The IT Boom and Other Unintended Consequences of Chasing the American Dream*

Gaurav Khanna†
Nicolas Morales‡

November 2016

Abstract

We use a general equilibrium model to study how US immigration policy coupled with the internet boom led to a tech boom in India. Specifically, we test the hypothesis that Indian students enrolled in engineering schools to gain employment in the rapidly growing US IT industry via the H-1B visa program. Those who could not join the US workforce, due to the H-1B cap, remained in India, enabling the growth of an Indian IT sector. Those who returned with acquired human capital and technology after the expiration of their H-1Bs also contributed to the growing tech-workforce in India. The increase in IT sector productivity allowed India to eventually surpass the US in IT exports. Our general equilibrium model captures firm-hiring across various occupations, innovation and technology diffusion, and dynamic worker decisions to choose occupations and fields of major in both the US and India. Supported by a rich descriptive analysis of the changes in the 1990s and early 2000s, we calibrate this model and perform counterfactual exercises. We find that the H-1B program induced Indians to switch to computer science (CS) occupations, increasing the CS workforce in India by 4.9% in 2010. It also induced US workers to switch to non-CS occupations, reducing the US native CS workforce by 7%. Indian IT output is higher by 1.8%, India's share in world IT production greater by 8.9%, and the combined income of both countries is higher by 1.5% because of this high-skilled migration.

JEL: I25, J30, J61

Keywords: High-skill immigration, H-1B visas, India, computer scientists, IT sector

*We thank John Bound, Andrei Levchenko, Sebastian Sotelo and seminar participants in Michigan, Purdue, Montevideo and New Delhi for insightful comments. Khanna would like to acknowledge the Alfred P. Sloan Foundation and the NBER Fellowship on High-Skill Immigration for generous research support.

†Center for Global Development and University of California – San Diego, gauravkhanna.info

‡University of Michigan, moralesn@umich.edu

Innovations in information technology allowed the US IT sector to expand rapidly in the early 1990s (Bound et al., 2015; Kerr, 2013a). A few years later, India experienced this growth: the IT-BPO (Information Technology and Business Process Outsourcing) sector increased its contribution to the Indian GDP from 1.2% in 1998 to 7.5% in 2012 (NASSCOM, 2012).[1] In the two decades spanning the 1990s and 2000s, India produced a growing supply of skilled IT workers, a large fraction of which were fed by a burgeoning student enrollment in both existing and newly built engineering colleges. While a fraction of these workers entered the US labor market via H-1B visas (and some subsequently became permanent members of the US workforce), many joined firms in India, supplying work to a rapidly growing IT sector. Given that by 2014, 70% of H-1B visas went to Indian workers, the H-1B program and the tech boom had a severe impact on the Indian IT sector. By the mid-2000s, the US-led boom spread to India, and India took over from the US as the major exporter of software. The human capital investments made by Indian students and workers in response to the boom, and the technology diffusion propagated by those who were forced to return after their visas expired, played a substantial role in the spread of this boom.

We develop a model of firm and worker decisions in both countries to test to what degree the following hypothesis can help explain the simultaneous growth in the US and Indian tech-industry. Starting in the early 1990s, innovation in the US IT sector led to a growth in computer science (CS) employment and wages, and enrollment in CS degrees (Figures 1a to 1c). An immigration policy that relatively favored high-skill immigrants led to an increasing fraction of foreigners in the US computer-science workforce (Figure 1d and 1e). By the mid-2000s more than half of all H-1B visas went to Indians, and by 2014 this number was as high as 70% (USCIS, 2014).[2] While all the top firms that hired H-1Bs are IT firms, the top 9 had India as their primary employment base (Table 1).

CS wages in the US are many times higher than in India (Figure 1f and Clemens (2013)). In search of these higher US wages, more and more Indian students started enrolling in engineering schools (Figure 2a). However, the number of available H-1B visas was capped, so a large number of potential foreign-workers then had to seek employment in India. Furthermore, since H-1Bs expired after 3 to 6 years, many of these workers returned to India, bringing with them their accumulated human capital and technological knowhow. This educated workforce in India enabled the Indian IT sector to grow rapidly, with new firms joining the race and older firms expanding, and over time, India became a major producer of software and eroded the US dominance on exports of IT products (Figure 2b and 2c).

Under our hypothesis, a lack of a tech-boom in the US, or more restrictive immigration policies for high-skilled workers may have led to fewer Indian engineering graduates and IT firms in both

[1] NASSCOM or the National Association of Software and Services Companies is a trade association of Indian Information Technology and Business Process Outsourcing industry established in 1988.

[2] The next highest category, China, received only 8% of H-1Bs in 2012.

countries. In so far as a newly educated Indian workforce played a major role in worldwide IT growth, our results are relevant in considering future immigration policy for high-skill labor into the US and other parts of the world, and for understanding structural transitions to high-skill workforces in emerging economies and the developing world.

In order to explore this hypothesis, we model firm-hiring and the forward-looking decisions of students and workers who dynamically choose majors and occupations. This model is grounded in and consistent with empirical trends, and we first provide a descriptive analysis of the IT sector growth in the two decades starting in the early 1990s. We then calibrate this model and perform counterfactual exercises to study the role played by the H-1B program.

In Section 1 we briefly describe the evolution of the IT sector in the US and India, and the changes in the labor market and education sector over this period. With time, an increasing proportion of the computer scientists (CS) and engineers in the US came from abroad. In the second half of the 1990s, the foreign fraction of CS workers increased considerably compared to both the foreign fraction of all workers with a bachelor degree and the foreign fraction of all workers in STEM occupations (Figure 1d). A substantial portion of this increase was driven by workers from India (Figure 2d). In general, the growth in the Indian IT sector was inextricably linked to the boom in the US. The bulk of the Indian growth in this sector was driven by exports to the US and the rest of the world. However, only a handful of other countries, like Israel and Ireland experienced booms of such magnitude. In this section we also discuss why the boom missed many other countries but settled on India. India has not only historically had high quality engineering schools that train potentially lower-wage (Arora et al., 2001), English-speaking workers but had also developed strong networks with the US sector during the earlier hardware boom (Bhatnagar, 2005).

In Section 2 we describe the model, which captures certain crucial features. First, we model how US firms hire both US and foreign workers, and Indian firms hire workers from India. Importantly, firms hire three different types of workers – computer scientists, non-CS college graduates and non college graduates. Second, the IT sector is monopolistically competitive with different varieties of products produced in each country. Third, computer scientists are also innovators and increase the overall productivity of firms in the IT sector via the generation of non-excludable ideas. Under this directed technological change, an increase in the relative size of the computer science workforce makes India relatively more productive over time. Fourth, to capture the trade patterns, we encapsulate the canonical Eaton and Kortum (2002) framework into our model of the labor market. Importantly, this allows us to capture the empirical patterns that show that even though the US had been the major exporter of IT, India takes over as the leader in the mid-2000s.

While the above four features capture the product and the labor demand aspects of the economy, the next two capture important labor supply decisions. The fifth feature is that students in both

countries are heterogeneous in preferences, and make dynamic decisions on choosing their college major given their expected future earnings in different occupations. Sixth, after graduation, workers (also with heterogeneous preferences), choose every year to either continue working in their current occupation or switch occupations given the labor demand shocks and their expected future benefits in each occupation. Student and worker dynamic decisions, therefore, depend on the innovation shocks and their preferences.

After calibrating the model in Section 3, we show that our model does a good job of matching both levels and trends in wages, employment and IT sector output in out of sample tests for both countries (Section 4). We then conduct counter-factual exercises to study the impact of more restrictive immigration policy on both the US and Indian IT sectors (Section 5). Our results indicate that US immigration policy did play a significant role in the spread of the IT boom from the US to India. The possibility of migrating to the US under the H-1B program incentivized students and workers in India to choose CS degrees and occupations. Those that returned after the expiration of their visas contributed to this growing CS workforce and enabled the increases in technological productivity in India. We show that the H-1B program was associated with an increase of 4.9% in the size of the non-migrant Indian CS workforce in 2010. However, the migration led US native CS workers to switch to non-CS occupations and is therefore associated with a fall in the US native CS workforce by as much as 7% in 2010.

An increase in the size of the Indian CS workforce also led to an increase in productivity in the Indian IT sector. Under the H-1B program, production shifts to India – the share of world IT output that comes from India is 8.9% higher, and Indian IT output increases by 1.8% in 2010. World IT output increases by 1.6% and the US-India combined incomes are higher by 1.5% under the H-1B regime. The shift in production to India, however, hurts some US workers.

In Section 6 we discuss this result in relation to the larger literature on trade, technological diffusion and migration. There is extensive work in trade that focuses on catch-up from the global South, and the role that technological diffusion plays in this catch-up (Krugman, 1979). Since migrants accumulate human capital and technical knowhow in the US and return with this knowledge to India, this speeds up technological diffusion and catch-up. As Davis and Weinstein (2002) highlight, immigration may also deteriorate the terms of trade for the country with superior technology – in this case, the US. On the other hand, Freeman (2006b) argues that immigration can actually help the US maintain its advantage by attracting global talent. This analysis, however, misses the incentives to invest in human capital, and the corresponding growth in production for sending countries like India – features that play an important role in our analysis.

Our paper is innovative in many ways. First, it addresses some crucial issues raised by the literature on the impacts of high-skill immigrants on the US economy. High-skill immigrants could

impart benefits to employers, complementary inputs used in production, and to consumers, and in general may be valuable innovators in technology (Foley and Kerr, 2013). However, they potentially impose some costs on domestic workers who are close substitutes (Borjas, 1999). On the other hand, the magnitude of these costs may be substantially mitigated if US high skill workers have good alternatives to working in sectors most impacted by immigrants (Peri and Sparber, 2011). If high skill immigrants contribute to the generation of knowledge and productivity through patenting and innovation, then this serves to shift out the production possibility frontier in the US, and may also slow the erosion of the US comparative advantage in high tech (Freeman, 2006b). We address these issues in a general equilibrium framework, similar to Bound, Khanna, and Morales (2016), to measure the entire distribution of welfare benefits for all the different types of workers.

Second, this is one of the first papers to look at the impact on countries that send immigrants rather than receive them. The Indian case is interesting in particular because not only does it contribute to a major portion of the H-1B visas, but the country experienced rapid growth in the two decades starting in 1991, a large proportion of which was led by the IT sector. The IT sector boom and immigration policy in the US, and the Indian growth-story are therefore closely linked, and studying this boom can help us understand how workforce skill transitions may come about in developing countries. Importantly, we show that US immigration policy can affect such a structural transition half-way across the world, in India.

Last, the paper addresses the debate regarding brain-drain and brain-gain. While many commentators worry about the fact that a large number of well-educated Indians leave the country for work in the US, this paper will show how better paid jobs may also incentivize students to choose certain majors and supply a highly-educated workforce to Indian firms as well.

A general equilibrium model of both the home and destination country is therefore essential to our analysis for pinning down the entire gamut of welfare effects for all the different types of workers. Changes in immigration policy, like lowering the H-1B cap or extending the work-period, does have effects in both countries. As we show in this paper, US immigration policy led to a boom half-way across the world in India, and a shift in the production of IT from the US to India.

1 The Tech Boom

1.1 A Brief History of Indian IT

The Indian software industry was born in the 1970s when the Tata Consultancy Services (TCS) opened shop and started sending Indian engineers abroad (mostly to the US) to do software programming, referred to as 'bodyshopping.' Some companies were helped at the start by Indian

Government policies that permitted duty-free imports of computer systems if importers would promise to export software and services worth twice the value of imports within a specified time period. In 1978, however, IBM had to exit the market in violation of Foreign Exchange regulations, serving a blow to firms that required their hardware.[3] There was little seen of the US hardware boom in India as the industry was not mature and could not acquire the technical knowhow held on to by firms like IBM that had virtually a monopoly in mainframe computers (Bhatnagar, 2005).[4]

In the 1980s, the government tried to open up the sector by deregulating the import licensing policy. Software policy, in the late 1980s, was made separate from policies governing the hardware sector, and the government tried to procure domestic software and imposed duties on software imports to try and promote the domestic industry. At around the same time, the crash in world-wide hardware prices lowered the costs of setting up the software sector. As the personal computer became more popular in the US and Europe, the demand for software programming services grew rapidly, especially for low-cost workers from India.[5] However, the lack of a domestic market and of the advanced technological capability required to write software meant that 90% of software revenues came from on-site work (NASSCOM, 2003). These counted towards exports, and exports grew from about $50 million in the late 1980s to about $200 million by 1993, at a rate of about 30% a year.

With large-scale economic reforms in the early 1990s, the industry was opened up even further, and there was a spurt in the entry of multinational firms and demand for software services. On-site work dominated because otherwise software had to be transported on tapes which 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 off-shoring of work to India. In 1993, the shift from B-1 to H-1 visas lowered the incentives to hire Indian engineers for on-site work, as they were to be paid the prevailing market wage. The fraction of on-site work went from 90% in 1988 to 66% in 1995 and 56% in 2000.[6] At the same time, import duties on computers were lowered and they were made free from import licensing. One estimate suggests that by 1996, India had 16% of the globalized market in customized software, and more than 100 out of the Fortune 500s had outsourced to them (Dataquest, 1996).

The boom in the US also affected the education sector in India. Bhatnagar (2005) notes that "*growth (in training and degrees) was also driven by larger salaries in the IT industry abroad.*" To meet the rising demand for workers, engineering schools introduced more computer science oriented degrees, and companies started their own training divisions in the 1980s, building

[3] They were back as a joint venture with Tata Information Systems only as late as 1992

[4] At this time personal computers were absent from the Indian market altogether

[5] In 1989, an Indian computer professional earned about $5486, whereas Microsoft was offering $40,000 plus relocation benefits and a green card for software engineers (Subramanian, 1992).

[6] Kumar (2001) notes another significant advantage for the Indian industry - the 12-hour time lag between India and the US virtually doubled the working time per day and cut the development life-cycle by half.

technical skills for the industry (Figure 2a).

In August 1995, the internet was introduced to households in the Indian metros, and the first mobile phone call was made. By 1998, when the government deregulated the internet-suppliers monopoly, there were already more than 1 million internet users in India. The Net allowed many more firms access to the markets abroad since it was cheaper to obtain phone lines than satellite links (Desai, 2003). The Y2K threat was not to the detriment of the Indian industry. In fact, "*Y2K projects were an important source of revenue for Indian firms*" (Arora et al., 2001), and this helped build reputation with their US counterparts. One commentator notes that the industry "*grew on the strength of Y2K and never looked back*" (Dataquest, 2003).

By 2001 exports had reached about $6 billion, growing at about 50-60% annually from the mid-1990s. By this time, only five of the top twenty exporters were subsidiaries of foreign firms, indicating that software exports were largely products of Indian firms.[7] NASSCOM estimates that while about 0.16 million software professionals worked in India in 1996, this number more than doubled to about 0.34 million by 2000, showing that the industry generated about 60,000 jobs a year around this period. In the three years around the dot-com crash, the compound annual growth of employment was about 28.5% (Kumar, 2001). By 2009, when 3G data services on mobile phones were introduced, there were already more than 60 million internet users. This number increased to about 122 million people online by 2011, making it the third largest base in the world, even though 90% of the population was still not connected.

1.2 The Internet Boom in The US

Starting in the mid 1990s, the usage of the Internet for commercial purposes grew rapidly in the US (Leiner et al., 1997).[8] This led to an increase in demand for computer scientists, and a rise in R&D expenditures for the firms. The share of total private R&D of the firms in the computer programming services and the computer related equipment sector increased from 19.5% to 22.1% between 1991 and 1998 (computations using Compustat data). The entry and growth of tech firms like Yahoo, Amazon and eBay helped sustain the "boom" in the IT sector till the end of the century.

These changes had a significant impact on the market for IT workers. The number of computer scientists or computer software developers (CS) increased by 161% between the years 1990 and 2000 (US Census), whereas during the same period, the total number of workers with at least a bachelor degree increased by 27%, while the number of workers in other STEM occupations increased by 14%. Table 2 shows computer scientists as a share of the college

[7] This is in stark contrast to the Irish software industry, where the bulk of the exports were US firms based in Ireland (Athreye, 2005).

[8] The decommissioning of the National Science Foundation Network in April of 1995 is considered crucial for introducing nationwide commercial traffic on the Internet.

educated workforce and the college educated STEM workforce. These shares were rising before 1990, and rises dramatically during the 1990s. By the turn of the century more than half of all STEM workers are computer scientists. In Figure 1a, we can see a similar pattern, additionally showing that the growth of CS employment started in the second half of the 1990s – the same period as the dissemination of the Internet.

A substantial fraction of this immigrant IT workforce was educated abroad. Table 3 shows the fraction of workers, and specifically IT workers, by the location of their highest degrees. Given that such a large proportion obtain their bachelor's degrees in other countries, we would expect the education sector abroad to play a major role in the US tech boom. There is also a growing number of students, especially from India, who do their bachelor's abroad, but come to the US to do a master's degree and enter the US labor force (Bound et al., 2014).

This growth in the demand for computer scientists increased their wages. In the Census, we see an 18% increase in the median real weekly wages of CS workers between 1990 and 2000. The Current Population Survey (CPS) presents similar patterns – in Figure 1c we see that wages of computer scientists increased considerably when compared to both workers with other STEM occupations and all workers with a bachelor degree. In fact, while during the beginning of the 1990s, the earnings of CS workers were systematically lower than other STEM occupations, the wage differential tends to disappear after 1998.

This rising demand for computer scientists also affected educational choices of US students. In Figure 1b, it is clear that the number of bachelor degrees awarded in CS relative to the total number of bachelor degrees, or the number of STEM major degrees increased dramatically. The CS share of total bachelors increased from about 2% in 1995 to more than 4% in 2002, showing that the decision to study computer science also responded to the Internet boom.

1.3 US workforce and the H-1B Visa

Employment adjustments for computer scientists disproportionately favored foreigners (Table 2 and Figure 1d). In the second half of the 1990s, the foreign fraction of CS workers increased considerably more than both the foreign fraction of all workers with a bachelors degree and the foreign fraction of all workers in a STEM occupation. In 1994, foreigners were less represented among individuals working as computer scientists than in other STEM occupations, but by the end of the decade, foreigners comprised 29.6% of the increase in CS workers.

