{\rho_{h,S}}{\rho_{h,S}-1})(\frac{1}{\rho_S}-1)}}{A_{l,S,k'}^{\frac{1}{\rho_S}-1}}$$

C.3 Unknown amenities

Given the distribution of population in each region, estimated migration costs and real wages, unknown region, field of education and sector specific amenities are backed out. The equilibrium population in location d of workers with degree in s working in sector S is given by:

$$L_{sdS} = \sum_j \mu_{jdss}^\theta \left(\frac{W_{sdS} u_{sdS}}{P_d}\right)^{\theta_s} \Phi_{sj}^{-1} L_{sj} \tag{44}$$

And

$$\Phi_{sj} = \sum_{d'S'} \mu_{ojs}^\theta \left(\frac{W_{sd'S'} u_{sdS}}{P_d}\right)^{\theta_s} \tag{45}$$

Equations 44 and 45 have $D * s * S + D * s$ unknowns and $D * s * S + D * s$ equations. We can solve these uniquely for the unknowns Φ_{sj} and $\left(\frac{W_{sdS} u_{sdS}}{P_d}\right)^{\theta_s}$. Using the obtained values, I run the following regression:

$$\left(\frac{W_{sdS} u_{sdS}}{P_d}\right)^{\theta_s} = \theta_s ln\left(\frac{W_{sdS}}{P_d}\right) + \theta_s u_{sdS}$$

Since wages are given from data and the sequence of regional prices P_d have already been estimated, one can run this regression to recover θ_s. Now local amenities are correlated with real wages in general equilibrium. I can again use the same instrument here by using a long-difference equation and using model predicted wages, holding amenities constant at old values, as an instrument.

To estimate the elasticity of movement for education, I use the population of people with degrees in field s in each location:

$$L_{js} = \sum_o l_{ojs} L_o = \sum_o \frac{(a_{js} \mu_{oj} E(V^{ijs}))^\gamma}{\Phi_o} L_0 \tag{46}$$

$$\Phi_o = \sum_{s,j} (a_{js} \mu_{oj} E(V^{ijs}))^\gamma \tag{47}$$

In the same way, I can solve for unknown quantities a_{js} and Φ_o. The utility cost of education has two components:

C.4 Existence of equilibrium proof

To show the existence of equilibrium I use the following theorem, proved in Allen et al (2019).

Theorem 1: Consider any $N \times K$ system of equations $F : R_{++}^{N \times K} R_{++}^{N \times K}$:

$$F(x)_{ik} \equiv \sum_j K_{ij,k} \prod_{l=1}^{K} (x_{j,l})^{\alpha_{k,l}} \prod_{l=1}^{K} (x_{i,l})^{\lambda_{k,l}} \prod_{m=1}^{M} Q_m(x_j)^{\gamma_{k,m}} \prod_{m=1}^{M} Q_m(x_i)^{\kappa_{k,m}}$$

where $Q_m(.)$ are nested CES aggregating functions:

$$Q_m(x_j) \equiv \left(\sum_{l \in S_m} \frac{1}{\mid S_m \mid} \left(\left(\sum_{n \in T_l} \frac{1}{\mid T_n \mid} (x_{j,n})^{\delta_{m,l}} \right)^{\frac{1}{\delta_{m,l}}} \right)^{\beta_m} \right)^{\frac{1}{\beta_m}}$$

where $\delta_{m,l} > 0$ and $\beta_m > 0$ for all m and l, $K_{ijk}, U_l, T_{j,n}$ are all strictly positive parameter values; S_m and $T_{l,m}$ are (weak) subsets of $1, ..., K$; and $\{\alpha_{k,l}, \lambda_{k,l}, \gamma_{k,m}, \kappa_{k,p}\}$ are all real-valued.

