Staff Working Paper ERSD-2021-10 **3 March 2021**

World Trade Organization

Economic Research and Statistics Division

THE IMPACT OF SERVICES LIBERALIZATION ON EDUCATION: EVIDENCE FROM INDIA*

Enrico Nano
The Graduate Institute Geneva

Gaurav Nayyar
World Bank

Stela Rubínová
ERSD, World Trade Organization

Victor Stolzenburg
ERSD, World Trade Organization

Manuscript date: 22 February 2021

---

* enrico.nano@graduateinstitute.ch (contact author). The findings, interpretations, and conclusions expressed in this paper are those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. They are not intended to represent the positions or opinions of the WTO or its members and are without prejudice to members' rights and obligations under the WTO. Any errors are attributable to the authors.

The Impact of Services Liberalization on Education: Evidence from India*

Enrico Nano (The Graduate Institute Geneva)

Gaurav Nayyar (World Bank)

Stela Rubínová (World Trade Organization)

Victor Stolzenburg (World Trade Organization)

Manuscript date: 22/02/2021

Abstract

This paper studies the impact of services liberalization on education and the gender education gap at the district level in India. We focus on the time period 1987 to 1999 and three services sectors - banking, insurance and telecommunications - which were all state monopolies, have been heavily liberalized in the time frame studied, have relatively high shares of female employment and require high education investments. Our hypothesis is that the national-level liberalization spurred higher investment in education, particularly girls' education, in districts with higher employment growth in these key services sectors. We employ a first difference strategy to control for unobserved time-invariant heterogeneity, use an IV procedure to eliminate other potential sources of bias and control for the simultaneous tariff liberalization. Our results indicate that employment growth in liberalized services sectors is a consistently significant determinant of both the average number of years of schooling (positively) and the gender education gap (negatively). These effects are at least as relevant as those of merchandise trade liberalization, are persistent and driven mostly by the banking and, to a lower extent, the telecommunications sectors. Looking at the transmission channels, we employ a 3SLS strategy and observe that both growing incomes and higher returns to education drive this relationship.

Keywords

Services, liberalization, education, gender, inequality, India

JEL codes

F63, I24, J16, L80, O12

---

* enrico.nano@graduateinstitute.ch (contact author). The findings, interpretations, and conclusions expressed in this paper are those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. They are not intended to represent the positions or opinions of the WTO or its members and are without prejudice to members' rights and obligations under the WTO. Any errors are attributable to the authors.

I. Introduction

There is a large literature that looks at the effects of macro shocks and, in particular, of trade liberalization episodes on micro or meso aspects of development. However, most papers focus on tariff liberalization, thereby limiting the analysis to the manufacturing sector. Being more difficult to measure, the effect of liberalization in the services sector has so far received little attention. This is especially relevant given the trend of premature deindustrialization (Rodrik, 2016) which also largely reflects the fact that services have grown faster than manufacturing in many developing countries (Hallward-Driemeier and Nayyar, 2017). Furthermore, the nature of liberalization in the manufacturing or agricultural sectors is unlikely to be informative for services, many of which need to be delivered face-to-face for technical or regulatory reasons. In addition, despite a growing recognition that gender inequality hinders economic growth, only few studies have analyzed the effects of services liberalization by gender.

In this paper, we jointly address these gaps in the literature by looking at the effect of the services sector's growth, induced by a major liberalization episode for banking, insurance and telecommunications services, on education and the gender education gap (GEG) in India from 1987 to 1999. These services are, on average, significantly more skill-intensive and pay more than agriculture or manufacturing (Nayyar, 2012) so that a rise in such services employment should raise the returns to education. As services employment is, on average, also less male-biased than manufacturing or agriculture (Ngai and Petrongolo, 2017), it might affect the returns to education differentially across genders. Hence, education and the gender education gap are natural indicators to look at when studying the effects of liberalization that spurred the growth in banking, insurance and telecommunication services.

To analyze the effect of services growth on average schooling levels and gender education gaps, we exploit differences across Indian districts in the employment expansion of these three key services subsectors. To obtain exogenous variation in district-level services employment growth, we rely on variation in the initial employment shares of these subsectors across districts. We discuss in detail below and in the next section why initial employment patterns can be assumed to be exogenous to subsequent changes in education levels and gender gaps. We find that the liberalized services subsectors' growth is a statistically significant and economically meaningful determinant of increases in the average years of schooling and declines in the gender education gap, contributing respectively to 4.4% and 8.3% of the overall change in each dependent variable. This finding is robust to instrumentation and various robustness checks, including the simultaneous tariff liberalization episode and expansion of the schooling offer. Looking at the channels behind these findings, we observe that both growing incomes and higher returns to education drive this relationship, with the former prevailing over the latter. Interestingly, we also find that employment growth in these services subsectors prior to the liberalization has no significant impact on our outcome variables highlighting the importance of the reforms for our results.