This trend in foreign representation in the workforce was sustained by a few developments. Freeman (2009), and Bound et al. (2014), attribute some of these changes to the dramatic increase in college educated (science and engineering) workers in India and China. In India, the number of first degrees conferred in science and engineering rose from about 176 thousand in 1990 to 455 thousand in 2000. The second development was The Immigration Act of 1990 which

established the H-1B visa program for temporary workers in "specialty occupations," defined as requiring *theoretical and practical application of a body of highly specialized knowledge in a field of human endeavor including, but not limited to, architecture, engineering, mathematics, physical sciences, social sciences, medicine and health, education, law, accounting, business specialties, theology, and the arts.*

In order to hire a foreigner on an H-1B visa the firm must first file a Labor Condition Application (LCA), and pay them the greater of the actual compensation paid to other employees in the same job or the prevailing compensation for that occupation. After which, the H-1B prospective must demonstrate to the US Citizenship and Immigration Services Bureau (USCIS) in the Department of Homeland Security (DHS) that they have the requisite amount of education and work experience for the posted positions. USCIS then may approve the petition for the H-1B non-immigrant for a period up to three years, which can be extended up to six years. Since US employers face both pecuniary and non-pecuniary costs associated with hiring H-1Bs, there must exist some productivity premium associated with them. The U.S. General Accounting Office 2011 survey estimates the legal and administrative costs associated with each H-1B hire to range from 2.3 to 7.5 thousand dollars. It therefore seems reasonable to assume that employers must expect some cost or productivity advantage when hiring foreigners.

Since 1990 the cap on H-1Bs was set at 65,000 visas per year. In the early years, the cap was never reached, but by the time the IT boom was starting in the mid-1990s, the cap started binding and the allocation was filled on a first come, first served basis. The cap was raised to 115,000 in 1999 and to 195,000 for 2000-2003, and then reverted back to 65,000 thereafter.[9] Figure 1e shows the growth in the number of H-1 visas issued over the last three decades, the stock of H-1 visas in the economy each year, and changes in the H-1B visa cap.

H-1B visas, therefore, became an important source of labor for the technology sector. The National Survey of College Graduates (NSCG) shows that 55% of foreigners working in CS fields in 2003 arrived in the US on a temporary working (H-1B) or a student type visa (F-1, J-1). The GAO (1992) claims that "computers, programming, and related occupations" corresponded only to 11% of the total number of H-1 visas in 1989, but according to the USINS (2000), the number of H-1B visas awarded to computer-related occupation in 1999 jumped to close to two-thirds of the visas, and the DOC (2000) estimated that during the late 1990s, 28% of programmer jobs went to H-1B visa holders.

1.4 The Impact of High Skill Immigrants on the US Workforce

Some commentators of the H-1B program argue that given the excess supply of highly qualified foreigners willing to take the jobs, and given the lack of portability of the H-1B visa, workers

[9] The 2000 legislation that raised the cap also excluded universities and non-profit research facilities from it, and a 2004 change added an extra 20,000 visas for foreigners who received a masters degree in the US

on an H-1B visa are not in a position to search for higher wages, allowing firms to undercut and replace US workers (Kirkegaard, 2005; Matloff, 2003). Critics of the H-1B program are arguing that, for the reasons discussed above, employers find hiring foreign high skilled labor an attractive alternative and that such hiring either "crowds out" natives from jobs or put downward pressure on their wages.

Immigrants may, on the other hand, have impacts on the innovative capacity of the firm. Kerr and Lincoln (2010) and Hunt and Gauthier-Loiselle (2010) provide evidence on the link between variation in immigrant flows and innovation measured by patenting, suggesting that the net impact of immigration is positive rather than simply substituting for native employment. Kerr and Lincoln (2010) also show that variation in immigrant flows at the local level related to changes in H-1B flows do not appear to adversely impact native employment and have a small, statistically insignificant effect on their wages.

Bound et al. (2015) proposes an alternative interpretation to Kerr and Lincoln (2010) results. Even when the labor demand is not close to perfectly elastic, if employers face costs to hire immigrant labor and are bound to pay the going wage, firms might disproportionately hire immigrants only when the demand for workers is increasing. In this case, immigrants would not replace incumbent workers or depress wages, but rather have a negative impact on the *growth* of wages and employment for natives. Bound et al. (2015) find that wages for computer scientists would have been 2.8-3.8% higher, and the number of Americans employed as computers scientists would have been 7.0-13.6% higher in 2004 if firms could not hire more foreigners than they could in 1994. In contrast, total CS employment would have been 3.8-9.0% lower, and consequently output smaller.

While the approach in Bound et al. (2015) is distinctly partial equilibrium in nature, Bound, Khanna, and Morales (2016) extends this analysis into a general equilibrium model of the US economy. Doing so allows them to conduct a comprehensive welfare analysis and study the distributional implications of the H-1B program. Importantly, by comprehensively modeling the firms' decisions, including the spillovers from technological innovation, they derive the labor demand curve for different types of workers. Bound, Khanna, and Morales (2016) find that even though US computer scientists are hurt by immigration, complements in production and firm entrepreneurs benefit substantially. Lastly, consumers benefit from lower prices and higher output, increasing US welfare as a whole.

While these papers focus only on the US labor market, we include the crucial role played by India as the largest contributor to this boom. Growth in the US not only helped the Indian IT sector, but also provided incentives for an educated workforce well-versed in technical knowhow for the expanding CS industry. The US boom may not have been easy to sustain for such a period without the contribution of this foreign educated workforce. For the purposes of studying only the direct impact on US wages, employment and output, these other papers do not model

the foreign side of things. However, to study the linkages across the countries and the feedback into the US industry, this paper will model what happened on both sides of the world.

1.5 The Indian IT Firm and its Relationship with the US

Unlike most Indian industries that focus on the large domestic market, the Indian IT firm is significantly export oriented; catering to a consumer base abroad that has the purchasing power for its products (Figure 2b). It is clear that most of the early-growth was export-led growth since by the turn of the century, software exports accounted for 26% of all exports, whereas in 1995 it was only 2% of all exports. Moreover, till about the end of the 1990s, most of these exports involved the physical presence of Indian workers at an overseas work-site. Over time, however, Indian IT firms moved from providing low-cost programming abroad to more comprehensive software development services for their overseas clients that was directly exported from India. Bhatnagar (2005) describes how, in 1995, 66% of all Indian IT exports involved an Indian worker on a foreign work-site, but this number fell to 29% by 2005.

The low-wage advantage is one of the earliest explanations advanced to describe the growth in Indian IT (Heeks, 1995). Arora et al. (2001) note that by the turn of the century, India had the largest number of people working in the industry and the highest revenue growth, but also the lowest revenue per employee. However, this low productivity could be a reflection of the types of services they provided – maintenance of data/legacy systems and other low-value services (Athreye, 2005).

A large part of the success of Indian firms is attributed to high-skilled Indian immigrants in the US. Bhatnagar (2005) notes that Indian professionals in Silicon Valley "*built personal networks and valuable reputations and used their growing influence within US companies to help Indian companies get a foot in the door*" of the expanding IT work. This reputation was largely built in the on-site consulting phase of the early 1990s. As Banerjee and Duflo (2000) note, reputation is essential in an industry like this because a lot of contracts are for customized software and can lead to hold-ups which a court of law may find difficult to arbitrate over.

Saxenian (1999) shows that in 1998, persons with Indian names headed 774 high-technology Silicon Valley firms overlooking sales of about $3.6 billion, which is roughly the same amount as the entire Indian IT sales bill in 1998. By the turn of the century, Indians headed 972 Silicon Valley firms, accounting for nearly 26,000 jobs. Bhatnagar (2005) claims that Indians headed about 3% of tech companies started between 1980 and 1985, but by 1995 they headed about 10% of them. At around the same time, NASSCOM estimated that about 200,000 Indian software professionals were working on H-1B visas.

The US has historically been the largest exporter of software products, and continues to produce the largest number of patents in the industry. US multinationals entered the Indian market by

setting up liaison offices and subsidiaries. While they initially intended to sell to the Indian market, they shifted to using India as a place for software development (Arora et al., 2001). By 1997, the US accounted for about 58% of the all export revenues whereas all the European countries combined accounted for only 21%. By the mid-2000s, however, India overtakes the US as the major exporter of IT products (Figure 2c).

Indian firms could use the H-1B program as a method to set up a base in the US with a ready supply of workers from India. Even as late as 2013, Indian firms are the largest sponsors of H-1B visas to the US. Even non-Indian firms are big employers of H-1Bs, some of which have Indians as their largest employment-base. Table 1 shows that 10 out of the top 11 H-1B firms have Indians as their primary employment base. Indian citizens, are therefore, the largest beneficiaries of the H-1B visa program, with about 70% of all H-1Bs in 2014 being awarded to Indians (USCIS, 2014).

These workers come in search of higher wages, as working in the US provides substantial productivity premiums (Clemens, 2013). However, there is a cap on H-1B visas, and the visas are only for 3-6 years and, a large number of potential engineers must seek work elsewhere. Workers who are unable to obtain an H-1B visa because of the cap, or return to India after their visa expires are a ready supply of labor for firms in India. This leads to a large skilled workforce in India, and enables the Indian IT sector to expand, tapping in to this growing educated workforce. The growth in Indian firms are therefore strongly linked to larger exports to the US, and in the ready supply of labor.

1.6 Indian Students and College Choice

While the abundant stock of programmers had induced recruiters to come to Indian in the early 1990s, this was sustained till the end of the decade by a steady increase in freshly trained programmers (Desai, 2003). In 2002, these software engineers were young with a median age of 26.5 years, and 58% of them had less than three years of experience (NASSCOM, 2003). In India, most programmers and the chief executives in IT companies are predominantly trained as engineers (Desai, 2003). Science graduates and those with master's degrees in computer applications make up the rest. A NASSCOM-Hewitt survey found that 88% of firms visited engineering colleges to recruit, and 47% recruited *only* there.

A survey by Arora and Athreye (2002) found that 80% of all software professionals employed had engineering degrees, and show that over time a number of engineering colleges have increased their emphasis on IT and even IT management. This, however, has meant that the number of PhDs in engineering disciplines has actually fallen from about 675 in 1987 to 375 in 1995. The industry is therefore attracting some of the brightest young graduates with little experience for highly-qualified jobs. The salaries are among the highest across industries, growing at a steady

rate, and some firms even offer stock options. Despite this, the attrition is quite high, as they "*migrate to better paid jobs in other countries*" (Kumar, 2006).

The bulk of Indian workers get their degrees at Indian universities. India has historically been better at technical education like engineering and medicine (Arora et al., 2001). Furthermore, it has a linguistic advantage over East Asian countries, due to a vast majority of Indians being fluent in English. Over the last few decades, there has also been consistent growth in the number of new undergraduate engineering schools being opened to cater to the burgeoning demand (NASSCOM, 2012). This growth in engineering schools drives the growth in undergraduate enrollment in technical institutions (Figure 2a), and a similar trend can be seen for graduate and diploma-granting institutions.

A large number of foreign students also come to the US for higher education purposes, plausibly also exploring this as a pathway to the US labor market. Many of these students stay on to obtain work visas (Bound et al., 2014). 20,000 H-1B visas are granted to students who obtained their Master's (or higher) degrees from US institutions, and this may incentivize students to obtain their degrees in US universities so that they may avail of the higher wages that come with an H-1B visa (Rosenzweig, 2006).

2 Model

We model the market for high-skill immigrants, focusing on decisions made by firms and workers in both the US and India. In each country firms are split up into two sectors – the IT sector and the final goods sector. IT sector firms hire computer scientists (CS) and other college graduate workers. Whereas firms in the final goods sector hire all three types of workers – CS college graduates, college graduates that are not computer scientists and non-college graduates. The output of IT sector firms is a direct input into the final good produced in the economy which is the only good consumed by workers.

We model trade and migration based on institutional restrictions and empirical patterns. High-skill immigration flows in one direction from India to the US and is restricted to computer scientists. Furthermore, US firms can outsource production to Indian firms or subsidiaries. These outsourced goods and services show up as imports from India to the US. Last, as we show in Figure 2c, while the US was the predominant exporter of IT goods for most of the 1990s, India takes over soon thereafter. To capture these features, IT firms are modeled in the style of Eaton and Kortum (2002), where there is a continuum of varieties and the market for each variety is perfectly competitive. The final goods sector is also perfectly competitive but produces a homogeneous good.

Importantly, we include the possibility of directed technological change (Acemoglu, 1998). Since

production in IT is heavily reliant on technology, this is an important driver of how technology spreads to India. Computers scientists in both countries are also innovators and increase the technological productivity in the IT sector. Workers that migrate from India to the US acquire human capital in the form of skills and technologies, and when they return they bring this knowledge with them. This spread of technology makes the Indian IT sector more productive, and over time the leading exporter of IT goods.[10]

On the labor supply side of the model, workers in both India and the US, choose college majors and occupations. They can pay switching costs to switch occupations based on changes in expected future wages. With a certain probability Indian CS workers get hired into the US IT sector. Given the large wage premium in working in the US IT sector (Clemens, 2013), changes in this probability will affect the choices made by Indian students and workers.

Our model, therefore, consists of two sections. In the first we discuss the how goods are produced and sold to consumers. The firm-side problem therefore details the demand for different types of workers. Every period firms maximize profit by choosing the optimal amount of the different types of labor. Indian and US firms trade in these goods with each other, and with the rest of the world, while consumers choose how much of the final good to consume every period and have no savings.

In the second section we model the labor supply decisions of college graduates. Since human capital investments and career choices have long term payoffs, workers in both countries are allowed to choose their fields of study and occupations based on the information they have today and their expected payoffs in the future. They are then allowed to switch occupations, by paying a switching cost, when a change occurs in the current or expected payoffs associated with any occupation. Given the labor supply decisions of different types of workers in both countries, the labor supply of immigrants, and the labor demand from firms in each sector, the market clears to determine the equilibrium wages for each type of worker. Equilibrium prices are determined in the product market, where the demand for the two types of goods from consumers and firms meets the supply of these goods from firms.

2.1 Product Market

2.1.1 The Household Problem

Consumers in each economy supply one unit of labor each, and have the same preferences over the final good $Y_{d,k}$ where $k = \{US, IN\}$ denotes the country and d represents that it is the

[10] We build on a growing contemporary literature on technological diffusion and directed technological change within the Ricardian trade model framework (Alvarez et al., 2013; Dasgupta, 2012; Kerr, 2013b; Lee, 2016; Perla et al., 2015; Somale, 2014)

demand for the good. The final good is assumed to be homogeneous, and consumer preferences can be represented with log utility as in the equation 1:

$$U(Y_{d,k}) = log(Y_{d,k}) \; for \; k = \{US, IN\} \tag{1}$$

While consumers have identical consumption preferences they do not receive the same labor income as they work in different occupations. Consumers maximize their utility subject to a budget constraint: $m_{occ,k} = P_{Y,k}Y_{d,k}$, where $m_{occ,k}$ is the income of the worker that has occupation *occ*, and $P_{Y,k}$ is the price of the final good. The optimal consumption is therefore $Y_{d,occ,k} = \frac{m_{occ,k}}{P_{Y,k}}$.

We outline the details of the labor-supply decisions in subsection 2.2, where we discuss how workers in each country choose their field of college-majors and occupations over time. The decision of whether to attend college or not is made outside this model which means that the supply of non college graduates $\bar{H}_k$ is exogenous, and so is the total supply of native college graduates ($\overline{L_{n,k} + G_k}$). Those who do get a college degree can choose whether to work as a computer scientists $L_{n,k}$, or in some other occupation that requires a college degree G_k.

Furthermore, workers in the US can either be native workers (denoted by a subscript n) or foreign workers (denoted by a subscript F). High skilled immigrants who come in to the US on H-1B visas can do so only if they meet the skill requirements of the visa and only if firms recruit them. Computer science H-1Bs $L_{F,US}$ make up the remaining part of the US workforce.

In the US, the size of the labor force in the economy is $\bar{H}_{US} + \overline{L_{n,US} + G_{US}} + L_{F,US}$ and total income m can be written as the sum of the labor income for the different types of workers as in equation 2:

$$m_{US} = w_{US}(L_{n,US} + L_{F,US}) + s_{US}G_{US} + r_{US}\bar{H}_{US} \; , \tag{2}$$

where w_{US} is the wage paid to computer scientists, s_{US} the wage earned by college graduate non computer scientists and r_{US} is the wage paid to non college graduates. Since there are no savings, this means that $m_{US,L} = w_{US}$ and so on.

As the wage differential between the US and India is very large, Indian computer scientists are always willing to come and work in the US. We assume they are marginally more productive (or are willing to work at marginally lower wages) than their US counterparts which means that firms first hire these foreign CS workers (subject to the H-1B cap), and native computer scientists face a residual demand curve after all available foreigners have been hired.

One way to think about this assumption in our model is that any extra productivity is almost entirely offset by the recruitment costs of hiring foreigners. Also, due to H-1B restrictions,

immigrants get paid the same wage as native computer scientists. In what remains of subsection 2.1 we will refer to foreign and native computer scientists as a single group, since from a firm's point of view they are indifferent between hiring the two at the going wage.

In India, on the other hand, there are two types of computer scientists. Those that return from the US R_{IN} after the expiration of their H-1Bs earn a wage w_R, and those natives $L_{n,IN}$ that never migrated to the US, earn a wage w_n. These workers can work in either the IT sector or the final goods sector. Therefore, in India, the size of the labor force is $\bar{H}_{IN} + \overline{L_{n,IN} + G_{IN}} + \bar{R}_{IN}$, and the total income in the economy is:

$$m_{IN} = w_{n,IN} L_{n,IN} + w_R R_{IN} + s_{IN} G_{IN} + r_{IN} \bar{H}_{IN} \ , \tag{3}$$

2.1.2 Final Goods Production

The final goods sector is perfectly competitive and produces a homogeneous good. The representative firm in this sector has a Cobb Douglas constant returns to scale technology over intermediate inputs from the IT sector $C_{y,k}$ and the labor aggregate X_k:

$$Y_k = C_{y,k}^{\gamma_k} X_{y,k}^{1-\gamma_k} \ , \tag{4}$$

where $C_{y,k}$ represents intermediate inputs from the IT sector and $X_{y,k}$ the labor aggregate. These firms have preferences over the different types of IT goods $c_{j,k}$, such that:

$$C_{y,k} = \left(\int_{j \in \Omega} c_{j,k}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}} \ , \tag{5}$$

where Ω is the set of varieties and σ is the elasticity of substitution between the varieties of IT. Given this setup, it is possible to write the price index P in the form of equation 6:

$$P_{c,k} = \left(\int_{j \in \Omega} p_{j,k}^{1-\sigma} dj \right)^{\frac{1}{1-\sigma}} \tag{6}$$

The optimal IT bundle can then be represented by equation 7:

$$\left(\frac{c_{j,k}}{C_{y,k}} \right)^{-\frac{1}{\sigma}} = \frac{p_{j,k}}{P_{c,k}} \tag{7}$$

This sector employs the three types of labor denoted by subscript y. $X_{y,k}$ is a labor aggregate that aggregates non college graduates $H_{y,k}$ and an aggregate of college graduates $Q_{y,k}$:

$$X_{y,k} = \left[\alpha^{y,k} H_{y,k}^{\frac{\tau-1}{\tau}} + (1-\alpha^{y,k}) Q_{y,k}^{\frac{\tau-1}{\tau}}\right]^{\frac{\tau}{\tau-1}} \tag{8}$$

Using a nested CES format, the aggregate of college graduates $Q_{y,k}$ can be represented by:

$$Q_{y,k} = \left[\delta_k L_{y,k}^{\frac{\lambda-1}{\lambda}} + (1-\delta_k) G_{y,k}^{\frac{\lambda-1}{\lambda}}\right]^{\frac{\lambda}{\lambda-1}} , \tag{9}$$

where $L_{y,k}$ is the number of computer scientists hired in the final goods sector, and $G_{y,k}$ is the number of non computer scientists hired in the final goods sector. Both sectors have the same elasticity of substitution between college and non college graduates (τ) and between computer scientists and non-CS college graduates (λ).