If $max_{k \in \{1,...,K\}} \left(\sum_{m=1}^{M} \mid \gamma_{k,m} \mid + \sum_{l=1}^{K} \mid \alpha_{k,l} \mid + \sum_{m=1}^{M} \mid \lambda_{k,m} \mid + \sum_{m=1}^{M} \mid \kappa_{k,m} \mid \right) < 1$, then there exists an unique fixed point $F(x*) = x*$

I can show that the equilibrium system of equations in my model falls into the framework considered by theorem 1.

The equilibrium conditions that govern enrollment are:

$$L_{o_2} = \sum_{o_1, f} l_{o_1 o_2 f} L_{o_1}$$

$$L_{o_2} = \sum_{o_1, f} \frac{\left(\left(\frac{a_{o_2 f} \mu_{o_1 o_2}^1}{P_{o_2}} \right)^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}}{\Phi_{o_1}^{\frac{\gamma}{IU}}} L_{o_1}$$

$$L_{o_2} \left(\left(\frac{a_{o_2 f}}{P_{o_2}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} = \sum_{o_1, f} \frac{\left((\mu_{o_1 o_2}^1)^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}}{\Phi_{o_1}^{\frac{\gamma}{IU}}} L_{o_1}$$

$$\Phi_{o_1}^{\frac{\gamma}{IU}} = \sum_{o_2', f'} \left(\left(\frac{(a_{o_2' f'} \mu_{o_1 o_2'}^1)}{P_{o_2'}} \right)^{IU} \Phi_{o_2' f'}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

Let the following hold for some value of κ

$$L_{o_2} \left(\left(\frac{a_{o_2 f}}{P_{o_2}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} = \kappa_f (\Phi_{o_2})^{\frac{\gamma}{IU}}$$

Thus, we get,

$$\kappa_f (\Phi_{o_2})^{\frac{\gamma}{IU}} = \sum_{o_1,f} \frac{\left((\mu^1_{o_1 o_2})^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}}{\Phi_{o_1}^{\frac{\gamma}{IU}}} L_{o_1}$$

$$= \sum_{o_1,f} \frac{\left((\mu^1_{o_1 o_2})^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}}{\Phi_{o_1}^{\frac{\gamma}{IU}}} L_{o_1}$$

$$= \sum_{o_1,f} \left((\mu^1_{o_1 o_2})^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}} \kappa_f L_{o_1}^{-1} (\frac{a_{o_1,f}}{P_{o_1}}^{\gamma}) L_{o_1}$$

Simplifying,

$$\Phi_{o_2}^{\frac{\gamma}{IU}} = \sum_{o_1,f} \left(\left(\frac{(a_{o_1 f} \mu^1_{o_1 o_2})}{P_{o_1}} \right)^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

The two equations then just boil down to one.

This allows us to consider a single non linear equation:

$$L_{o_2} = \kappa_f \left((\frac{a_{o_2 f}}{P_{o_2}})^{IU} \right)^{\frac{-\gamma}{IU}} \sum_{o_1,f} \left(\left(\frac{(a_{o_1 f} \mu^1_{o_1 o_2})}{P_{o_1}} \right)^{IU} \Phi_{o_2 f}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}} \tag{48}$$

Substitute

$$(\Phi_{o_2 f}) = \sum_{d'' f''} \mu^\theta_{o_2 d''} (\frac{W_{fd''S''} u_{fd''S''}}{P''_d})^\theta$$

$$L_{o_2} = \kappa_f \left((\frac{a_{o_2 f}}{P_{o_2}})^{IU} \right)^{\frac{-\gamma}{IU}} \sum_{o_1,f,d'',f''} \left(\left(\frac{(a_{o_1 f} \mu^1_{o_1 o_2})}{P_{o_1}} \right)^{IU} (\mu^\theta_{o_2 d''} (\frac{W_{fd''S''} u_{fd''S''}}{P''_d})^\theta)^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

The equilibrium condition in the internally traded sector is given by:

$$Y_{dS}^{\sigma_S} Q_{dS}^{1-\sigma_S} = \sum_j \tau_{dj}^{1-\sigma_S} P_j^{\sigma_S - 1} Y_j$$