India is an optimal case study for this research question for several reasons. Firstly, India has been the poster child for services-led development due to the rapid growth of its export-oriented services sector since the 1990s (Basu, 2015). This has created large variations across Indian districts with respect to their dependence on services employment. Secondly, India underwent a series of comprehensive policy reforms in the early

1990s that were, to a large extent, unexpected since they were externally imposed by the IMF in response to a severe balance-of-payment crisis. Thirdly, the services sector in India prior to the reforms was dominated by state monopolies and faced heavy regulation, which led to an arbitrary and highly inefficient allocation of services employment across districts whereby less attractive locations were oversupplied with certain services (Baldwin and Forslid, 2020). As a result, we can use initial district employment shares as an instrument for subsequent employment growth since these shares are essentially unrelated to economic fundamentals. The most relevant regulation was the 1:4 branch licensing policy in place from 1977 to 1990 whereby for every branch a bank opened in a location that already had a bank branch, it was required to open four branches in unbanked areas (Cole, 2009). This policy led to employment levels in the finance and insurance subsectors that were almost arbitrarily located across the country.

Our paper relates to several strands in the literature, beginning with studies that analyze at the effects of trade and globalization on the accumulation of human capital. Methodologically, we build particularly on Topalova (2007) and Edmonds, Pavcnik and Topalova (2010) who use the IMF-induced Indian reforms to show that schooling increased less in rural districts that faced larger tariff declines, particularly for girls. They exploit a local labor market approach (see Section IV), which has recently received a lot of attention in the trade literature due to the work by Autor, Dorn and Hanson (2013) in relation to the China shock. Along the same lines, Greenland and Lopresti (2016) find that Chinese import competition in the US led to an increase in high school graduation rates and a decrease in employment rates for the least educated ones. Mirroring this result, Atkin (2016) finds that during a period of major trade reforms in Mexico the school dropout rate increased with the expansion of less-skilled export-manufacturing jobs, as it raised the opportunity cost of schooling for students at the margin.

Focusing on India, Aghion, Burgess, Redding and Zilibotti (2008) argue that the dismantling of the "License Raj" in 1991 (see Section II) positively impacted manufacturing industries, as also did the tariff and FDI reforms (Harrison, Martin and Nataraj, 2012). However, Mehta and Hasan (2012) and Arnold, Javorcik, Lipscomb and Mattoo (2014) suggest that these are not the only liberalization channels that affected India's manufacturing sector: policy reforms in services also played a large role, as we will show in relation to education. Studying the boom of IT services centers outsourced to India, Shastry (2012) finds that linguistic differences between districts resulted in different propensities to learn English and thus affect the ability of districts to take advantage of the new employment opportunities. From a gender point of view, Oster and Steinberg (2013) show that the IT revolution boosted enrolment in English-speaking schools of girls and boys equally. However, using a randomized controlled trial (RCT), Jensen (2012) finds that an increase in labor market opportunities in the business process outsourcing industry increased education and health outcomes of girls, boosted career aspirations, and delayed marriage and fertility decisions of young women. Similar findings are reported by Heath and Mobarak (2015), who analyze the garment industry in Bangladesh.

More in line with our focus, the literature on the effects of trade liberalization on gender inequality has focused so far on the gender wage gap (GWG). While Saurè and Zaobi (2014), following Galor and Weil (1996), argue that trade can increase the gender wage gap, Juhn, Ujhelyi and Villegas-Sanchez (2015) report the opposite: trade causes more productive firms to upgrade their technology which benefits low-skilled women since it lowers the demand for physical skills. Moreover, Becker's theory suggests that import competition in concentrated industries lowers costly discrimination, and thus the gender wage gap (Becker, 1957; Black and Brainerd, 2004; Menon and Rodgers, 2009).

The remainder of the paper is organized as follows. Section II presents the historical background and the conceptual framework on which the paper relies. Section III describes the data and provides some descriptive statistics. Section IV discusses the empirical strategy. Section V shows the main results and their robustness. Section VI analyzes potential mechanisms behind the findings. Section VII concludes.

II. Background and conceptual framework

Following a balance of payment crisis in August 1991, India experienced a large and relatively sudden wave of liberalization in the 1990s, which was externally induced by the IMF. To make the economy more market- and services-oriented, several sectors were liberalized across four dimensions. First and most studied, tariff cuts were implemented across almost all manufacturing industries and the average tariff dropped by more than 60% between 1991 and 1997 (Topalova, 2007). In addition, average tariffs faced by India dropped substantially in this period because many other countries were also undergoing a phase of liberalization. Second, the progressive dismantling of the "License Raj" – a complex system of licenses that were required to operate a business dating back to 1947 – also started in 1991. The liberalization of foreign direct investment, allowing foreign investors to hold equity of a business up to 100% in many sectors, and labor market deregulation, which, depending on the sector, turned the regulation either more pro-worker or more pro-employer (Aghion, Burgess, Redding and Zilibotti, 2008) were the third and fourth dimensions of liberalization.