In India, on the other hand, firms pay different wages to native CS workers and the return migrants. The $L_{y,k}$ computer science labor is an aggregate over the native $L_{n,y,k}$ and return migrants $R_{y,k}$:

$$L_{y,k} = \left[\Psi L_{n,y,k}^{\frac{\epsilon-1}{\epsilon}} + (1-\Psi) R_{y,k}^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \text{ for } k = \{IN\} , \tag{10}$$

The first order conditions determine the demand for the intermediate inputs and the different types of labor in this sector. Together with the demand for labor from the IT sector we can derive the aggregate labor demand for each worker.

2.1.3 Production in the IT sector

The IT sector produces a continuum of varieties $c_{j,k}$ where each variety is indexed by type j. Following the framework introduced by Dornbusch et al. (1977) and Eaton and Kortum (2002), each country k will have a different level of efficiency in producing an IT variety, denoted by $z_{k,j}$. In order to capture the different the contribution of the different types of labor, the IT firm has a CES technology in the labor aggregate as in equation 11:

$$c_{j,k} = z_{j,k} A_k \left[(\delta_k + \Delta_k) \ell_{j,k}^{\frac{\lambda-1}{\lambda}} + (1-\delta_k-\Delta_k) g_{j,k}^{\frac{\lambda-1}{\lambda}}\right]^{\frac{\lambda}{\lambda-1}} , \tag{11}$$

where $\ell_{j,k}$ is the number of CS workers and $g_{j,k}$ the non-CS college graduates employed by firm j. Here λ is the elasticity of substitution between the CS workers and non-CS college graduates and $\delta_k + \Delta_k$ is the distributional parameter of the CES function. We impose $\Delta_k > 0$ to indicate that the IT sector is more intensive in computer scientists than the final goods sector.

In India, we assume that native and return migrant CS workers are not perfect substitutes as return migrants may have a different set of skills:

$$\ell_{j,k} = \left[\Psi \ell_{n,j,k}^{\frac{\epsilon-1}{\epsilon}} + (1-\Psi) R_{j,k}^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \text{ for } k = \{IN\} \ , \tag{12}$$

where $\ell_{n,j,IN}$ is the number of CS workers that never went abroad and $R_{j,k}$ the number of return migrant CS workers employed by firm j. Here ϵ is the elasticity of substitution between the native CS workers and return migrants.

A_k represents the country-level firm productivity in IT. We model the relative productivity between the US and India to be a function of the total number of CS IT workers in in India, relative to the US: $\frac{A_{IN}}{A_{US}} = A\left(\frac{L_{n,c,IN}+R_{IN}}{L_{n,c,US}+L_{F,c,US}}\right)$. This captures the notion that relative productivity in the IT sector depends on the relative number of the computer scientists in this sector. If computer scientists are innovators, they increase the productivity in the IT sector.

For each IT variety we assume that there are infinitely small firms with constant returns to scale technology willing to produce the good. The marginal cost of production for each firm is therefore constant and together with the individual productivity is equal to the price as in equation 13:

$$p_{j,k} = MC_{j,k} = \frac{\xi_{j,k}(w_k, s_k, A_k)}{z_k} \ , \tag{13}$$

Like Eaton and Kortum (2002) we assume that country k's efficiency in producing good j is the realization of the random variable Z_k, drawn independently for each j from a distribution $F_k(z)$. We assume that the productivity $z_{j,k}$ comes from the Frechet (Type II extreme value) distribution:

$$F_k(z) = e^{-z^{-\theta}}, \tag{14}$$

where $\theta > 1$ reflects the variation within the distribution.[11] Given this set-up, we can calculate the aggregate demand for different types of workers in the IT sector by integrating the individual firm demand across all varieties. In the US we get aggregate variables $L_{C,US}$ and $G_{C,US}$ while in India, given there are two types of computer scientists we get aggregates for $L_{n,C,IN}$, $R_{C,IN}$ and $G_{C,IN}$.

2.1.4 International Trade

The US and Indian economies are the only producers of IT goods, and they export different varieties to each other and to the rest of the world (W). While we do not explicitly model outsourcing decisions, we do allow for the fact that Indian IT goods can be exported to the

[11] Note that this is isomorphic to a case where $A_k = 1$ and $F_k(z) = e^{-T_k z^{-\theta}}$, where a higher T_k increases the likelihood of drawing a higher efficiency for good j.

US, and are an intermediate input into the US final good. The fraction of IT goods imported from IT producing country k depends on the marginal costs of production, and therefore the productivity parameter A_k, in each country. Note that under the absence of trade costs, the country that is more efficient at producing a specific IT variety will be the sole supplier of such variety in our economy.

In the absence of any trade costs, perfect competition implies that buyers in country k will buy the good from the cheapest source, such that the price of the good in country k is:

$$p_{k,j} = \min\{p_{US,j}, p_{IN,j}\} \ , \tag{15}$$

Since countries in the rest of the world do not produce IT goods, they will also choose to buy from either India or the US depending on which country sells the good at a lower price.

Following the setup of a two-country Eaton-Kortum model we can get a closed form solution for the probability that country $k = \{US, IN\}$ supplies a particular IT variety to the rest of the world which is equivalent to the probability that country k sells the good at the lowest price. The probability that a country (the US, India or anywhere else in the world) buys a good produced in country k is given by π_k:

$$\pi_k = \text{Prob}(p_{j,k} < p_{j,-k}) = \frac{(\xi_{j,k}(w_k, s_k, A_k))^{-\theta}}{(\xi_{j,k}(w_k, s_k, A_k))^{-\theta} + (\xi_{j,-k}(w_{-k}, s_{-k}, A_{-k}))^{-\theta}} \ \ for\ k = \{US, IN\}, \tag{16}$$

Consumers in the rest of the world (W) have the same utility function as in India and the US $U_W(Y_W) = log(Y_W)$, and the final goods production in the rest of the world also takes on the same form:

$$Y_W = C_{y,W}^{\gamma_W} X_W^{1-\gamma_W} \tag{17}$$

This implies that the demand for the IT good from the rest of the world is characterized by:

$$C_{y,W} = \frac{\gamma_W \ m_W}{P_{c,W}} \tag{18}$$

Therefore, on net, IT firms export final goods to consumers in other countries $C_{y,W}$, whereas US and Indian consumers import the final good from the rest of the world Y_W. Imports into the US and India from the rest of the world are represented by $Y_{IM,k}$ for $k = \{US, IN\}$. For convenience we assume trade is balanced implying that the value of imports must equal the value of exports in both India and the US. Since India and the US trade between each other

in IT goods only, this implies that the value of IT good flows from India to the US equals the value of IT good flows from the US to India. Similarly, the value of IT good flows from the US to the rest of the world must equal the value of the final goods flowing from the rest of the world to the US.

Section 2.2 describes the supply of the different types of workers, and Section 2.3 describes the equilibrium, where we also detail how the labor demand curve in the US shifts over time given the technological boom in the 1990s.

2.2 The Supply of Workers in India and the US

The labor supply side of the model is related to Bound et al. (2015) who model the US labor market for computer scientists. College graduates in the US and India make two types of decisions along their career in order to maximize the expected present value of their life time utility. At age 20, individuals in college choose the field of study that influences their initial occupation after graduation, and from age 22 to 65, workers choose between working as a computer scientist or in another occupation. Individuals have rational, forward looking behavior and make studying and working decisions based on the information available at each period. Indian workers that return from the US always work as computer scientists.

2.2.1 Field of Study Decision

At age 20, an individual i draws idiosyncratic taste shocks for studying computer science or another field: $\eta^c_{i,k}$ and $\eta^o_{i,k}$,respectively (where $k = \{IN, US\}$ indexes the Indian students in India and US students in the US). This student also has expectations about the prospects of starting a career in each occupation after graduation (age 22), which have a values $V^c_{22,k}$ and $V^o_{22,k}$ respectively. With this information, an individual chooses between pursuing computer sciences or a different choice of major at the undergraduate level.

The utility of a student is modeled as a linear function of the taste shocks and career prospects in each sector. There is also a taste attractiveness parameter $\zeta_{o,k}$ for studying a different field from computer science and individuals discount their future with an annual discount factor β. With these assumptions, the field of study decision for $k = \{IN, US\}$ is represented by:

$$max\{\beta^2 \mathbb{E}_t V^c_{22,k} + \eta^c_{i,k},\ \beta^2 \mathbb{E}_t V^o_{22,k} + \zeta_{o,k} + \eta^o_{i,k}\} \tag{19}$$

It is assumed that $\eta^c_{i,k}$ and $\eta^o_{i,k}$ are independently and identically distributed and for $s = \{c, o\}$, can be defined as $\eta^s_{i,k} = \sigma_{0,k} v^s_{i,k}$, where $\sigma_{0,k}$ is a scale parameter and $v^s_{i,k}$ is distributed as

a standard Type I Extreme Value distribution. This distributional assumption is common to dynamic discrete choice models (Rust, 1987) and it is convenient because it allows the decisions of agents to be smoothed out, a desired property that will be used in the characterization of the equilibrium of the model.

Given the distributional assumption of idiosyncratic taste shocks, it follows that the probability of a worker *graduating* with a computer science degree can be written in logistic form:

$$q^c_{t,k} = [1 + exp(-(\beta^2 \mathbb{E}_{t-2}[V^c_{22,k} - V^o_{22,k}] - \zeta_{o,k})/\sigma_{0,k})]^{-1} \tag{20}$$

Note that the important parameter for how studying choices of workers are sensitive to different career prospects is the standard deviation of taste shocks. Small values of $\sigma_{0,k}$ imply that small changes in career prospects can produce big variations in the number of students graduating with a computer science degree.

The next step to characterize the supply of young computer scientists is to map the graduating probability described above to employment. Defining $M^a_{t,k}$ as the exogenous number of college graduates with age a in time period t, the number of recent graduates with a computer science degree in year t is represented by $L^{grad}_{t,k} = q^c_{t,k} M^{22}_{t,k}$ for $k = \{IN, US\}$.

2.2.2 Occupational Choice

The field of study determines if an individual enters the labor market as either a computer scientist or with a different occupation. However, individuals can choose to switch occupations along their careers. Specifically, at the beginning of each period, individuals between ages 22 and 65 choose to work in CS or another type of job in order to maximize the expected present value of their lifetime utility.

A feature of the model is that switching occupations is costly for the worker. A justification for this assumption is that workers have occupational-specific human capital that cannot be transferred (Kambourov and Manovskii, 2009). It is assumed that the cost to switch occupations is a quadratic function of a worker's age. Note that this assumption implies that it becomes increasingly harder for workers to switch occupations as they get older. Additionally, there is no general human capital accumulation and wages do not vary with the age of a worker.

In the equations below, the value functions for Indian workers in India ($k = IN$) and US workers in the US ($k = US$) are represented. There is no value function for Indian workers in the US as it is assumed that (a) they would rather work in the US than in India because of the wage premium, and (b) they cannot shift occupations as their H-1B is tied to their employer.

Finally, it is assumed that workers have linear utility from wages, taste shocks and career prospects. Furthermore, wages must be totally consumed in that same year and workers cannot save or borrow. The Bellman equations of worker i at age a between 22 and 64 at time t if he starts the period as a computer scientist or other occupation are respectively:

$$V^c_{t,a,k} = max\{w_{t,k} + \beta \mathbb{E}_t V^c_{t+1,a+1,k} + \varepsilon^c_{i,t,k},\ s_{t,k} - \chi_k(a) + \beta \mathbb{E}_t V^o_{t+1,a+1,k} + \varepsilon^o_{i,t,k} + \zeta_{1,k}\} \tag{21}$$

$$V^o_{t,a,k} = max\{w_{t,k} - \chi_k(a) + \beta \mathbb{E}_t V^c_{t+1,a+1,k} + \varepsilon^c_{i,t,k},\ s_{t,k} + \beta \mathbb{E}_t V^o_{t+1,a+1,k} + \varepsilon^o_{i,t,k} + \zeta_{1,k}\} \tag{22}$$

where $\chi_k(a) = \chi_{0,k} + \chi_{1,k}a + \chi_{2,k}a^2$, is the monetary cost of switching occupation for an age a worker, and $\zeta_{1,k}$ is the taste attractiveness parameter for not working as a computer scientist. In the model, all workers retire at age 65 and their retirement benefits do not depend on their career choices. As a consequence, workers at age 65 face the same decision problem but, without consideration for the future.

The current wage in the other occupation $s_{t,k}$ is exogenous and perfectly anticipated by the workers. While $w_{t,US}$ is the wage in the US, $w_{t,IN}$ is a weighted average of the wages in the US ($w_{t,US}$) and India ($w_{t,n,IN}$) where the weights are the probability of finding employment in the US:

$$w_{t,IN} = q^c_{t,US} w_{t,US} + (1 - q^c_{t,US}) w_{t,n,IN} \tag{23}$$

Here, the probability of getting a US job is determined by the fraction of Indian computer science workers that are recruited to work in the US every year (determined by the H-1B cap $\bar{h1b}$):

$$q^c_{t,US} = \frac{\bar{h1b}}{\bar{h1b} + L_{n,t,IN}} \tag{24}$$

As in the college-major decision problem, idiosyncratic taste shocks play an important role in working decisions of an individual. Once more, it is assumed that taste shocks are independently and identically distributed and for $s = \{c, o\}$ can be defined as $\varepsilon^s_{i,t,k} = \sigma_{1,k} v^s_{i,t,k}$ where $\sigma_{1,k}$ is a scale parameter and $v^s_{i,k}$ is distributed as a standard Type I Extreme Value distribution.

Defining $p^{sS}_{t,a,k}$ as the probability that a worker at age a between 22 and 64 moves from occupation s to occupation S, it follows from the error distribution assumption that the migration probabilities can be represented as:

$$q^{oc}_{t,a,k} = [1 + exp(-(w_{t,k} - s_{t,k} - \chi_k(a) - \zeta_{1,k} + \beta\mathbb{E}_t[V^c_{t+1,a+1,k} - V^o_{t+1,a+1,k}])/\sigma_{1,k})]^{-1} \tag{25}$$

$$q^{co}_{t,a,k} = [1 + exp(-(s_{t,k} - w_{t,k} - \chi(a) + \zeta_{1,k} + \beta\mathbb{E}_t[V^o_{t+1,a+1,k} - V^c_{t+1,a+1,k}])/\sigma_{1,k})]^{-1} \tag{26}$$

and the occupational migration probabilities of workers at age 65 are the same without discounting future career prospects. Note that the switching probabilities depend upon both the current wage differential and expected future career prospects at each occupation. The standard deviation of the taste shocks, the sector attractiveness constant and the cost of switching occupations will effect the extent to which changes in relative career prospects affect the movement of US residents across fields.

A feature of dynamic models with forward looking individuals is that working decisions depend upon the equilibrium distribution of career prospects. As in the dynamic choice literature with extreme value errors (Rust, 1987), one can use the properties of the idiosyncratic taste shocks distribution to simplify the expressions for the expected values of career prospects. As a result, the expected value function for an individual at age a between 22 and 64 working as a computer scientists or in another occupation are respectively:

$$\mathbb{E}_t V^c_{t+1,a+1,k} = \sigma_{1,k}\mathbb{E}_t[\varpi + ln\{exp((w_{t+1,k} + \beta\mathbb{E}_{t+1}V^c_{t+2,a+2,k})/\sigma_{1,k}) + exp((s_{t+1,k} - \chi_k(a) + \zeta_{1,k} + \beta\mathbb{E}_{t+1}V^o_{t+2,a+2,k})/\sigma_{1,k})\}] \tag{27}$$

$$\mathbb{E}_t V^o_{t+1,a+1,k} = \sigma_{1,k}\mathbb{E}_t[\varpi + ln\{exp((s_{t+1,k} + \zeta_{1,k} + \beta\mathbb{E}_{t+1}V^o_{t+2,a+2,k})/\sigma_{1,k}) + exp((w_{t+1,k} - \chi_k(a) + \beta\mathbb{E}_{t+1}V^c_{t+2,a+2,k})/\sigma_{1,k})\}] \tag{28}$$

where $\varpi \cong 0.577$ is the Euler's constant and the expectations are taken with respect to future taste shocks. Workers at age 65 face the same expected values but don't discount the future.

In transferring the occupational migration probabilities to employment, the first step is to determine the CS supply of recent college graduates. After leaving college, individuals can start their careers in the occupation correspondent to their field of study with no cost. However, we also allow workers at age 22 to pay the switching costs and get their first job in an occupation different from their field of study. As a consequence, the number of computer scientists at age 22 is a function of the number of recent graduates with a computer science degree and the migration probabilities:

$$L^{22}_{t,n,k} = (1 - q^{co}_{t,22,k})L^{grad}_{t,k} + q^{oc}_{t,22,k}[(L^{22}_{t,n,k} + G^{22}_{t,k}) - L^{grad}_{t,k}] \tag{29}$$

$$G^{22}_{t,k} = (1 - q^{o,cs}_{t,22,k})[(L^{22}_{n,t,k} + G^{22}_{t,k}) - L^{grad}_{t,k}] + q^{cs,o}_{t,22,k}L^{grad}_{t,k} \tag{30}$$

where $L^{grad}_{t,k}$ is the number of recent graduates with a computer science degree, and $(L^{22}_{t,n,k} + G^{22}_{t,k}) - L^{grad}_{t,k}$ is the number of college graduates with any other degree. In the same way, the supply of native computer scientists at age a from 23-65 is a function of past employment in each occupation and the occupational migration probabilities:

$$L^{a}_{t,n,k} = (1 - q^{co}_{t,a,k})L^{a-1}_{t-1,n,k} + q^{oc}_{t,a,k}[G^{a-1}_{t-1,k}] \tag{31}$$

$$G^{a}_{t,k} = (1 - q^{o,cs}_{t,a,k})G^{a-1}_{t-1,k} + q^{cs,o}_{t,a,k}[L^{a-1}_{n,t-1,k}] \tag{32}$$

where $G^{a}_{t,k}$ is the number of workers at age a working in the residual sector and $L^{a}_{n,t,k}$ is the number of native CS workers at age a.