We can rewrite the internal gravity equation 12 as:

$$Y_{dS} = \sum_j \tau_{dj}^{1-\sigma} p_{dS}^{1-\sigma} P_J^{\sigma-1} (\alpha_j Y_j)$$

Multiplying both sides by $Q_{dS}^{1-\sigma}$, we get,

$$Y_{dS}Q_{dS}^{1-\sigma} = \sum_j \tau_{dj}^{1-\sigma}p_{dS}^{1-\sigma}P_j^{\sigma-1}Q_{dS}^{1-\sigma}(\alpha_S Y_j)$$

$$= \sum_j \tau_{dj}^{1-\sigma}P_j^{\sigma-1}Y_{dS}^{1-\sigma}(\alpha_S Y_j)$$

Simplifying, the above:

$$Y_{dSS}^{\sigma}Q_{dS}^{1-\sigma} = \sum_j \tau_{dj}^{1-\sigma}P_j^{\sigma-1}(\alpha_S Y_j)$$

We can rewrite 13,

$$P_j^{1-\sigma} = \sum_k \tau_{jk}^{1-\sigma}p_{kS}^{1-\sigma}Q_{kS}^{1-\sigma_S}Q_{kS}^{\sigma_S-1}$$

$$= \sum_k \tau_{jk}^{1-\sigma}Y_{kS}^{1-\sigma}Q_{kS}^{\sigma-1}$$

Suppose that the following relationship holds true for some scalar κ

$$Y_{dS}^{\sigma}Q_{dS}^{1-\sigma} = \kappa P_d^{1-\sigma}$$

In that case, as I show below, I can express equations 12 and 13 as a single equation. Equation 12 is given below:

$$Y_{dS}^{\sigma}Q_{dS}^{1-\sigma} = \sum_j \tau_{dj}^{1-\sigma}P_j^{\sigma-1}(\alpha_S y_j)$$

Substituting $Y_{dS}^{\sigma}Q_{dS}^{1-\sigma} = \kappa P_d^{1-\sigma}$ in the above we get back equation 13

$$P_d^{1-\sigma} = \sum_j \tau_{dj}^{1-\sigma}(Y_{jS}^{1-\sigma})Q_{jS}^{\sigma-1}$$

This allows us to consider a single non-linear equation:

$$Y_{dS}^{\sigma}Q_{dS}^{1-\sigma} = \kappa \sum_{j\in N} \tau_{dj}^{1-\sigma}(Y_{jS}^{1-\sigma})Q_{jS}^{\sigma-1}$$

Now substitute the price index

$$P_d = (\kappa)^{\frac{-1}{1-\sigma}}Y_{dS}^{\frac{\sigma}{1-\sigma}}Q_{dS}$$

in 48

$$L_{o_2} = \kappa_f \left(\left(\frac{a_{o_2 f}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_2 S}^{\frac{\sigma}{1-\sigma}} Q_{o_2 S}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} \sum_{o_1, f} \left(\left(\frac{(a_{o_1 f} \mu_{o_1 o_2}^1)}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_1 S}^{\frac{\sigma}{1-\sigma}} Q_{o_1 S}} \right)^{IU} \right.$$
$$\left(\sum_{d'' f''} \mu_{o_2 d''}^{\theta} \left(\frac{W_{fd'' S''} u_{fd'' S''}}{P_d''} \right)^{\theta} \right)^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

Simplifying,

$$L_{o_2} = \kappa_f \left(\left(\frac{a_{o_2 f}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_2 S}^{\frac{\sigma}{1-\sigma}} Q_{o_2 S}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} \sum_{o_1, f} \left(\left(\frac{(a_{o_1 f} \mu_{o_1 o_2}^1)}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_1 S}^{\frac{\sigma}{1-\sigma}} Q_{o_1 S}} \right)^{IU} \right.$$
$$\left(\sum_{d'' f''} \mu_{o_2 d''}^{\theta} \left(\frac{W_{fd'' S''} u_{fd'' S''}}{P_d''} \right)^{\theta} \right)^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$
$$= \sum_{o_1, f, d'', f''} \kappa_f \left(\left(\frac{a_{o_2 f}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_2 S}^{\frac{\sigma}{1-\sigma}} Q_{o_2 S}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} \left(\left(\frac{(a_{o_1 f} \mu_{o_1 o_2}^1)}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_1 S}^{\frac{\sigma}{1-\sigma}} Q_{o_1 S}} \right)^{IU} \right.$$
$$\left(\mu_{o_2 d''}^{\theta} \left(\frac{W_{fd'' S''} u_{fd'' S''}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{d'' S}^{\frac{\sigma}{1-\sigma}} Q_{d'' S}} \right)^{\theta} \right)^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

Finally, substitute the expression for wages,

$$W_{fd'' S''} = Y_{d'' S''} Q_{d'' S''}^{\frac{1-\rho_f}{\rho_f}} Q_{hd'' S''}^{\frac{1}{\rho_{h, S''}} - \frac{1}{\rho_S''}} A_{fd'' S''} (\tilde{L}_{fd'' S''})^{\frac{-1}{\rho_h}}$$

$$L_{o_2} = \sum_{o_1, f, d'', f''} \kappa_f \left(\left(\frac{a_{o_2 f}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_2 S}^{\frac{\sigma}{1-\sigma}} Q_{o_2 S}} \right)^{IU} \right)^{\frac{-\gamma}{IU}} \left(\left(\frac{(a_{o_1 f} \mu_{o_1 o_2}^1)}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{o_1 S}^{\frac{\sigma}{1-\sigma}} Q_{o_1 S}} \right)^{IU} \right.$$
$$\left(\mu_{o_2 d''}^{\theta} \left(\frac{Y_{d'' S''} Q_{d'' S''}^{\frac{1-\rho_f}{\rho_f}} Q_{hd'' S''}^{\frac{1}{\rho_{h, S''}} - \frac{1}{\rho_S''}} A_{fd'' S''} (\tilde{L}_{fd'' S''})^{\frac{-1}{\rho_h}} u_{fd'' S''}}{(\kappa)^{\frac{-1}{1-\sigma}} Y_{d'' S}^{\frac{\sigma}{1-\sigma}} Q_{d'' S}} \right)^{\theta} \right)^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$$

We are thus able to express the equilibrium conditions in the form required for theorem 1. An equilibrium thus exists by the contraction mapping theorem.

C.5   Proofs of Frechet propositions

Proposition 1: If $\eta \sim Frechet(\theta)$, then $\eta^{\alpha} \sim Frechet(\frac{\theta}{\alpha})$

Proof:

Since $\eta \sim Frechet(\theta)$, thus

$$F_\eta(x) = P(\eta \le x) = (exp(-x^{-\theta}))$$

Let $z = \eta^\alpha$

$$\begin{aligned} F_z(x) = P(z \le x) &= P(\eta^\alpha \le x) \\ &= P(\eta \le (x)^{\frac{1}{\alpha}}) \\ &= exp(-(x^{\frac{1}{\alpha}})^{-\theta}) \\ &= exp(-(x^{\frac{-\theta}{\alpha}})) \end{aligned}$$

Thus z follows Frechet with dispersion parameter $\frac{-\theta}{\alpha}$

Proposition 2: If $\eta_i \sim Frechet(\theta)$, then $E(max_i(a_i \times \eta_i)) = (\sum_i a_i^\theta)^{\frac{1}{\theta}} \Gamma(1 - \frac{1}{\theta})$

Proof:

Let $z_i = max_i(a_i\eta_i)$

$$\begin{aligned} F_{Z_i}(z) &= Pr(Z_i \le z) \\ &= Pr(a_i\eta_i \le z \forall i) \\ &= Pr(\eta_i \le \frac{z}{a_i}, \forall i) \\ &= \Pi_i F(\frac{z}{a_i}) \\ &= \Pi_i exp(-\frac{z}{a_i})^{-\theta} \\ &= exp(-z^{-\theta}(\sum_i (\frac{1}{a_i})^{-\theta})) \\ &= exp(-(\sum_i a_i^\theta))z^{-\theta} \end{aligned}$$