The aggregate domestic labor supply of computer scientists and other workers is the sum across all ages:

$$L_{n,t,k} = \sum_{a=22}^{a=65} L^{a}_{n,t,k} \tag{33}$$

$$G_{t,k} = \sum_{a=22}^{a=65} G^{a}_{t,k} \tag{34}$$

Here we can see that the labor supply in each occupation depends on past employment, new college graduates and on wages through the occupational switching probabilities.

2.3 Equilibrium

Equilibrium in each period can be defined as a set of prices and wages ($P_{c,t,k}$, $p_{j,t,k}$, $P_{Y,t,k}$, $w_{t,k}$, $w_{n,t,k}$, $w_{R,t,k}$, $s_{t,k}$, $r_{t,k}$), quantities of output and labor ($C^{*}_{y,t,k}$, $Y^{*}_{t,k}$, $C^{*}_{t,k}$, $Y^{*}_{d,t,k}$, $Y^{*}_{IM,t,k}$, $L^{*}_{n,t,k}$, $L^{*}_{F,t,US}$, $R^{*}_{t,IN}$, $G^{*}_{t,k}$, $H^{*}_{t,k}$), share of IT goods produced ($\pi^{*}_{t,k}$), and the level of productivity ($A_{t,k}$) such that:[12]

[12] Note that we've introduced a t subscript to each of the variables to denote that there is a different equilibrium for each time period

- Consumers in the US, India and the rest of the world, maximize utility by choosing $Y_{t,k}$ taking prices as given, and choose their field of major and occupations taking wages as given.
- Final good producers choose the optimal $C_{t,k}$ given the prices of each variety.
- Firms in both the IT sector and the final goods sector maximize profits taking wages and prices as given.
- $\pi_{t,IN}$ fraction of the IT expenditure is spent on goods from India, and similarly the fraction of IT expenditure on US produced goods is $\pi_{t,US}$.
- Output and labor markets clear as in equations 35 - 43

Total consumer expenditure equals labor income (equations 35-36):

$$P_{y,t,k}Y^*_{d,t,k} = m_{t,k} = w_{t,k}(L^*_{n,t,k} + L^*_{F,t,k}) + s_{t,k}G^*_{t,k} + r_{t,k}H^*_{t,k} \; for \; k = \{US\} \tag{35}$$

$$P_{y,t,k}Y^*_{d,t,k} = m_{t,k} = w_{n,t,k}L^*_{n,t,k} + w_{R,t,k}R^*_{t,k} + s_{t,k}G^*_{t,k} + r_{t,k}H^*_{t,k} \; for \; k = \{IN\} \tag{36}$$

Total IT production in the US and India equals domestic consumer demand in each country, including the rest of the world (equation 37):

$$C^*_{t,IN} + C^*_{t,US} = C^*_t = C^*_{y,t,IN} + C^*_{y,t,US} + C^*_{y,t,W} \tag{37}$$

Total quantity produced and imports in the final goods sector, equals domestic consumer demand (equation 38):

$$Y^*_{d,t,k} = Y^*_{t,k} + Y^*_{IM,t,k} \; for \; k = \{US, IN\} \tag{38}$$

Trade in IT goods between the US and India is balanced:

$$\pi_{t,IN}P^*_{c,t,IN}C^*_{y,t,US} = \pi_{t,US}P^*_{c,t,US}C^*_{y,t,IN} \tag{39}$$

Trade with the rest of the world is balanced:

$$\pi_{t,k}P^*_{c,t,k}C^*_{y,t,W} = P^*_{y,t,W}Y^*_{IM,t,k} \; for \; k = \{US, IN\} \tag{40}$$

Given that the supply of non college graduates is inelastic $\bar{H}_t$ and they can only be employed in the final goods sector, their labor market clears as in equation 41:

$$\bar{H}_{t,k} = H^*_{y,t,k} \; for \; k = \{US, IN\} \tag{41}$$

Total labor supply for college graduates (CS and non-CS) is fixed, such that total demand for college graduates has to be equal to total supply in each period (equation 42-43):

$$\overline{L_{n,t,US} + G_{t,US}} + \bar{L}_{F,t} = L^*_{t,US} + G^*_{t,US} = L^*_{c,t,US} + L^*_{y,t,US} + G^*_{c,t,US} + G^*_{y,t,US} \tag{42}$$

$$\overline{L_{n,t,IN} + G_{t,IN}} + \bar{R}_t = L^*_{t,IN} + G^*_{t,IN} = L^*_{c,t,IN} + L^*_{y,t,IN} + G^*_{c,t,IN} + G^*_{y,t,IN} \tag{43}$$

Native college graduates face the decision of whether to work as computer scientists or in some other occupation that requires a college degree. This decision has an inter-temporal dimension which requires the definition of the dynamic equilibrium in the labor market for college graduates. As in Bound et al. (2015), this equilibrium is characterized by the system of equations (19 - 34) and a stochastic process $Z_{t,k}$. In particular, equations 27 and 28 characterize the expectations of workers with respect to future career prospects and equations 33 and 34 describe the dynamic labor supply of US computer scientists and other college graduates respectively.

A unique equilibrium is pinned down each period by an aggregate labor demand curve for US computer scientists relative to other college graduates that comes from the product market model. Even though this labor demand curve from the two sectors has no closed form solution we will express it as in equation 44-45, a setup that will prove to be useful for the calculations in the following sections.

$$\frac{L_{n,t,k}}{G_{t,k}} = Z_{t,k} + \Upsilon_k \left(\frac{w_{t,k}}{s_{t,k}} \right) \; for \; k = \{US\} \tag{44}$$

$$\frac{L_{n,t,k}}{G_{t,k}} = Z_{t,k} + \Upsilon_k \left(\frac{w_{n,t,k}}{s_{t,k}} \right) \; for \; k = \{IN\} \tag{45}$$

where $\Upsilon_k(.)$ is a baseline relative demand curve that depends on the relative wage. $Z_{t,k}$ is a shifter that can be thought of as a combination of the productivity shocks from the IT boom, that shifts out the relative demand for computer scientists every year and the cap of foreign computer scientists $\bar{L}_F$ that moves the relative demand curve every period differently for each country. For US workers, the relative inflow of foreign workers $\frac{L_{F,t,US}}{G_{t,US}}$ shifts in the relative demand curve. Whereas, for Indian workers, the relative prospects of foreign jobs, $\frac{L_{F,t}}{G_{t,IN}}$ shifts out the relative demand curve every year. $Z_{t,k}$ is assumed to follow a random walk process with high persistence such that:

$$Z_{t,k} = 0.999 Z_{t-1,k} + 0.001 \bar{Z}_k + \nu_{t,k} \tag{46}$$

where $\bar{Z}_k$ is the steady state value of $Z_{t,k}$ and $\nu_{t,k}$ is an i.i.d. shock.[13]

The flow of workers, though, also shifts the relative productivity of the IT sector A_k, where the relative productivity depends on the relative CS workforce:

$$\frac{A_{IN}}{A_{US}} = A \left(\frac{L_{n,IN} + R_{IN}}{L_{n,US} + L_{F,US}} \right) , \tag{47}$$

We abstract from explicitly modeling the return migration decision. Due to the large wage premiums, it is reasonable to expect that Indian CS workers would not have returned in the absence of regulations under the H-1B law. After the 6-year limit on the H-1B visa, firms that wish to hold on to a worker must sponsor them for a green-card. This is an administratively and monetarily costly process, and once the worker receives the green-card she is free to switch workers or work for multiple employers. We assume that the number of workers who return are a fixed fraction of the number of H-1B visas that expire that year. This fraction is represented by the parameter ϱ.

The equilibrium in the labor market can be expressed by a mapping from the state variables: $s = \{\mathrm{R_{t,k}}, L^{grad}_{t,k}, L^{22}_{n,t-1,k}, ..., L^{64}_{n,t-1,k}, G^{22}_{t-1,k}, ..., G^{64}_{t-1,k}, Z_{t-1,k}\}$ and exogenous productivity shock $\nu_{t,k}$ to the values of $L_{n,t,k}$, $w_{t,k}$, $w_{R,t,k}$, $w_{n,t,k}$, $G_{t,k}$, $s_{t,k}$ and $\mathbf{V_t}$, the vector of career prospects at different occupations for different ages, that satisfies the system of equations 19 to 34 as well as each period's relative demand curve.

3 Calibration

We now proceed to calibrate the parameters of the model. In total we have 25 parameters in the product market: τ, λ, ϵ, σ, θ, $h\bar{1}b$, ϱ, $\alpha_{y,US}$, $\alpha_{y,IN}$, δ_{US}, δ_{IN}, Δ_{US}, Δ_{IN}, Ψ, $\bar{L}_{n,US}$, $\bar{G}_{US}$, $\bar{H}_{US}$, $\bar{N}_{IN}$, $\bar{R}_{IN}$, $\bar{G}_{IN}$ and $\bar{H}_{IN}$ in addition to the relative productivity parameters. For the labor suply of college graduates we have 8 parameters for India and the US: $\sigma_{0,k}, \zeta_{0,k}, \sigma_{1,k}, \zeta_{1,k}, \chi_{0,k}, \chi_{1,k}, \chi_{2,k}$,and β.

We follow a sequential approach in order to back out all the parameters of the model. As a first step we take the labor supply of computer scientists, non-CS college graduates and non college graduates as given and calibrate the parameters of the product market in order to match certain features of the data. Once we have our product market parameters we derive a demand curve for computer scientists relative to non-CS college graduates, for every year. Calibration

[13] We assume workers consider both the technological progress from the IT boom as well as the flow of migrants from India to the US to be a series of highly persistent shocks.

of these parameters and the demand curve are summarized in subsection 3.1. In a second step, we use the shifts in the relative demand curve to calibrate the labor supply parameters and trace out the labor supply curve as explained in subsection 3.2. Finally, we use all calibrated parameters to run our counterfactual simulations.

3.1 Product Market Calibration

We calibrate the model each year between 1994-2010 and present a summary of all calibrated parameters for selected years in Table 4. A detailed summary of all our data sources can be found in Appendix A.

3.1.1 Production Function Parameters

We set the elasticity of substitution between different factors to be time invariant and the same across all countries and sectors in our model. We calibrate the elasticity of substitution between college and non college graduates, $\tau = 1.7$ based on an average of different papers that estimate that parameter such as Katz and Murphy (1992), Card and Lemieux (2001) and Goldin and Katz (2007).[14] For the elasticity of substitution between computer scientists and non-CS college graduates we set $\lambda = 2$ which is within the estimates of Ryoo and Rosen (2004) who estimate the elasticity of substitution between engineers and other college graduates to be between 1.2 and 2.2. In the case of India we also need to calibrate the substitution between computer scientists who never worked abroad and those who return from the US. We would expect this elasticity to be greater than the elasticity between CS and non-CS graduates, and set $\epsilon = 3$.

For the substitution between varieties of IT in each country, we follow Bernard et al. (2003) who estimate the elasticity of substitution across US plants to be 3.79 and set $\sigma = 4$. Finally, we also calibrate the Frechet dispersion parameter $\theta = 8.28$ using the Eaton and Kortum (2002) preferred value.

We calibrate the rest of the parameters in the production function separately by country and year in order to capture the differential changes in technology that are going on in the US and India during our period of analysis.

The Cobb Douglas parameters γ_k represent the share of income from the final goods sector spent on varieties of the IT sector. We calibrate the parameters for $k = \{US, IN\}$ using the share of IT GDP to total GDP in each country and get values: $\gamma_{US} = 0.021$ and $\gamma_{IN} = 0.005$ in 1995. By calibrating these parameters every year we want to capture the changes in demand

[14] These papers estimate the overall substitution between college and non college graduates, while our parameter is sector specific. However, when calculating the overall substitution between college and non college graduates our estimates are indistinguishable from our assigned value of τ.

for IT varieties as an input into the final good production, and we can see that while it is increasing for both countries, it shows a larger increase for India over this period.

For the demand of IT goods from the Rest of the World W we use exports and GDP data from the OECD to first calculate the imports of IT products from the US and India as a share of the GDP from the rest of the world and calibrate γ_W. We then calculate the relative GDP of the rest of the world with respect to the combined GDP of US and India to match the size of the rest of the world with respect to the US and India.

The final goods production function distributional parameter $\alpha_{y,k}$ is calibrated in India and the US such that it matches the share of expenditures from the final goods sector in non college graduates. More specifically, from the March CPS for the US and the NSS data for India, we calculate the share of expenditures on non college graduates $\vartheta_{y,k}$ and the number of college and non college graduates in the final goods industry $\bar{H}_{y,k}$, $\bar{Q}_{y,k}$. We calibrate the parameter to be 0.47 in both the US and India in 1995 using equation 48:

$$\vartheta_{y,k} = \frac{\alpha_{y,k}\bar{H}_{y,k}}{\alpha_{y,k}\bar{H}_{y,k} + (1-\alpha_{y,k})\bar{Q}_{y,k}} \tag{48}$$

Importantly, in Table 4 it is clear that $\alpha_{y,k}$ decreases over time, capturing how skill-biased technological change shifts production to college graduate occupations over time.

The distributional parameter between CS and non-CS college graduates δ_k is calibrated so that it matches the relative wages between computer scientists and non-CS college graduates observed in the data. We calibrate values of δ to be 0.176 in the US and 0.147 in India in 1995. This parameter increases over time capturing how shifts in skill-biased technological increase the labor share of CS workers.

The additional distributional parameter in the IT sector Δ_k captures the extra intensity of CS in the IT sector. We calibrate Δ_k such that it matches the share of expenditures of the IT sector in computer scientists in the US for 1994 and get a value close to 0.2. For simplicity we use that value for both US and India throughout the period. Finally, for India we assign a value of 0.5 for Ψ, the distributional parameter between CS that never migrated to the US and CS that return from the US.

3.1.2 Labor quantities

We calculate the number of each type of worker for each country. For the US, we use the March CPS data to get the number of computer scientists, non-CS college graduates and non-college graduates each year to match the shares of employment in each occupation. For tractability, we normalize the total population in 1994 to be 100 and then let the population grow at the same rate that the economically active population grows in the data. We use the CPS-ORG

data to find the share of foreign computer scientists that are Indian and use that to calculate our foreign computer scientists measure. All foreigners who are not Indian computer scientists are considered native workers for the current purposes of our study.

We calculate the average yearly change of Indian CS in the US between 1994-2010 and define the average change as the H-1B cap: $h\bar{1}b$, the number of foreign computer scientists who enter the US every year. According to Lee (2016), the OECD estimates that $\varrho = 23.5\%$ of high-skill immigrants in the US return to their home countries after a 5-year period so we can write the law of motion for foreign CS in the US as in equation 49:

$$L_{F,t+1,US} = L_{F,t,US} + (1 - \varrho)h\bar{1}b \tag{49}$$

For India we follow a slightly different approach. We use the National Sample Survey (NSS) to get the share of each group in each year and we use World Bank data to get at total active population each year. Multiplying the shares by the total population we get the total employment under each occupation. We then normalize the total size of the population in India every year to make it's relative size with respect to the US match the data.

The NSS does not provide information that allows us to distinguish between computer scientists that never migrated to the US and those that are return migrants. To tackle this issue we create the series of return migrants based on the Indian computer scientists that are working in the US since 1980. We use the 1980 and 1990 US censuses to calculate, on average, how much has the Indian CS population been increasing every year. From the average yearly increase between 1980-1994 we assume that each year, $\varrho\%$ of that average return to India and create the series as the cumulative of those that go back in 1980 up until those that go back in 1994.

Once we get our initial stock for 1994, every year we assume that the number of return migrants evolves according to equation 50:

$$R_{t+1,IN} = R_{t,IN} + \varrho\ h\bar{1}b \tag{50}$$

A summary of the calibrated employment can be found in table 5.

3.1.3 Relative productivity changes

In a first stage we calibrate the relative productivity of the IT sector in the US with respect to India ($\frac{A_{US}}{A_{IN}}$) such that it matches the relative wage of computer scientists in the US with respect to India. One of the key assumptions of our model is that given that computer scientists are innovators, the productivity differences in the sector between the two countries depend on the relative size of the computer science workforce working for the sector. To capture how

relative productivity evolves over time, if our assumption holds true, we regress the calculated $\frac{A_{US}}{A_{IN}}$ on the log of the number of CS working for IT in the US relative to the CS working for IT in India. We then use the estimated coefficients to characterize the relationship between relative productivity and the relative size of CS working for IT when doing our counterfactual simulations.

3.1.4 Labor demand

Once we calibrate all the product market parameters and the relative productivity process we proceed to derive the labor demand curve for our baseline year 1994. As the notation is slightly different, we first describe the process for the US and then describe the differences of deriving the relative labor demand for India. Since our demand curve has no closed form solution we approximate it through the following process: we change the relative quantity of computer scientists to non-CS college graduates ($\frac{\bar{L}_{US}}{\bar{G}_{US}}$) leaving the total quantity of college graduates and all other parameters fixed and calculate, for each $\frac{\bar{L}_{US}}{\bar{G}_{US}}$ what is the relative wage $\frac{w}{s}$ predicted by the model.

We calculate the relative wages for relative labor quantities between 0.03-0.055 in order to capture the shape of the demand curve across the relative labor values we observe in the data during our period. We then run a regression of relative quantities of CS to non-CS college graduates on a quadratic of predicted relative wages.[15] This allows us to express the demand curve as in equation 51:

$$\frac{\hat{L_{US}}}{G_{US}} = \hat{f}(\frac{w_{US}}{s_{US}}) \tag{51}$$

Where $\hat{f}(.)$ are the estimated coefficients of the relative labor demand curve for the baseline year 1994. We can then calculate the relative labor demand shifter for years 1995-2010 as in equation 52, by calculating the difference between the relative labor quantities observed in the data and the predicted ones using equation 51 and the relative wage in each period.

$$\Lambda_{t,US} = \frac{L_{t,US}}{G_{t,US}} - \hat{f}(\frac{w_{t,US}}{s_{t,US}}) \tag{52}$$

The calculated Λ_t represent the relative demand shifters for all US workers and picks up the occupational-biased technological change that occurred during the period which shifts labor demand towards a more intensive use of computer scientists. Innovation shocks that drive the tech-boom disproportionately increase the demand for CS workers, and this is captured by Λ_t.

[15] We experiment with higher order polynomials and our results do not change.

In the case of the US, we want to use the shifts in labor demand to trace out the relative supply curve of native computer scientists to non-CS college graduates. However, the Λ_t represent the shifts in the total relative labor demand which includes foreign computer scientists. Given our assumption that foreigners are marginally more productive than natives, firms hire them first making US computer scientists experience the shifts $Z_{t,k}$ of a residual labor demand such as in equation 53:

$$Z_{t,US} = \Lambda_{t,US} - \frac{L_{F,t,US}}{G_{t,US}} \tag{53}$$

We assume $Z_{t,US}$ follows an AR(1) process as in equation 46, so we can use the calculated $Z_{t,US}$ from the data and calculate the $\nu_{t,US}$ for every year as the unexpected shocks the native US college graduates face.