$z = max_i(a_i\eta_i)$ thus follows a Frechet distribution with dispersion parameter θ and position parameter $(\sum_i a_i^\theta)$

According to the properties of the Frechet distribution, the mean of z will thus be

$E(z) = E(max_i(a_i\eta_i)) = (\sum_i a_i^\theta)^{\frac{1}{\theta}} \Gamma(1 - \frac{1}{\theta})$

To understand how the propositions apply to the maximization problem at hand, consider $(1 - IU) = \alpha$, $a_i = (w_{f,dS} \cdot P_d \cdot u_{f,dS} \cdot \mu_{o_2d}^2)^{1-IU}$

D   Identification and estimation

D.1   Migration cost estimation

In column 1, I report the results from the PPML estimation, the same one reported in the main text of the paper. In column 2, I report the results from regressing the log flows of people migrating for education on the relevant distance measures, where zeros are replaced by the minimum across all migration flows. In column 3, I repeat the same estimation as in column 2, but with zeros excluded. In column 4, I follow a more traditional estimation where the combinations of same state and neighbor dummies have been replaced with just a same state dummy. Across all specifications, the effect of state borders is huge: the effect is 269.5%, 481%, 293.5% and 1039% in specifications 1,2,3 and 4 respectively. Specification 1 is the preferred specification. In specification 2 pairs of districts that do not record any migration flows receive a small value in order to avoid getting thRoWn out of the sample, which induces downward bias in the estimated cost. In specification 3, the pairs of districts with zero values have been ignored which introduces the selection problem. In specification 4, the estimated state border effects are huge because districts in different states share a border with less frequency than districts in the same state, which was previously accounted for by the neighborhood dummy and now gets loaded on to the state dummy.

Table XVIII: **Gravity estimation on district to district migration when reason for migration is education**

	(1)	(2)	(3)	(4)
	migration flows	log migration flows	log migrationflows	migration flows
log distance centroid	-0.603***	-0.616***	-0.620***	-0.771***
	(-91.50)	(-136.98)	(-87.57)	(-159.26)
log common	0.307***	0.044***	0.106***	0.323***
	(25.39)	(22.86)	(24.60)	(25.04)
Same state, neighbors	3.647***	4.898***	2.686***	
	(75.31)	(192.11)	(95.61)	
Same state, not neighbors	2.380***	2.014***	1.057***	
	(59.10)	(179.62)	(67.88)	
Not same state, neighbors	2.340***	3.138***	1.316***	
	(25.27)	(71.90)	(30.78)	
Same state				2.433***
				(56.23)
N	280900	280900	45362	280900

In table 7 above, the same specifications are repeated when people migrate for work.

Table XVII: Gravity estimation on district to district migration when reason for migration is work

	(1) migration flows	(2) log migration flows	(3) log migration flows	(4) migration flows
log distance centroid	-0.548***	-1.294***	-1.006***	-0.722***
	(-75.44)	(-217.22)	(-171.13)	(-135.72)
log common	0.332***	0.164***	0.194***	0.342***
	(39.70)	(63.88)	(65.24)	(36.47)
Same state, neighbors	3.025***	2.663***	2.320***	
	(64.04)	(78.82)	(82.73)	
Same state,not neighbors	1.773***	2.112***	1.342***	
	(50.17)	(142.09)	(104.13)	
Not same state, neighbors	2.629***	2.702***	2.059***	
	(42.00)	(46.72)	(44.81)	
Same state				1.751***
				(44.21)
Clustering	State-year State	District	State	State-year
N	280900	280900	127709	280900

Table XIX: **Gravity estimation on district to district migration when reason for migration is education**