For India the process is almost identical. The key difference is that we just focus on the relative demand curve of CS workers that never migrate with respect to non-CS college graduates $\frac{L_{n,IN}}{G_{US}}$, taking the number of return migrant CS, $\bar{R}_{t,IN}$, as given. We calculate the predicted relative wages by the model for relative labor quantities between 0.001-0.03 to cover the range of values observed in the data. The process follows as it does in the US, by fitting a second order polynomial to get a closed form solution of the relative labor demand curve and calculate the aggregate demand shifts $\Lambda_{t,IN}$ in a similar fashion to equation 52. This gives us shocks to the relative demand curve of Indian firms. To get to the shocks that Indian workers experience, we add the possibility that Indian workers have of migrating into the US. This enters equation 54 as an additional shock to the relative labor demand experienced by workers.

$$Z_{t,IN} = \Lambda_{t,IN} + \frac{L_{F,t,US}}{G_{t,IN}} \tag{54}$$

Once again, assuming $Z_{t,IN}$ follows an AR(1) process and using the values of $Z_{t,IN}$ from the data we can get the shocks $\nu_{t,IN}$ that Indian college graduates face.

3.2 Calibrating Labor Supply

On the labor supply side of the model, we have eight parameters that need to be calibrated - $\{\sigma_{0,k}, \zeta_{0,k}, \sigma_{1,k}, \zeta_{1,k}, \chi_{0,k}, \chi_{1,k}, \chi_{2,k}, \beta\}$. Of these, we pick the annual discount rate to be $\beta = 0.9$, and calibrate the other parameters to match the data. In our model we assume the total quantities of non college graduates $\bar{H}_{t,k}$, native college graduates $\overline{(L_n + G)}_{t,k}$, foreign computer scientists $\bar{L}_{F,t}$ and return migrants R_t are determined outside the model.

In the way we set-up the model, changes in enrollment, employment and wages are driven by the exogenous technology shocks that shift out the demand curve for the different types of labor

over this decade. As the demand curve shifts, it traces out the labor supply curve for workers. The technological developments that drive these shifts in the labor demand are assumed to not affect the parameters of the workers' labor supply decisions.

We use data on relative wages, employment, enrollment and age shares to calibrate the remaining seven parameters. The first three series compare computer scientists to non-CS college graduate workers. For example, relative wages compare the wages for CS workers with wages for non-CS college graduates. To match data on wages, employment and age-shares in the US context, we use the March Current Population Survey (CPS). To match enrollment in CS degrees in the US, we use the Integrated Postsecondary Education Data System (IPEDS).

For India, to match wages, employment and age shares we use the largest and most comprehensive nationally representative labor force survey, called the National Sample Survey (NSS). For enrollment in engineering degrees we use yearly counts from the Ministry of Human Resources and Development. Details of the sample used in all these datasets and specific variable definitions can be found in Appendix A.[16]

We simultaneously match wages, employment and enrollment in three equally spaced years 1997, 2002 and 2007. We also match the share of computer science workers that are young (between 22 and 40) for the year 2007.[17] The series we use from the data are as follows:[18]

1. $\frac{L_{n,t,k}}{G_{t,k}} = \frac{\text{Computer scientists}}{\text{Non-CS college educated workers}}$ for $t = \{1997, 2002, 2007\}$
2. $\frac{w_{t,k}}{s_{t,k}} = \frac{\text{Median weekly wages for computer scientists}}{\text{Median weekly wages for non-CS college educated}}$ for $t = \{1997, 2002, 2007\}$
3. $\frac{q^{cs}_{t+2,k}}{q^{o}_{t+2,k}} = \frac{\text{Computer science/Engineering college degrees awarded (lagged 2 years)}}{\text{non-CS college degrees awarded (lagged 2 years)}}$ for $t = \{1997, 2002, 2007\}$
4. $age^{22,40}_{t,k} = \frac{\text{Computer scientists with age between 22 and 40}}{\text{CS}^{22,40} + \text{CS}^{41,65}}$ for $t = \{2007\}$

To simultaneously find parameter values which solve the model under these data restrictions, we use a Nelder-Mead simplex method. While the system uses all the data at the same time, there is strong intuition behind the identification of each parameter. For example, the relative enrollment data should help identify the taste parameters for field of major decisions ($\sigma_{0,k}$ and $\zeta_{0,k}$), whereas the relative employment data should help pin down the occupation specific tastes

[16] For the CPS, We exclude imputed wages, and multiply top-coded values by 1.4. Bollinger and Hirsch (2007) show that including imputations can lead to biased results. Whereas the top-coding adjustment is standard in the literature (Lemieux, 2006). For both the NSS and the CPS, we smooth the raw data over three-year moving averages as follows: $X_{t,smooth} = \frac{1}{3}(X_{t-1,raw} + X_{t,raw} + X_{t+1,raw})$

[17] Given that in our labor supply model we impose all cohorts are the same size, we normalize the number of computer scientists of a given age group dividing by the total number of college graduates in that age group before calculating the age shares.

[18] We have an exactly identified system as we use ten data moments to recover ten parameters - $\{\sigma_{0,k}, \zeta_{0,k}, \sigma_{1,k}, \zeta_{1,k}, \chi_{0,k}, \chi_{1,k}, \chi_{2,k}\}$ and three implied values of technology in the years we match the wage/employment data $\{Z_{97}, Z_{02}, Z_{07}\}$

($\sigma_{1,k}$ and $\zeta_{1,k}$). The age shares in computer science employment together with enrollment and employment help identify the occupation switching cost parameters that depend on age ($\chi_{0,k}$, $\chi_{1,k}$ and $\chi_{2,k}$).

3.2.1 Labor Supply Calibration Results

Figure 3 shows the data used and the model fit from this exercise. The figures report both the path of the variables of interest predicted by the model, and the data we use for these series. We match three equally spaced years (1997, 2002 and 2007) for employment, wages and enrollment, and the remaining years plotted are an out of sample test of our method. The years in between include years where there were observed changes to immigration laws, and other potentially structural changes that may make it difficult for the data to fit perfectly.

The employment series in Figures 3a and 3b fit well at the start and end of the period, but it misses some years in between. In India, the wage series fits well, particularly towards the end and the start of the series (Figure 3d), whereas in the US we do a better job of matching the years in between (Figure 3c). Lastly, the enrollment series can be seen in Figures 3e and 3f. In India, we match the rapid increase in enrollment relatively well, whereas in the US we match the enrollment rise only during the 1990s – while our model predicts a secular decline in the 2000s, it is at a slower rate than the slump in the data.

Table 6 presents the values of the calibrated parameters for India and the US. On average, we can see that in the US there is a mean taste for not working or studying in CS occupations, but this is the other way around in India where there is a greater dispersion in occupational tastes. In both countries, however, the sector switching costs are convex with age.

These calibrated parameters allow us to trace out the labor supply curve for computer-scientists relative to non-CS college educated workers. In order to do this, we use the model set-up and the parameters, and vary the relative wage to measure the response in relative quantities of labor. This derives the relative supply curve which we then use in the labor market to find the equilibrium wage.[19]

4 Endogenous Variables and Model Fit

In this section we study the evolution of the endogenous variables in our model over time, and evaluate how well it matches our data. In order to evaluate the fit of our model we compare our simulated results with features from the data as out of sample tests. In the calibration exercise we explicitly match certain data points or trends, whereas here we discuss how well

[19] Our estimated relative labor-supply elasticities lie between 4.3 and 14.5.

our model matches the data on items we do not explicitly calibrate. Figures 4-5 show that we match fairly well some of the key aspects that we are trying to capture.

As can be seen in Figure 4, relative wages in the US and India for our three type of workers closely match the data. In Figure 4a we plot the college skill premium within the US, where we construct the graduate wage to be an employment weighted average of the CS and non-CS college graduate wages. We do a similar exercise in India with Figure 4b where our weighted average also includes the separate earnings of return migrants, and once again do a good job of matching the college skill premium.

In Figures 4c-4e we study the relative wages for each occupation but across countries. In the initial step of the product market calibration we backed out the relative productivity in the IT sector to match relative wages of computer scientists between the two countries. Figure 4c shows that after running the specification of our full model, we still match the level differences and the trends in relative wages which show a closing of the gap between India and the US. These numbers are also very close to quasi-experimental results that show a 5.57-fold increase in wages for 2007 lottery winners and a 6.31-fold increase in wages for 2008 lottery winners (Clemens, 2013). While we never explicitly match the wages for non-CS college graduates and non college graduates, we can see in Figures 4d and 4e that the model does fairly well in predicting the trends and level differences between the wages in the US and India for these type of workers.

Figure 4f shows that the model does a good job predicting the levels and the trend in the share of IT output of the US in total IT production. As we can see, over time India captures more of the IT world market share, although we do predict India to catch up at a faster rate than what we observe in the data. The overall trend in our model and the data, however, indicates that India erodes the US comparative advantage in IT production and becomes a major played in the export market.

We calibrated the labor supply curve such that we predict the relative number of computer scientists to non-CS college graduates in each country for given wages. Figures 5a and 5b show that our final model does a good job in matching the equilibrium quantities of computer scientists to non-CS college graduates observed in the data. We also match the levels and trends in the total number of CS workers employed in both countries, which show a steady increase over this period (Figures 5c-5d).

5 Counterfactual Exercises

In order to evaluate the impact of the H-1B program on US and Indian economy we conduct a counterfactual exercise where we prohibit Indian workers from entering the US on an H-1B visa.

Those workers who would have gone to the US are forced to work in India. This restriction is expected and therefore had already influenced the enrollment and occupational decisions of workers in India. In other words, in the restricted H-1B regime, we are in a world where students, firms and workers in both India and the US would make decisions knowing that there would be no foreign CS workers from India.

Using the calibrated parameters and the given set up we can trace out what happens to all endogenous variables between 1994 and 2010. We describe the regime where Indians are allowed to enter the US on H-1B visas as "with immigration," and the counterfactual regime as "without immigration." In Figures 6 and 7 we presents the results for this exercise, where we plot the percentage change in the endogenous variable when going from the without immigration to the with immigration regime. That is, for variable X in year t we plot $\frac{X_{with\ immigration,t}-X_{without,t}}{X_{without,t}}$.

5.1 Employment and Wages

In Figure 6 we see how the labor market is affected when moving from the restricted immigration regime to the one with the H-1B program. Figure 6a looks at the size of the CS workforce in India. The opening up of immigration possibilities to the US increased the size of the CS workforce in India by as much 50% in 1994. While this is a large increase, it is important to keep in mind that the base is small in 1994, and over time in the long-run, by 2010 the increase in the total CS workforce is 11.2%. However, this includes those workers who had migrated pre-1994 and have now returned from the US. If we restrict it to non-migrants, then the increase in the Indian CS workforce due to the H-1B program in 2010 is 4.86%. The US CS wage in the data and our model is about 9 times that of the India CS wage in 1994 (Figure 4c). The prospect of, therefore, migrating to the US and earning such a high wage leads students to enroll in CS degrees and workers to switch into CS occupations in India.

In the US, on the other hand, Figure 6b shows that native employment in CS is lower in a world where they allow Indian CS workers to enter on H-1Bs. In the early years, in 1994, native employment is lower by about 1%, but this steady declines to about 7% in 2010. Total employment in CS is, however, higher under the H-1B regime by about 0.65% in 2010. When Indian CS workers enter the US on H-1B visas, this tends to lower the relative CS to non-CS wage hurting close substitutes like native CS workers and potentially benefiting complements like native non-CS college graduates. This tends to encourage native CS workers to switch into non-CS college graduate occupations, increasing the employment in this group by 0.4%.

These employment shifts are accompanied with changes in wages for each of these groups. In India, Figure 6c shows how the drastic increase in the CS workforce leads to an initial fall in the CS wage by as much as 9.9% in 1994. However, a larger CS workforce also leads to more productivity in the IT sector, raising wages for all workers and especially CS workers over time.

By the end of the period, in 2010, CS wages in India are actually higher by 3% due to these productivity increases.

Wages for other workers in the Indian economy will be higher when there are more CS workers because of two reasons. First, other workers are complements in production, but more importantly, CS workers are innovators and raise productivity in the IT sector. Many of the H-1B workers are trained in the US and acquire technology that they bring back to India with them, also raising productivity in the Indian IT sector. Since the IT sector output is an intermediate good into the final sector, this raises productivity in the entire economy, raising all wages. The wages for non-CS workers in the Indian economy are higher by between 9.5% and 9.7% in 2010 under the H-1B regime.

In the US, on the other hand, Figure 6d shows how there are very mild negative effects on real wages for all types of workers. This not only captures the effect of labor-market crowd out, but also the fact that output prices in the open economy may be higher. Furthermore, under the H-1B regime, the IT sector grows in India and production shifts away from the US negatively impacting all types of workers. Even though CS workers are the worst affected, their wages only fall at most by 0.5% in 2010. As workers switch into non-CS work, their earnings dip by at most 0.36% in 2010. The least worse off are the non graduates as their wages in 2010 are lower by only 0.1%.

5.2 Income, and IT sector output

In Figure 7 we look at how the IT sector and total income evolves in the US and India when moving from the H-1B regime to the restricted immigration one. The H-1B regime incentivizes students and workers in India to switch to CS occupations, growing the IT sector in India (Figure 7a). In 1994, the quantity of IT sector output is higher by as much as 5%, and at the end of this period it is still about 1.8% higher under the H-1B regime. Employment in the Indian IT sector is also higher by about 4.77% in 1994, but steadily falls, and by 2010 the IT employment is actually lower by 0.79% under the H-1B regime. Since we model the IT sector output as an intermediate input into final goods production, we are introducing a degree of complementarity in production. So an increasing CS workforce, and therefore increasing IT productivity, may have ambiguous effects on total IT sector employment.

It is no surprise that total income and total IT output in both countries combined is higher under the H-1B regime (Figure 7b). The gains from immigration are large in this context, with the combined income of the US and India being higher by about 1.5%. Total IT output rises steadily under the H-1B regime to about 1.63% in 2010.

As IT output rises, production shifts to India (Figure 7c). In the data, we see that over time India takes over as the major exporter of IT. Under the H-1B regime, due to the large wage

premium in the US, Indian students and workers switch to CS occupations and degrees raising the size of the CS workforce. This increasing CS workforce increases productivity in the Indian IT sector. Furthermore, those who return from the US bring with them technical knowhow also increasing productivity and growing the Indian IT sector. Under the H-1B regime, therefore, in 1994 the share of world IT output and the value of Indian IT output is higher by about 17%, and by 2010 this increase is still at least 8.9%.

The shift in production to India, however, hurts the US IT sector. Since the IT sector output is an important input into the final goods sector, it mildly lowers income in the US (Figure 7d). The relative rise in the Indian CS workforce also lowers the relative productivity in the US IT sector. By the end of this period, in 2010, US IT sector output is lower by 0.24% and total income is lower by 0.04%.

5.3 Disentangling the Effects of Return Migration and the Incentives to Work as CS

In our model there are two main mechanisms through which we explain the increase in computer scientists in India due to US immigration policy. The first mechanism is that the large wage differential for CS between US and India causes more Indians to join the CS labor force with the expectation of migrating to the US. A second mechanism is that a share of those Indians who migrate return to India after their visa expires with enhanced skills that allow them to work as a more specialized type of CS. In Figure 8 we show the with and without immigration differences in wages for our full model, where these two mechanisms are in place and a restricted model where we shut down the possibility of return-migration to India once they get to work in the US. This would be analogous to a policy that grants a green card to each foreigner that comes to work in the US.

We can see that while the US-India wage differential seems to be the main driver of our results, the absence of return migration does mitigate the changes observed between a situation with and without immigration. In the US, we see that shutting down return migration to India reduces the losses for all three type of workers. In fact, non college graduates benefit under immigration in the later period as the complementarity effect outweighs the negative relative productivity effect. In India we see that the wage increase is lower for non-CS college graduates and non college graduates when return migration is not allowed. The lower number of CS in India, caused by shutting down return migration, decreases relative productivity, making gains for complements smaller. For native CS that never migrate shutting down return migration creates losses for this group, as the competition effect of extra CS workers is now larger than the increases in productivity due to the larger CS workforce.

5.4 Distributional Gains and Losses

In Table 7 we see that world income per capita increases when immigration is allowed but that the income changes are different for different types of workers. As expected, those who gain the most from the H-1B program are Indian CS workers who get to migrate to the US – their per capita income increases by $31,838 in 2010. Indian CS who were already in the US lose from allowing immigration as an increase in close substitutes drives their wages down.

Those who migrate to the US and return to India after 1994 also benefit significantly from immigration, as they get the chance to work in India as a more specialized type of labor than regular CS, increasing their income by $10,925 in 2010. On the contrary, those who already returned from the US before 1994 lose when immigration is allowed since they lose their monopoly over the specialized occupation they are in as more Indians return as specialized CS workers from the US.

Non college graduates in India gain from positive spillovers of the IT boom given that they are complements in production and benefit from the increase in CS workers working in India as well as the increase in relative productivity in the IT sector. The same thing happens to non-CS college graduates, but to a greater extent, since they also benefit from more college workers deciding to become CS – this reduces the competition for non-CS college graduate jobs and drives up their wages.

We split CS workers that never leave for the US into two groups: those who would have been CS regardless of immigration policy and those who would not have been CS if immigration is banned, but do choose CS when immigration is allowed. The first group initially loses from immigration since there are more native workers who get into CS with the prospects of going to the US and drive CS wages down. This mechanism is offset by the increase in productivity generated by the larger size of the CS workforce under immigration, which compensates for any negative effect of immigration. These workers are better off by $231 per capita in 2010. Those that switch to CS when immigration is allowed also gain from immigration as they access better jobs when they switch occupations.

In the US we see that all types of workers lose from immigration, even though these losses are small compared to the gains experienced by India. Non college graduates lose around $19 per capita due to the decrease in relative productivity of the US with respect to India, while non-CS college graduates lose around 100 dollars per capita given that in addition to the productivity effect they face more competition from CS workers that switch into non-CS college graduate occupations when immigration is allowed. Those that lose the most are native CS who, in addition to the productivity effect, face direct competition from the Indian CS who migrate to the US, driving down their wages. Those native CS that switch occupations to non-CS college graduate occupations when immigration is allowed lose less than those who stay in CS given

that they mitigate the competition effect from foreigners by switching occupations.

6 Discussion

India experienced a dramatic expansion in IT employment and output in the 1990s and early 2000s. Many factors contributed to this boom but, our work suggests that, surprisingly, policies from halfway around the world played a critical role. We study how US immigration policy, combined with high wages and technical expertise in the US, helped enable the IT boom in India. To do this, we describe the IT boom in the US and India with the help of a general equilibrium model. In our model, the prospect of high wages in the US incentivized students and workers in India to choose CS degrees and occupations. Those returning from the US after the expiration of their H-1Bs also contributed to the growing Indian workforce. These movements increased overall IT sector productivity in India and shifted the production of IT goods away from the US.

In this paper, we are explicitly testing the explanatory power of only certain particular mechanisms by which US policy and conditions may have stimulated growth in the Indian IT sector. We do this by specifically focusing on four features of the US in this period that created important incentives and constraints for Indian students and workers. First, technological innovations and changing consumer preferences generated strong demand for IT workers in the US. Second, and not unrelated, the wage differential between the US and India was large, especially for IT workers. Third, US immigration policy, as embodied by the H-1B visa program, strongly favored skilled migrants. Finally, H-1B visas only last 3-6 years, obligating many migrant workers to return to India with accumulated human capital and technical knowhow. Together, these features help spread the boom across the world from the US to India.

The H-1B program has significant distributional consequences, where certain workers are adversely affected while others benefit. However, as we find in this paper, overall gains outweigh the losses as the combined incomes of the US and India rise under the H-1B program by about 1.5% or about $40 bill. This net gain is consistent with a long literature reviewed in Clemens (2011).[20]

However, one somewhat striking result is that as production shifts to India, US income actually falls. A driving feature of this result is that an increase in the size of the Indian CS workforce increases the relative productivity of India's IT sector. Such reductions in US income have already been discussed by a rich literature on the economics of trade and migration. Krugman (1979) describes a North-South general equilibrium trade model where the North initially has

[20]Specifically, see Klein and Ventura (2007); Moses and Letnes (2004); van der Mensbrugghe and Roland-Holst (2009); Walmsley and Winters (2005). Iregui (2005) shows that movement of skilled labor can increase world GDP by between 6-11% of GDP.

a monopoly over new products given its technological superiority and rate of innovation.[21] The South catches due to technological diffusion and over time starts exporting to the North the very same products the North used to export. As the rate of technological diffusion increases, or the rate of innovation in the North declines, living standards will actually fall in the North. With quality differentiation in products, Flam and Helpman (1987) generate richer trade dynamics, but also show that technical progress in the South brings about a decline in the North's wage rate. Therefore, as Samuelson (2004) notes, such technical progress in the South erodes the US comparative advantage and can permanently lower per capita incomes in the US.

The labor economics literature has also emphasized these channels. For instance, Johnson and Stafford (1993) show how the effect of foreign competition from abroad lowers aggregate real incomes in the US. In fact, Freeman (2006a) focuses on the global job market for high-tech workers and argues that the growth in such labor abroad adversely affects US industry and workers. In this analysis, immigration can help maintain the US's lead by attracting overseas talent. However, the analysis does not account for the effect of immigration on incentives to invest in India and the role of return migration, which we show to be important determinants of the shift in production abroad.

Our results, therefore, quantitatively confirm many of the theoretical results in the literature. Davis and Weinstein (2002) show how in a Ricardian trade framework, such as ours, a country that experiences immigration due to technological superiority always loses from such migration through a deterioration in the terms of trade. Free mobility will tend to equalize wages across countries and, therefore, hurt workers in the country with superior technology. They estimate that in 1998, the losses to US natives alone were about 0.8% of GDP, and about 0.88% of GDP for the economy as a whole including immigrants. While our numbers are considerably smaller – 0.16% of GDP in 1998, and 0.04% of GDP by 2010 – we are only focusing on high-skill immigration from India.

While consistent with a lot of the literature, our result does heavily rest on one crucial assumption – relative technology depends on the relative size of the computer science workforce in each country. If this were not true, then US workers would be less adversely affected and Indian workers would benefit less from the H-1B program. As our focus is on certain mechanisms, there may be other features of these economies that drive the boom in tandem. One clear shortcoming of our exercise is that we maintain that the price elasticity of demand would have been the same in the absence of the H-1B program. One may consider alternative models of consumer demand that include network externalities or learning-by-using inherent in IT goods that may prevent the fall in IT production in the US. We leave such an analysis for future research.

[21] This is a formalization of Vernon (1966).

References

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics*.

Alvarez, F. E., F. J. Buera, and R. E. Lucas (2013). Idea Flows, Economic Growth, and Trade. *NBER Working Paper Series, No. 19667*.

Arora, A. and S. Athreye (2002). The Software Industry and Indias Economic Development. *Information Economics and Policy, Vol. 14, No. 2, pp. 25373. Social Science Research Network, Rochester.*.

Arora, A., A. Gambardella, and S. Torrisi (2001). In the Footsteps of the Silicon Valley? Indian and Irish software in the international division of labour. *Indian Institute of Management, Ahmedabad*.

Athreye, S. S. (2005). The Indian Software Industry and its Evolving Service Capability. *Industrial and Corporate Change Advance Access*.

Banerjee, A. and E. Duflo (2000). Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry. *The Quarterly Journal of Economics*.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003, September). Plants and Productivity in International Trade. *The American Economic Review 93*(4), 1268–1290.

Bhatnagar, S. (2005). Indias Software Industry. *Indian Institute of Management Ahmedabad*.

Bollinger, C. and B. Hirsch (2007). How Well are Earnings Measured in the Current Population Survey? Bias from Nonresponse and Proxy respondents. In *North American Summer Meetings of the Econometric Society, Duke University*.

Borjas, G. J. (1999). The Economic Analysis of Immigration. *Handbook of Labor Economics*.

Bound, J., B. Braga, J. Golden, and G. Khanna (2015). Recruitment of Foreigners in the Market for Computer Scientists in the US. *Journal of Labor Economics*.

Bound, J., M. Demirci, G. Khanna, and S. Turner (2014). Finishing Degrees and Finding Jobs: U.S. Higher Education and the Flow of Foreign IT Workers. *Innovation Policy and the Economy, Volume 15*.

Bound, J., G. Khanna, and N. Morales (2016). Understanding the Economic Impact of the H-1B Program on the US. *Working Paper*.

Card, D. and T. Lemieux (2001, May). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics 116*(2), 705–746.

Clemens, M. A. (2011, Summer). Economics and Emigration: Trillion-Dollar Bills on the Sidewalk. *Journal of Economic Perspectives 25*(3), 83–106.

Clemens, M. A. (2013). Why Do Programmers Earn More in Houston than Hyderabad? Evidence from Randomized Processing of U.S. Visas. *American Economic Review Papers & Proceedings 103:3*.

Dasgupta, K. (2012, July). Learning and Knowledge Diffusion in a Global Economy. *Journal of International Economics 87*(2), 323–336.

Dataquest (1996, 31 July). pp. 43–44.

Dataquest (2003). The Made-in-India Brigade. *Vol. 20 No. 15, Cyber Media (India) Limited, Gurgaon*.

Davis, D. R. and D. E. Weinstein (2002, May). Technological Superiority and the Losses from Migration. *NBER WP No. 8971*.

Desai, A. (2003). The dynamics of the Indian information technology industry. *DRC Working Paper No. 20, Centre for New and Emerging Markets, London Business School*.

DOC (2000). Digital economy 2000. *U.S. Department of Commerce.*.

Dornbusch, R., S. Fischer, and P. Samuelson (1977, December). Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review 67*(5), 823–39.

Eaton, J. and S. Kortum (2002). Technology, Geography and Trade. *Econometrica 70*, 1741–1779.

Flam, H. and E. Helpman (1987, December). Vertical Product Differentiation and North-South Trade. *The American Economic Review 77*(5), 810–822.

Foley, F. and W. Kerr (2013). Ethnic Innovation and US Multinational Firm Activity. *Management Science 59:7*.

Freeman, R. (2006a). Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership? *Innovation Policy and the Economy Volume 6, Adam B. Jaffe, Josh Lerner, and Scott Stern*.

Freeman, R. (2006b). People Flows in Globalization. *Journal of Economic Perspectives*.

Freeman, R. (2009). What does global expansion of higher education mean for the us? *Working Paper 14962, National Bureau of Economic Research.*.

GAO (1992). Immigration and the Labor Market: Nonimmigrant Alien Workers in the United States. *U.S. General Accounting Office, Washington, D.C.*.

Goldin, C. and L. F. Katz (2007). Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing. *Brookings Papers on Economic Activity* (2), 135–65.

Heeks, R. (1995). From Regulation to Promotion: The States Changing but Continuing Role in Software Production and Export. *Development Policy and Practice Working Paper, No 30, Development Policy and Practice Research Group, Faculty of Technology, The Open University, UK*.

Hunt, J. and M. Gauthier-Loiselle (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics 2:2*.

Iregui, A. M. (2005). Efficiancy Gains from the Elimination of Global Restrictions on Labour Mobility. *Poverty, International Migration and Asylum*, 211–258. ed. George J. Borjas and Jeff Crisp. New York: Palgrave Macmillan.

Johnson, G. E. and F. P. Stafford (1993, May). International Competition and Real Wages. *The American Economic Review 83*(2), 127–130.

Kambourov, G. and I. Manovskii (2009, 02). Occupational Specificity Of Human Capital. *International Economic Review 50*(1), 63–115.

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

Kerr, W. (August 2013a). U.S. High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence. *Harvard Business School Working Paper, No. 14-017*.

Kerr, W. and W. Lincoln (2010). The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention. *Journal of Labor Economics*.

Kerr, W. R. (2013b). Heterogeneous Technology Diffusion and Ricardian Trade Patterns. *NBER Working Paper Series, No. 19657*.

Kirkegaard, J. (2005). Outsourcing and Skill Imports: Foreign High-Skilled Workers on H-1B and L-1 visas in the United States. *Working Paper no. 05-15, Peterson Institute for International Economics,Washington, DC.*.

Klein, P. and G. Ventura (2007). TFP Differences and the Aggregate Effects of Labor Mobility in the Long Run. *The B.E. Journal of Macroeconomics 7*(1).

Krugman, P. (1979). A Model of Innovation, Technology Transfer, and the World Distribution of Income. *Journal of Political Economy Vol. 87, No. 2*, 253–266.

Kumar, N. (2001). Indian Software Industry Development: International and national perspective. *Economic and Political Weekly, Sameeksha Trust Publication, Mumbai*.

Kumar, N. (2006). Emerging Multinationals : Trends, Patterns and Determinants of Outward Investment by Indian Enterprises. *Microeconomics Working Papers 22108, East Asian Bureau of Economic Research.*.

Lee, H. (2016). Quantitative Impact of Reducing Barriers to Skilled Labor Immigration: The Case of the US H-1B Visa. *Working Paper*.

Leiner, B. M., V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. S. Wolff (1997). The Past and Future History of the Internet. *Commun. ACM 40 (2), 102-108*.

Lemieux, T. (2006). Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *The American Economic Review 96*, 461–498.

Matloff, N. (2003). On the Need for Reform of the H-1B Non-Immigrant Work Visa in Computer-Related Occupations. *University of Michigan Journal of Law Reform 36:4*.

Moses, J. W. and B. Letnes (2004). The Economic Costs to International Labor Restrictions: Revisiting the Empirical Discussion. *World Development 32*(10), 1609–1626.

NASSCOM (2003). The IT Industry in India - Strategic Review. *New Delhi*.

NASSCOM (2012). The IT Industry in India - Strategic Review. *New Delhi*.

Peri, G. and C. Sparber (2011). Highly-Educated Immigrants and Native Occupational Choice. *Industrial Relations 50:3*.

Perla, J., C. Tonetti, and M. E. Waugh (2015). Equilibrium Technology Diffusion, Trade, and Growth. *Working Paper*.

Rosenzweig, M. (2006). Global Wage Differences and International Student Flows. *Brookings Trade Forum*.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica 55*(5), pp. 999–1033.

Ryoo, J. and S. Rosen (2004, February). The Engineering Labor Market. *Journal of Political Economy 112*(S1), S110–S140.

Samuelson, P. A. (2004, Summer). Where Ricardo and Mill Rebut and Confirm Arguments of Mainstream Economists Supporting Globalization. *Journal of Economic Perspectives 18*(3), 135–146.

Saxenian, A. (1999). *Silicon Valley's New Immigrant Entrepreneurs*. San Francisco CA: Public Policy Institute of California.

Somale, M. (2014). Comparative Advantage in Innovation and Production. *Working Paper*.

Subramanian, C. (1992). India and the Computer: A Study of Planned Development. *Oxford University Press: New Delhi.*.

USCIS (2014). Characteristics of H-1B Speciality Occupation Workers. *US Department of Homeland Security*.

USINS (2000). Characteristics of Specialty Occupation Workers (H1B). *U.S. Immigration and Naturalization Service, Washington, D.C.*.

van der Mensbrugghe, D. and D. Roland-Holst (2009). Global Economic Prospects for Increasing Developing-Country Migration into Developed Countries. *United National Development Programme Human Development Research Paper 50*.

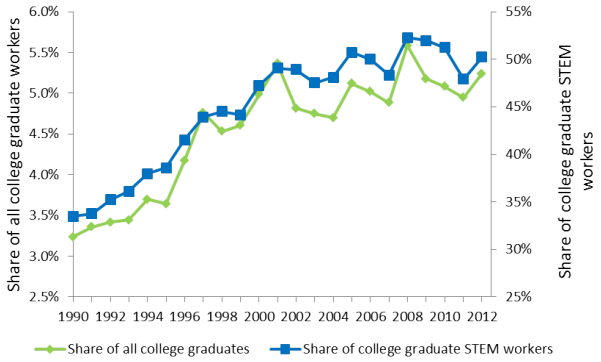
Vernon, R. (1966, May). International Investment and International Trade in the Product Cycle. *Quarterly Journal of Economics 80*, 190–207.

Walmsley, T. L. and A. L. Winters (2005). Relaxing the Restrictions on the Temporary Movement of Natural Persons: A Simulation Analysis. *Journal of Economic Integration 20*(4), 688–726.

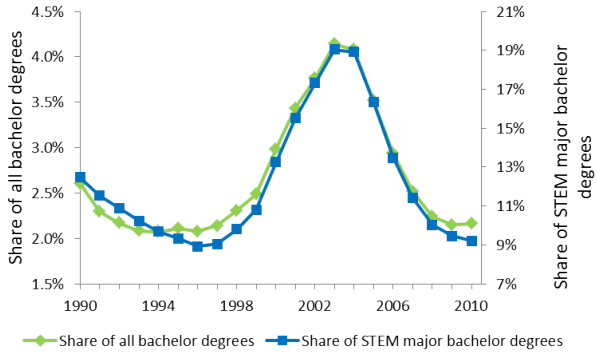
7 Tables and Figures

Figure 1: Descriptive: High-Skill Immigration and the IT Boom

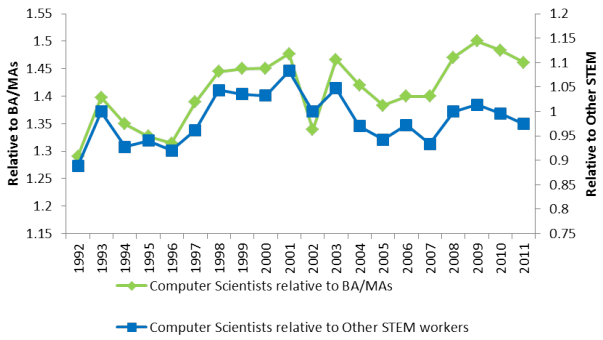
(a) Fraction of Computer Scientists in US Workforce



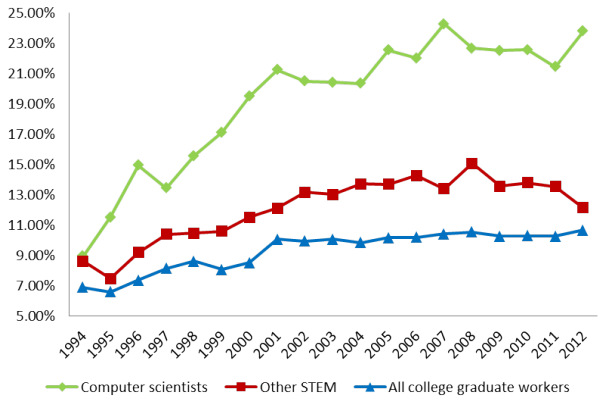
(b) Computer Science Fraction of Bachelor Degrees in US



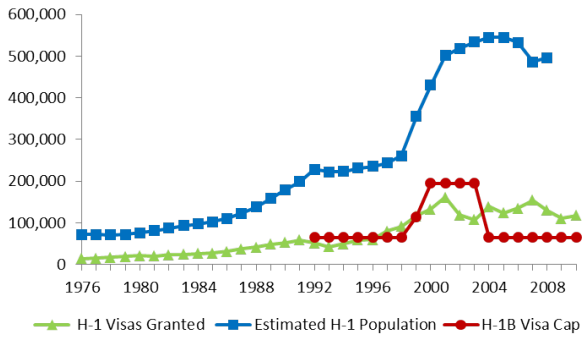
(c) Earnings of Computer Scientists Relative to Other groups



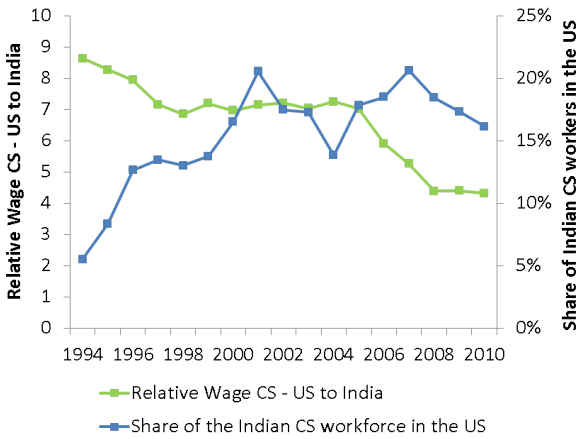
(d) Immigrants as Fraction of Workers by Occupation



(e) H-1 Visas



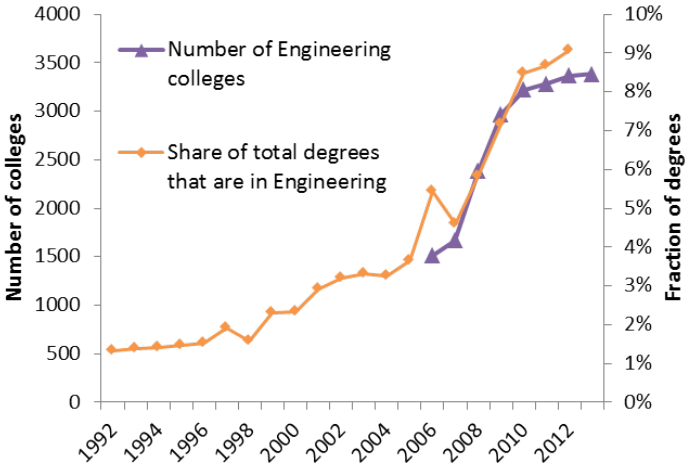
(f) CS Wage Differential and Share of Indian CS in US



Sources: Figure 1a, 1c and 1d March Current Population Survey (CPS). Figure 1b is from IPEDS (The Integrated Postsecondary Education Data System). Figure 1e author's calculations updating Lowell (2000). Figure 1f are based on author's calculations using CPS and the National Sample Survey (NSS) of India.