	(1)	(2)	(3)	(4)
	migrationflows2	logmigrationflowsNoZero	logmigrationflowsWithZero	migrationflows2
logdistance_centroid	-0.609***	-0.614***	-0.617***	-0.777***
	(-92.37)	(-137.31)	(-87.89)	(-162.01)
log common	0.303***	0.044***	0.106***	0.317***
	(25.05)	(23.03)	(24.64)	(24.94)
Same state, neighbors	3.611***	4.893***	2.685***	
	(74.11)	(191.91)	(95.67)	
Same state, not neighbors	2.364***	2.010***	1.051***	
	(58.61)	(179.22)	(67.49)	
Not same state, neighbors	2.315***	3.155***	1.329***	
	(24.76)	(72.34)	(31.13)	
Same state				2.408***
				(55.75)
N	280900	280900	45362	280900

Table XX: **Gravity estimation on district to district migration when reason for migration is work**

	(1)	(2)	(3)	(4)
	migrationflows2	logmigrationflowsNoZero	logmigrationflowsWithZero	migrationflows2
logdistance econcenter	-0.553***	-1.284***	-0.995***	-0.727***
	(-76.85)	(-216.57)	(-170.23)	(-138.81)
log common	0.331***	0.166***	0.195***	0.340***
	(39.11)	(64.70)	(65.87)	(35.75)
Same state, neighbors	2.992***	2.669***	2.331***	
	(63.82)	(78.96)	(83.06)	
Same state, not neighbors	1.753***	2.107***	1.339***	
	(49.89)	(141.68)	(103.77)	
Not same state, neighbors	2.611***	2.750***	2.099***	
	(39.89)	(47.55)	(45.67)	
Same state				1.723***
				(44.13)
N	280900	280900	127709	280900

Table XXI: **Gravity estimation on district to district migration when reason for migration is education, clustering at origin level**

	(1) migrationflows2 b/t	(2) logmigrationflowsNoZero b/t	(3) logmigrationflowsWithZero b/t	(4) migrationflows2 b/t
logdistance_centroid	-0.603***	-0.616***	-0.620***	-0.771***
	(-68.52)	(-41.63)	(-46.61)	(-67.15)
logcommon	0.307***	0.044***	0.106***	0.323***
	(15.40)	(9.52)	(19.19)	(15.53)
Same state, neighbors	3.647***	4.898***	2.686***	
	(53.89)	(72.24)	(83.06)	
Same state, not neighbors	2.380***	2.014***	1.057***	
	(34.12)	(28.98)	(41.89)	
Not same state, neighbors	2.340***	3.138***	1.316***	
	(19.69)	(34.73)	(21.09)	
Same state				2.433***
				(35.48)
N	280900.000	280900.000	45362.000	280900.000

Table XXII: **Gravity estimation on district to district migration when reason for migration is work, clustering at origin level**

	(1) migration flows b/t	(2) log migration flows no zero b/t	(3) log migrationflows With Zero b/t	(4) migrationflows2 b/t
logdistance_centroid	-0.548***	-1.294***	-1.006***	-0.722***
	(-52.50)	(-56.99)	(-54.57)	(-65.96)
logcommon	0.332***	0.164***	0.194***	0.342***
	(26.63)	(16.79)	(25.63)	(22.91)
(mean) one_one	3.025***	2.663***	2.320***	
	(43.02)	(30.25)	(43.16)	
(mean) one_zero	1.773***	2.112***	1.342***	
	(36.95)	(45.09)	(42.27)	
(mean) zero_one	2.629***	2.702***	2.059***	
	(22.46)	(32.15)	(27.26)	
(mean) same_state				1.751***
				(31.88)
N	280900.000	280900.000	127709.000	280900.000

Table XXIII: **Gravity estimation on district to district migration when reason for migration is education, clustering at destination level**