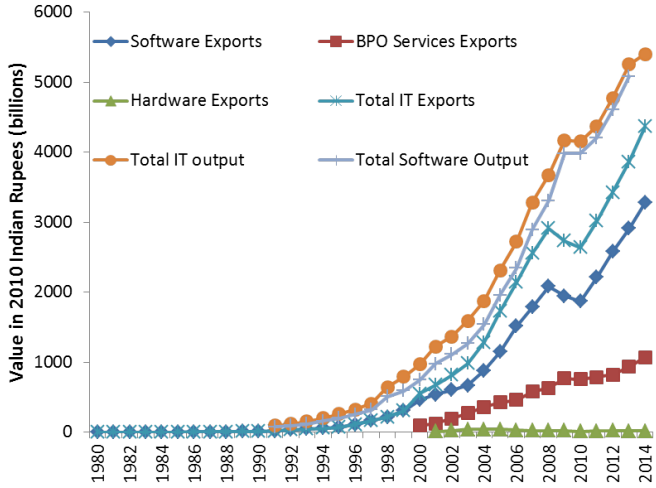
Figure 2: Descriptive: High-Skill Immigration and the IT Boom

(a) India: Growth in Engineering in the Education Sector



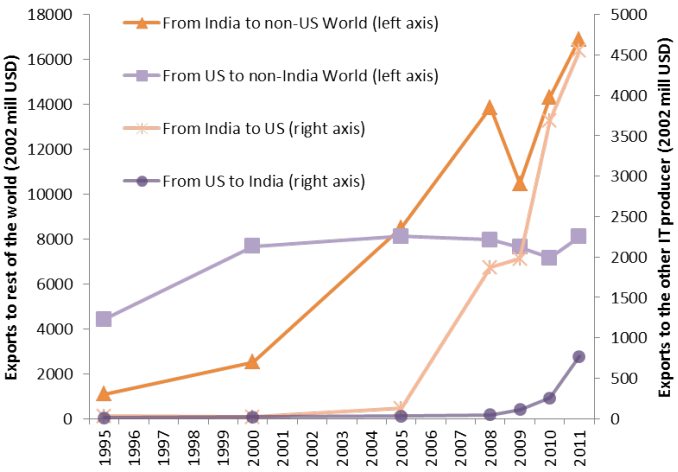
Source: Degrees come from a combination of sources – Ministry of Human Resources and Development, the National Association of Software and Service Companies (NASSCOM) and the All India Council for Technical Education (AICTE); missing years are interpolated. Number of universities is from Ministry of Human Resources and Development.

(b) Growth in the Indian IT sector



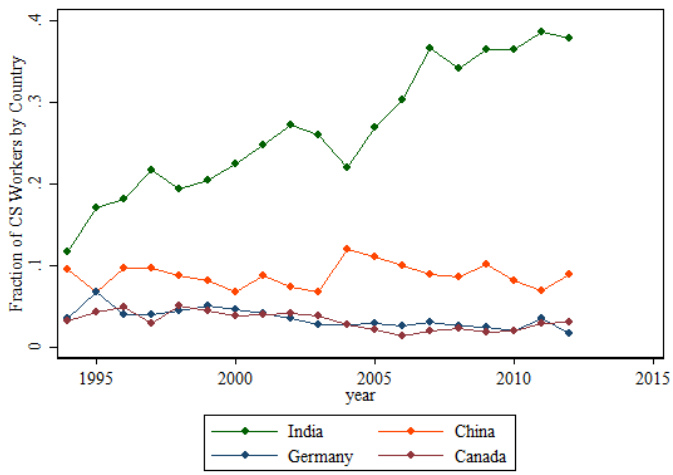
Source: Electronic and Information Technology Annual Reports, Indian Department of Electronics reports, National Association of Software and Service Companies (NASSCOM). Much of this data has been collated and standardized by the Center for Development Informatics at the University of Manchester, UK.

(c) IT Exports Over Time



Source: OECD Trade in Value Added statistics for the "C72: Computer and Related Activities" industry. The data is only available for 1995, 2000, 2005 and 2008-2011.

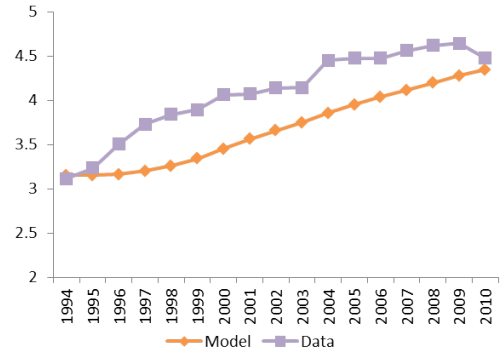
(d) Fraction of US Computer Scientists by Country



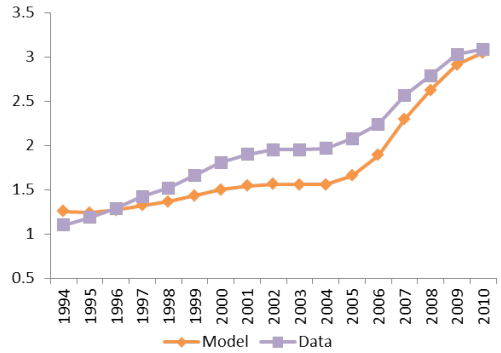
Source: Current Population Survey – Outgoing Rotation Group Survey. Fraction of US CS workers by Country of Birth (variable 'penatvty' in CPS-ORG).

Figure 3: Calibration: Labor Supply Parameters

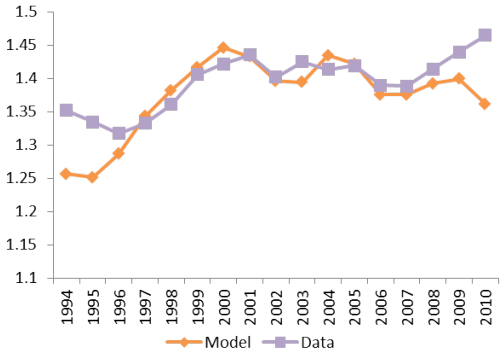
(a) Relative Labor Supply – US



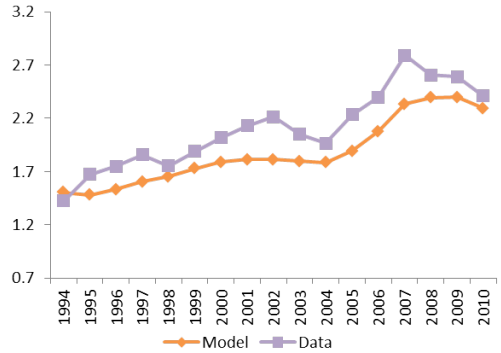
(b) Relative Labor Supply – India



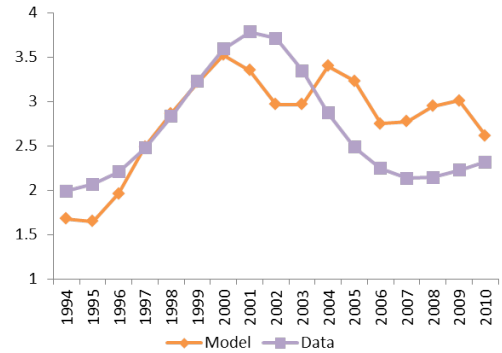
(c) Relative Wages – US



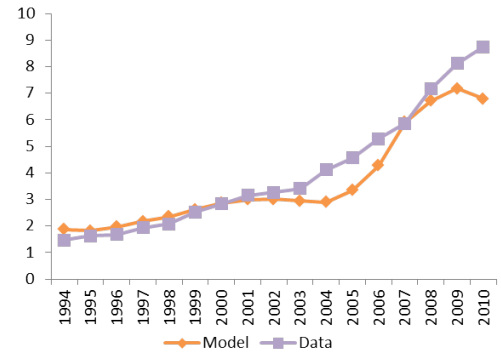
(d) Relative Wages – India



(e) Relative Enrollment – US



(f) Relative Enrollment – India



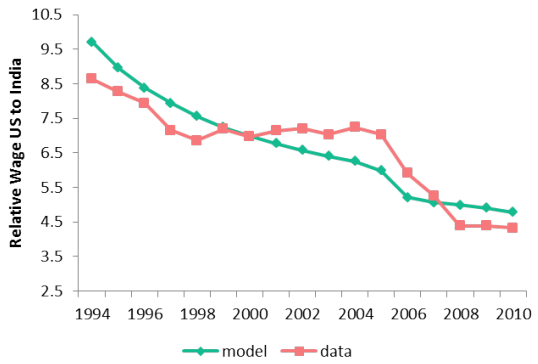
In the calibration exercise, the years 1997, 2002 and 2007 were used to match the data for employment, wages and enrollment. The years in between are an out-of-sample test. In the US, wage and employment data come from the March CPS, whereas enrollment data is from IPEDS. In India, wage and employment data come from the NSS, whereas enrollment data is from HRD Ministry. See Appendix A for more details.

Figure 4: Model Fit: Relative Wages and Production – US and India

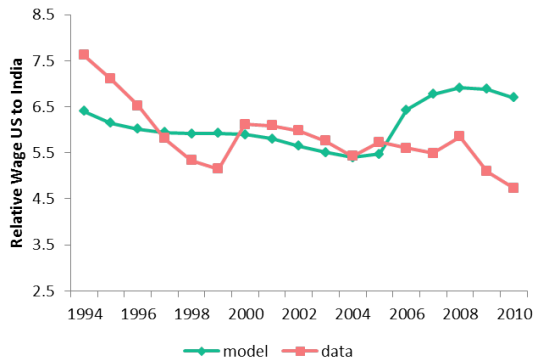
(a) Relative Wage within the US: College Graduates to non College Graduates

(b) Relative Wage within India: College Graduates to non College Graduates

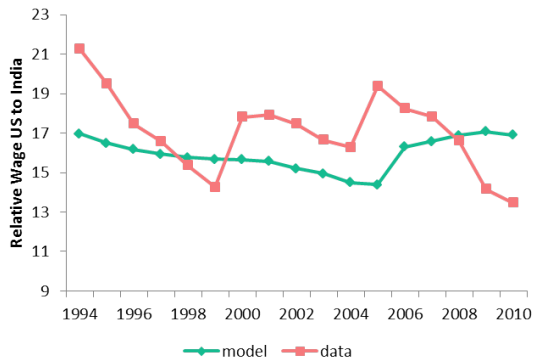
(c) Relative Wage for Computer Scientists: US to India



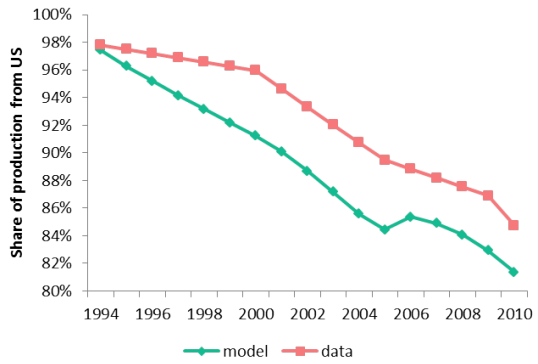
(d) Relative Wage for non-CS College Graduates: US to India



(e) Relative Wage for non College Graduates: US to India



(f) Share of US production in IT

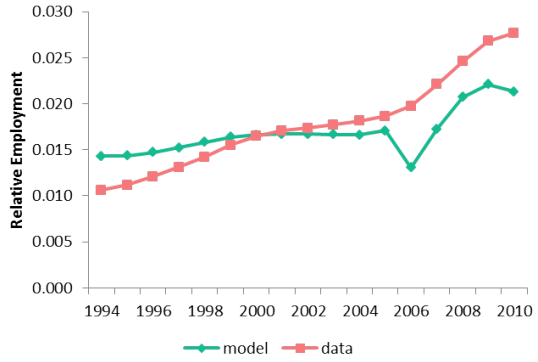


Figures plot the simulated model output and the actual data for the endogenous variables of interest. For data sources please refer to Data Appendix A

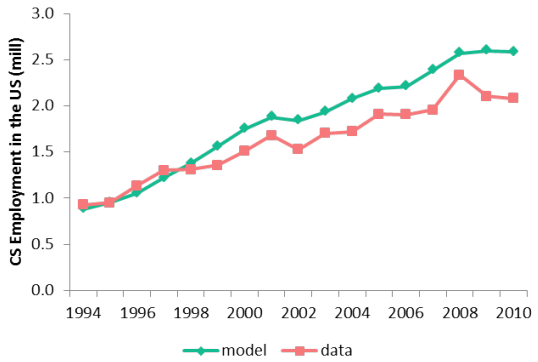
Figure 5: Model Fit: Employment Patterns and IT production by Country

(a) Computer Scientists to non-CS College Graduates – US

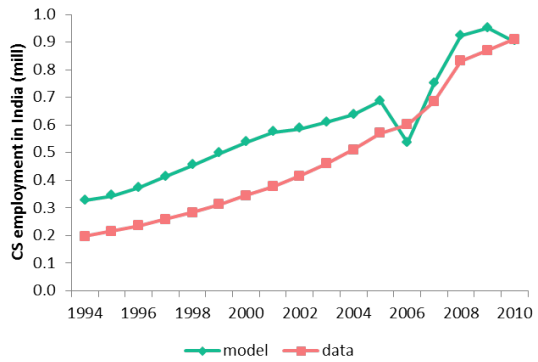
(b) Computer Scientists to non-CS College Graduates – India



(c) Number of Computer Scientists in the US



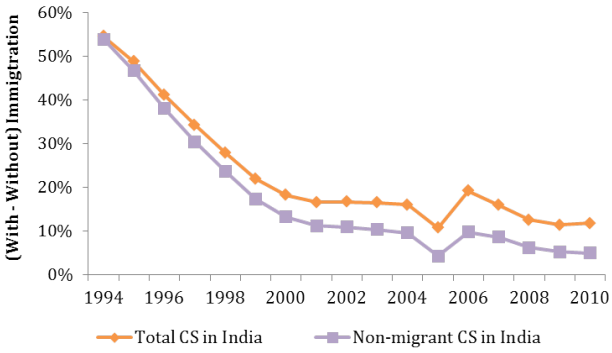
(d) Number of Computer Scientists in India



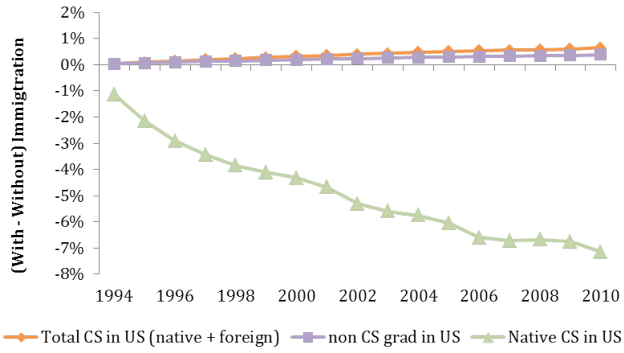
Figures plot the simulated model output and the actual data for the endogenous variables of interest. For data sources please refer to Data Appendix A

Figure 6: Counterfactuals: Employment and Wages by Country

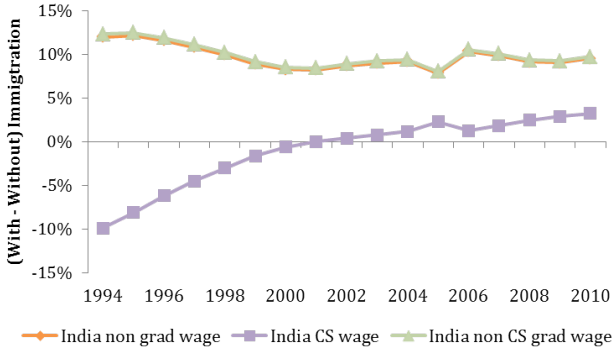
(a) Computer Scientists in India



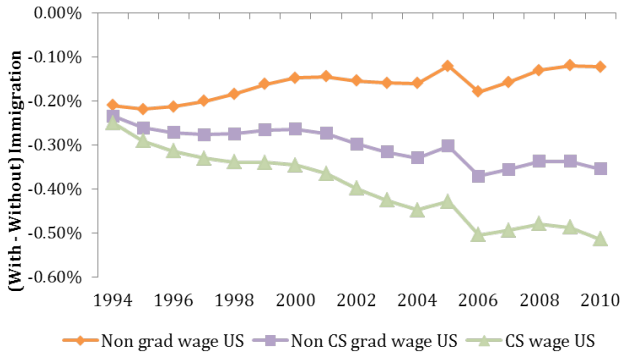
(b) College Graduates in the US



(c) Real Wages in India



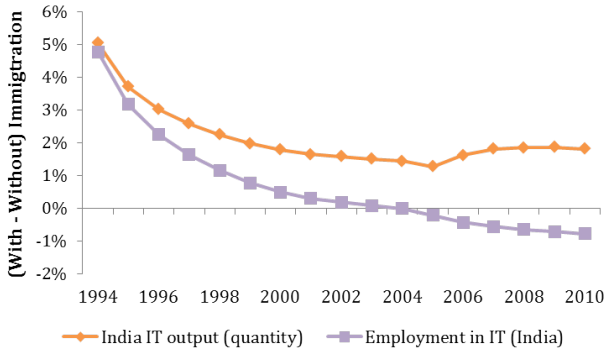
(d) Real Wages in the US



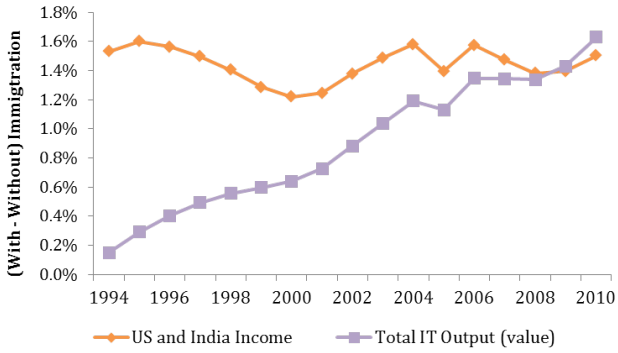
Figures plot the difference between the real scenario under the H-1B regime, and the counterfactual scenario where Indian CS workers are no longer allowed to migrate to the US under the H-1B program. For variable X in year t we plot $\frac{X_{with\ immigration,t} - X_{without,t}}{X_{without,t}}$. Nominal variables are in units of the final good.