	(1) migrationflows b/t	(2) logmigrationflowsNoZero b/t	(3) logmigrationflowsWithZero b/t	(4) migrationflows b/t
logdistance_centroid	-0.603***	-0.616***	-0.620***	-0.771***
	(-62.04)	(-25.21)	(-36.35)	(-81.47)
logcommon	0.307***	0.044***	0.106***	0.323***
	(17.77)	(5.81)	(8.61)	(18.82)
Same state, neighbors	3.647***	4.898***	2.686***	
	(45.48)	(66.78)	(61.44)	
Same state, not neighbors	2.380***	2.014***	1.057***	
	(27.49)	(24.07)	(21.56)	
Not same state, neighbors	2.340***	3.138***	1.316***	
	(17.16)	(34.00)	(24.02)	
same_state				2.433***
				(27.58)
N	280900.000	280900.000	45362.000	280900.000

Table XXIV: **Gravity estimation on district to district migration when reason for migration is work, clustering at destination level**

	(1) migration _ lows	(2) log_ migration_flows_no_zero	(3) log_migration_flows_with_zero	(4) migration_flows
logdistance_centroid	-0.548***	-1.294***	-1.006***	-0.722***
	(-46.73)	(-73.33)	(-64.70)	(-70.47)
logcommon	0.332***	0.164***	0.194***	0.342***
	(23.27)	(20.08)	(24.11)	(23.00)
Same state, neighbors	3.025***	2.663***	2.320***	
	(37.83)	(31.91)	(41.97)	
Same state, not neighbors	1.773***	2.112***	1.342***	
	(23.00)	(47.22)	(40.93)	
Not Same state, neighbors	2.629***	2.702***	2.059***	
	(28.49)	(33.42)	(30.45)	
Same_state				1.751***
				(22.66)
N	280900.000	280900.000	127709.000	280900.000

D.2 Trade cost estimation

<div align="center">

Table XXV: **First Stage Regression**

</div>

	(1) Observable price
Hisorical Software exports	-6.139***
	(-2.83)
Remoteness	-0.0000135
	(-1.36)
log English to hindi speakers	0.213***
	(3.31)
linguistic distance	-390.901**
	(-2.39)
Constant	-10.088*
	(-1.83)
N	482.000

Robust Standard errors are used. t statistics reported in parenthesis

E Theoretical extensions

1. Quality differences in education: Let the idiosyncratic preference shock be drawn from a Frechet distribution with mean T_{o_2}, where T_{o_2} depends on regional average quality of education.

$$G(\zeta_{io_2f}) = exp(-T_{o_2}\zeta_{io_2f}^{-\gamma})$$

Given this, the proportion of people migrating for education is given by:

$$l_{o_1o_2f} = \frac{T_{o_2}\left(\left(\frac{a_{o_2f}\mu_{o_1o_2}^1}{P_{o_2}}\right)^\beta \Phi_{o_2f}^{\frac{(1-\beta)}{\theta}}\right)^{\frac{\gamma}{\beta}}}{\Phi_{o_1}} \tag{49}$$

The new Φ_{o_1} is scaled by T_{o_2}. The destination fixed effects capture the average quality of education in that region. It therefore behaves in the exact same manner as amenities for education. Note that, this does not change the migration equation for work.

2. International migration: International migration is introduced in the model by adding one region where people can migrate to but from where people cannot migrate out. This region is closer to some regions of India and further from others. Education facilities and job opportunities are both better in this region than in any

region of India. The introduction of this region increases both the aggregate welfare as well as the regional inequality.

3. Endogenous agglomeration and congestion:

$$A = \overline{A}L^{\alpha}$$

$$U = \overline{U}L^{-\beta}$$

I take $\alpha = .3$ and $\beta = -.2$. For this parametric configuration, overall inequality as a result of the IT boom increases.

4. Differential mobility costs for skilled and unskilled workers: In this extension, the migration costs for unskilled workers are taken to be double that of skilled workers, with the average migration costs being the same as the original estimated migration costs. In this extension, unskilled workers lose in about a third of Indian districts. This is because skilled and unskilled workers are complements in the production function. As skilled workers start migrating out of certain districts that did not see much of the IT boom, this brings down the marginal productivity of unskilled workers as skilled and unskilled workers are complements in the production function.