Figure 7: Counterfactuals: IT Sector Output and Total Income

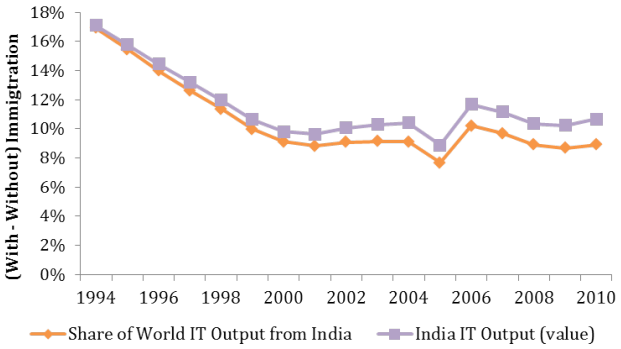
(a) The IT sector in India



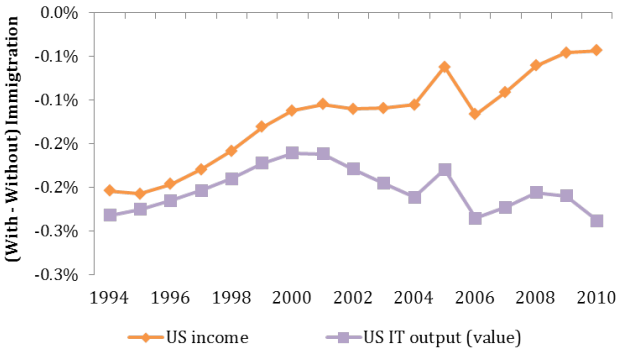
(b) IT Output and Income for IT Producers



(c) India's share of IT Output



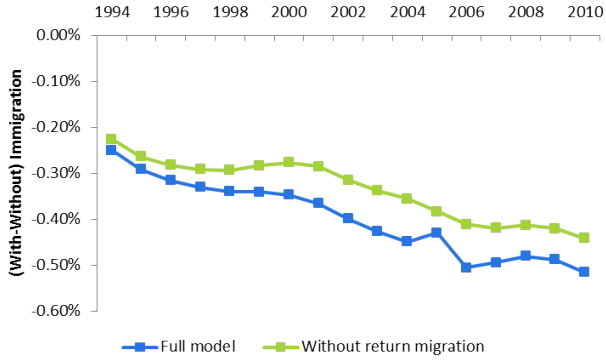
(d) US Income and IT Output



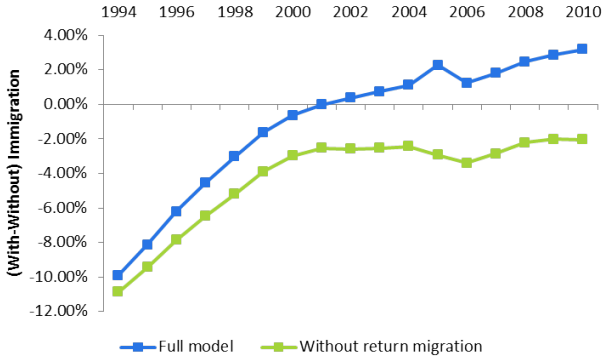
Figures plot the difference between the real scenario under the H-1B regime, and the counterfactual scenario where Indian CS workers are no longer allowed to migrate to the US under the H-1B program. For variable X in year t we plot $\frac{X_{with\ immigration,t} - X_{without,t}}{X_{without,t}}$. Nominal variables are in units of the final good.

Figure 8: Counterfactuals: With and Without Return Migration

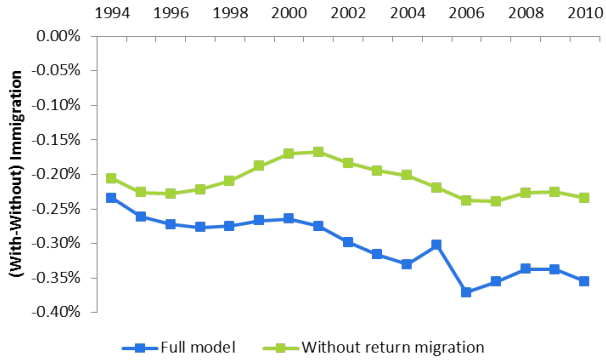
(a) CS Wage - US



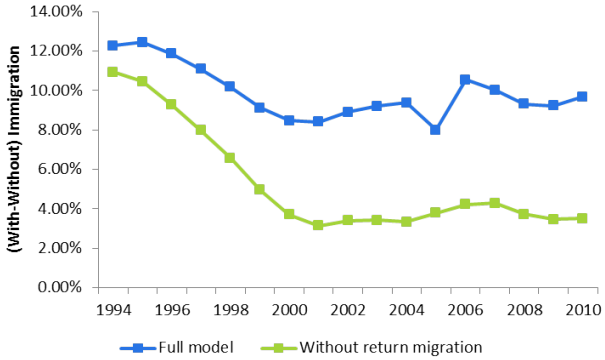
(b) CS Wage - India



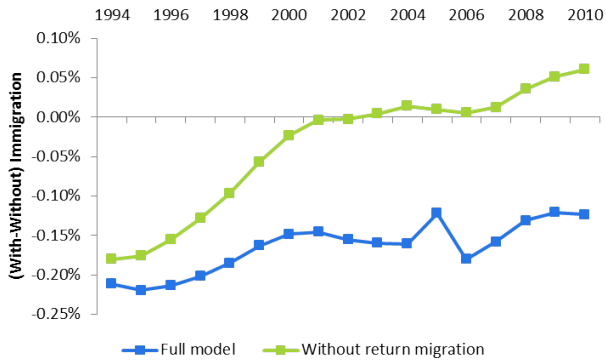
(c) Non-CS College Graduates Wage - US



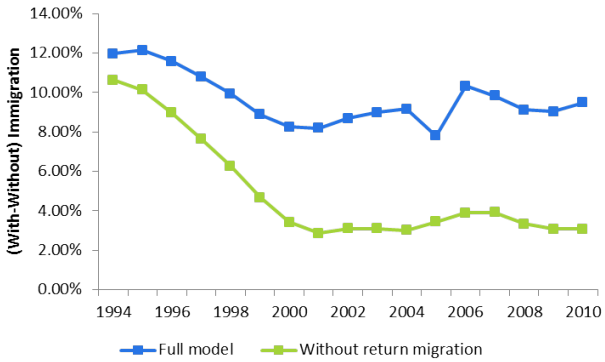
(d) Non-CS College Graduates Wage - India



(e) Non College Graduates Wage - US



(f) Non College Graduates Wage - India



Figures plot the difference between the real scenario under the H-1B regime, and the counterfactual scenario where Indian CS workers are no longer allowed to migrate to the US under the H-1B program. For variable X in year t we plot $\frac{X_{with\ immigration,t}-X_{without,t}}{X_{without,t}}$. For the model "Without return migration" we set the return rate parameter $\varrho = 0$. Nominal variables are in units of the final good.

Table 1: Number of H-1Bs by Firm (Approved)

2013 Rank	Company	Headquarters	Primary Employment Base	2012	2013
1	Infosys	India	India	5600	6298
2	Tata Consultancy Services	India	India	7469	6258
3	Cognizant	USA	India	9281	5186
4	Accenture Inc	Bahamas	India	4037	3346
5	Wipro	India	India	4304	2644
6	HCL Technologies Ltd	India	India	2070	1766
7	IBM(India, Private Ltd.)	USA	India	1846	1624
8	Mahindra Satyam	India	India	1963	1589
9	Larsen & Toubro Infotech	India	India	1832	1580
10	Deloitte	USA	US	1668	1491
11	IGATE(Patni)	USA and India	India	1260	1157

Source: Author's calculations using USCIS reports. The last two columns indicate the number of H-1B visas that were approved for each year.

Table 2: Immigration and the Computer Science Workforce

Year	1970	1980	1990	2000	2010
Computer Scientists as a fraction of workers with a BA/MA	1.68%	1.83%	3.30%	5.66%	5.28%
Computer Scientists as a fraction of STEM college graduates	16.86%	23.60%	35.99%	53.31%	54.90%
Immigrants as a fraction of BA/MAs	2.10%	5.43%	6.86%	8.41%	12.77%
Immigrants as a fraction of Computer Scientists	2.37%	7.09%	11.06%	18.59%	27.82%
Immigrants as a fraction of Other STEM workers	3.63%	9.72%	10.71%	12.69%	18.21%

Note: Sample restricted to employed workers with a Bachelor's or a Master's degree. Definition of Computer Scientists and STEM workers determined by occupational coding (for details see Data Appendix A). Immigrant defined as one born abroad, and migrated to the US after the age of 18.
Source: US Census (years 1970 to 2000); ACS (2010)

Table 3: Location of Highest Degree by Type of Worker for those on Temporary Work Visas

	All Workers			**IT Workers**		
Location of Highest Degree	**Abroad**	**US**	**Total**	**Abroad**	**US**	**Total**
Bachelor's Highest degree	599,500	35,604	635,104	132,600	12,744	145,344
Percentage	94%	6%		91%	9%	
Master's Highest degree	207,964	61,751	269,715	68,301	22,003	90,303
Percentage	77%	23%		76%	24%	

Source: National Survey of College Graduates (2010). Sample of foreign-born workers working in the US on temporary work visas in 2010.

Table 4: Product Market Parameters

Time invariant parameters			
τ	1.7	Ψ	0.5
λ	2	$h\bar{1}b$	0.011
σ	4	ϱ	0.230
θ	8.28	Δ_{US}	0.2
ϵ	3	Δ_{IN}	0.2

Time varying parameters				
	1995	**2000**	**2005**	**2010**
γ_{US}	0.021	0.025	0.028	0.028
γ_{IN}	0.005	0.013	0.022	0.031
γ_W	0.0001	0.0002	0.0003	0.0003
$m_W/(m_{US}+m_{IN})$	3.106	2.301	2.708	3.508
$\alpha_{y,US}$	0.474	0.442	0.416	0.406
δ_{US}	0.176	0.207	0.217	0.226
$\alpha_{y,IN}$	0.466	0.434	0.415	0.411
δ_{IN}	0.147	0.174	0.178	0.271

Time invariant parameters: τ: Elasticity of substitution between college and non college graduates; λ: Elasticity of substitution between CS and non-CS college graduates; σ: Elasticity of substitution between IT varieties; θ: Dispersion parameter of Frechet distribution; ϵ: elasticity of substitution between CS that never migrated to the US and CS that return from the US; Ψ: distributional parameter of CS aggregate in India; $h\bar{1}b$: H-1B cap in the US; ϱ: return rate to India and Δ_k: Extra CS intensity of IT sector compared to Final Goods Sector.

Time varying parameters: γ_k: Cobb-Douglas parameter of IT goods in Final goods sector; $m_W/(m_{US}+m_{IN})$: Relative income of the Rest of the World to India and US; $\alpha_{y,k}$: Distributional parameter of non college graduates in Final goods sector; δ_k: Distributional parameter of CS on both sectors.

Table 5: Normalized Labor Quantities

	1995	**2000**	**2005**	**2010**
		US		
$L_{n,US}$	0.83	1.22	1.46	1.62
$L_{F,US}$	0.02	0.06	0.11	0.15
G_{US}	25.8	30.2	32.8	36.3
H_{US}	74.7	80.0	80.4	76.1
Total US	101.3	111.4	114.7	113.9
		India		
$L_{n,IN}$	0.23	0.41	0.56	0.89
R_{IN}	0.005	0.018	0.031	0.044
G_{IN}	20.6	25.9	31.7	33.7
H_{IN}	291.8	319.6	361.0	365.1
Total India	312.6	345.9	393.3	399.7

US population in 1994 is normalized to 100. India and US populations grow at the rate observed in the data. We use the March CPS and the National Sample Survey to match the shares of each occupation to those observed in the data every year.

Table 6: Labor Supply Calibrated Parameters

		US	India
σ_0	Std dev of study-area taste shocks	0.585	0.102
σ_1	Std dev of occupation taste shocks	0.669	3.427
ζ_0	Mean taste for not studying CS	2.401	-0.799
ζ_1	Mean taste for not working in CS	0.888	-0.113
χ_0	Sector switching cost (constant)	2.311	0.614
χ_1	Sector switching cost (linear in age)	0.616	0.700
χ_2	Sector switching cost (quadratic in age)	0.053	0.086

In the calibration exercise, the years 1997, 2002 and 2007 were used to match the data for employment, wages and enrollment. In the US, wage and employment data come from the March CPS, whereas enrollment data is from IPEDS. In India, wage and employment data come from the NSS, whereas enrollment data is from HRD Ministry. See Appendix A for more details.

Table 7: Real Per Capita Income Changes by Worker Type in going from No Immigration to Immigration (In 1999 US dollars)

		Share of Total Population 2010	**1995**	**2000**	**2005**	**2010**
	Total Income per capita	100%	1885	1936	2056	1743
US	Native CS that stay in CS	0.35%	-128	-146	-177	-202
	Native CS that switch from CS to Other College	0.03%	-105	-113	-133	-151
	Other College graduates that never switch to CS	7.00%	-83	-80	-89	-100
	Non College Graduates	14.80%	-39	-24	-18	-19
India	Indian CS who migrate to US before 1994	0.002%	-128	-146	-177	-202
	Indian CS who migrate to US after 1994	0.03%	38717	36410	34870	31838
	Native CS that stay in CS	0.10%	-418	-37	139	231
	Native CS that switch from Other College to CS	0.04%	77	182	266	299
	Other College graduates that never switch to CS	6.60%	571	400	393	366
	CS that return to India before 1994	0.0005%	-6215	-14505	-20028	-27856
	CS that return to India after 1994	0.01%	10590	9515	9312	10925
	Non College Graduates	71.06%	116	78	74	76

Our income changes are rescaled by total labor income in US and India as well as working population in both countries. Results are in per capita real 1999 US dollars."Native CS that stay" are those that are in the CS occupation with and without immigration, "Native CS that switch" are those that are in one occupation when immigration is allowed but switch occupations when immigration is not, "Other College graduates that never switch" are those that are in other college non-CS occupations with and without immigration. Indian CS who migrate to the US are split in two: those who migrated before 1994 so are not affected by the immigration restriction and those who migrate after 1994 who are affected by the restriction in immigration. Indian CS that return from the US are also divided in two groups: those who returned before 1994 and those who return after 1994, for which if immigration is restricted are assumed to have stayed in India working as regular computer scientists.

A Details of the Data Used

A.1 US Data

Data on earnings, domestic employment and foreign employment used in the calibration procedure and in the descriptive figures come from the March CPS, obtained from the IPUMS and NBER websites. The sample consists of employed persons with at least a BA degree. A person is defined as foreign if he/she was born outside the United States and immigrated after the age of 18. Earnings are deflated to 1999 dollars, and top-coded values are multiplied by 1.4.

In our analysis we drop imputed earnings. In order to identify these imputed values, we use a methodology similar to (Bollinger and Hirsch (2007)). From the IPUMS database we use the qinclongj and qincwage variables, and from the NBER database we use the FL665 flag to identify imputations. The database also contains ten Census Bureau flags that identify a small fraction (less than 1%) of earnings as allocated. Over the period under study around 26% of earnings were allocated. This fraction of imputations varies over time - between 19.14% (in 1994) and 29.47% (in 2003). These numbers are consistent with (Bollinger and Hirsch (2007)) who find that between 1998 and 2006, the non-response rate was about 20%. The small difference in our numbers arises both from using a different sample (restricted to those with BA/MA degree) and because non-response is not the only reason the CPS imputes earnings.

In order to define workers in Computer Science we use the occupational codes and the crosswalk given the categories in the CPS Outgoing Rotation Group (CPS-ORG) data set. The occupational coding in the CPS-ORG up to 2002 uses the 1990 Census definition. We consider as Computer Scientists those under the occupational titles of: "064 Computer systems analysts and scientists" and "229 Computer programmers". For the years 2000-2 the CPS-ORG reports codes using both the 1990 Census definition and the 2000 Census definition. This allows us to create a crosswalk where we weight the 2000 occupational codes by the 2 occupational categories in the 1990 Census.

College enrollment data is based on Integrated Post-secondary Education Data System (IPEDS) Completions Survey. It consists of bachelor's degrees awarded by the NSF population of institutions. We consider enrollment in computer science and electrical engineers as the number of degrees awarded in these fields lagged by 2 years. For 1994 and 1995, enrollment in electrical engineering was not available by native and foreign students but only shown together with all engineering degrees. We impute the data for these two years by looking at the average growth in electrical engineering for 1996-2002.

In some descriptive statistics, we compare the computer science workforce to STEM workers. STEM occupations are defined as engineers, computer systems analysts and computer scientists, computer software developers, operations and systems researchers and analysts, actuaries, statisticians, mathematicians and mathematical scientists, physicists and astronomers, chemists, atmospheric and space scientists, geologists, physical scientists n.e.c., agricultural and food scientists, biological scientists, foresters and conservation scientists, and medical scientists.

We use data on the prices, quantities, costs and value added from the Bureau of Economic Analysis

(BEA) since this source allows us to look into data for specific industry groups. Data on firm entry and exit comes from the Business Dynamic Statistics (BDS), and the 1992 Census' Statistics of U.S. Businesses (SUSB). In these data sets we define the IT sector as the sub-sectors of "Publishing industries, except Internet (includes software)," "Data processing, Internet publishing, and other information services" and "Computer systems design and related services" according to the NAICS 2002 classification. The Non IT sector is defined as all other sectors in the economy.

A.2 Trade Data

Information on Imports, Exports and Rest of the World consumption of IT products from the US and India come from the OECD Trade in Value Added statistics. We use gross exports, gross imports and total GDP data for the "C72: Computer and Related Activities" industry in addition to aggregate numbers by country across all industries. The data is only available for 1995, 2000, 2005 and 2008-2011 so we interpolate the missing years.

A.3 India Data

Data on earnings, employment in different occupations and sectors, and age-shares in each occupation comes from the National Sample Survey (NSS). We use the Employment and Unemployment labor force surveys that come between rounds 50 through 66. These rounds cover 1994 through 2010 with gaps in between, for which we interpolate the macro moments.

NSS is a nationally representative survey used by many researchers studying India. It is the largest household survey in the country, asks questions on weekly activities for up to five different occupations per person, and records earnings during the week for each individual in the household. Computer scientists are defined as "systems analysts, programmers, and electrical and electronic engineers" based on the 3-digit National Occupational Codes (NOC). We use the earnings data for the primary occupation only. The IT sector is restricted to be "software" (code 892 in the National Industrial Classification).

There are various sources for education data, the most comprehensive of which is the Ministry of Human Resources and Development that records number of degrees and universities by type of degree (for example, engineering degrees). We combine this with reports from the All India Council for Technical Education (AICTE) and the National Association of Software and Service Companies (NASSCOM) to also look at the growth in Masters for Computer Application (MCA) degrees.

To get total IT output (and export numbers which we corroborate with our exports data), we use data from the Electronic and Information Technology Annual Reports, and the Indian Department of Electronics reports. Much of this data has been collated and standardized by the Center for Development Informatics at the University of Manchester, UK. The remaining data for summary tabs and graphs are from National Association of Software and Service Companies (NASSCOM).