We focus on banking, insurance and telecommunications services that were state monopolies but heavily liberalized in the time period studied, require higher education investments, and are relatively female-friendly. These characteristics would be expected to increase average schooling and lower the gender education gap. These sectors have also been previously studied by Arnold, Javorcik, Lipscomb and Mattoo (2014) who analyzed the impact of liberalization in these sectors on Indian manufacturing firms.[1] Table 1 summarizes their qualitative liberalization index, which goes from 0 to 5 depending on the intensity of reforms. Of main interest to the present study is that all these subsectors were not liberalized at all before 1993 but largely liberalized by 2003. We note that our main period of interest is 1987 to 1999 and liberalization in the insurance sector only began in 1999, but we include the sector nevertheless in light of potential anticipatory effects and for robustness checks extending the period studied to 2005.

[1] Arnold et al. (2014) also studies the transport sector, which we do not include in our analysis as it has completely different characteristics compared to the other three, such as very low returns to education and a strong bias towards male employment.

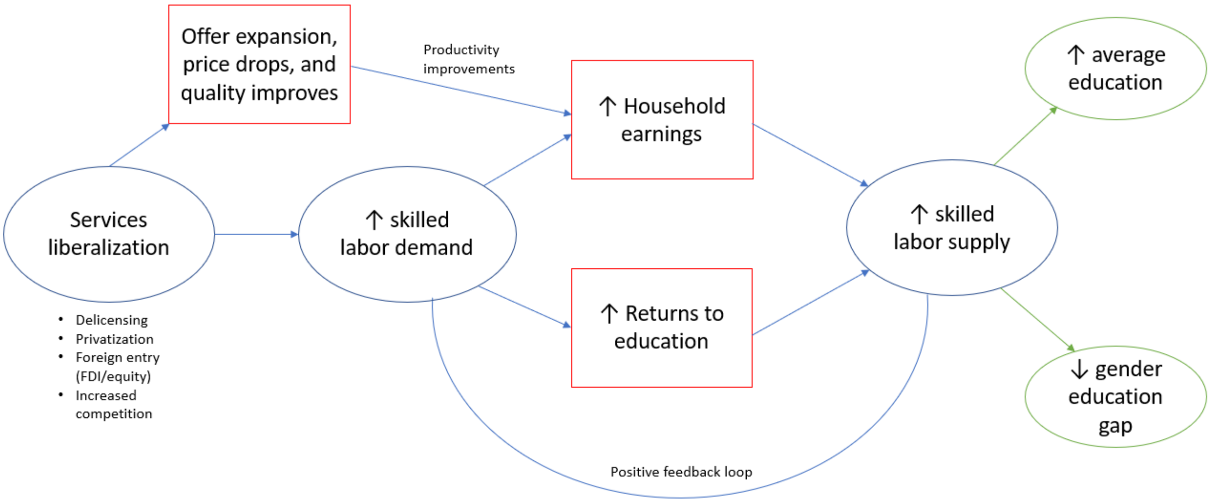
Table 1. Liberalization index of selected sectors in India

	Banking	Insurance	Telecommunications
0-1	**1993/4** Legislation passed to signal government's in-principle approval of new private entry into banking sector. Seven new banks enter the market. FDI up to 20% but foreign banks are barred. Banks given more freedom to allocate their inventories and receivables across different items.	**1999/00** Bill passed to open up the insurance sector to private entry, including foreign equity participation up to 26%. Substantial freedom with respect to pricing but strict regulatory supervision. Discretionary entry permission was required and no acquisitions possible due to public sector dominance.	**1993/4** The first private networks in industrial areas were licensed and put in operation. Licensing process for cellular service begins, envisaging the possibility for foreign participation.
1-2	**2000/1** Discretionary barriers to entry into banking sector are lowered significantly. State signals its intent to withdraw from the banking sector eventually.		**1994/5** Private mobile phone service providers emerge in major cities, all of which have some foreign equity. Process of issuing further licences to private sector begins. New Telecom Policy announced to define framework for further private sector participation FDI possible up to 49%.
2-3	**2001/2** Major interest rate deregulation allows banks to set prices more freely. Private sector banks gain more relevance as lenders and begin to crowd out public sector banks in some instances.	**2002/3** Entry of 12 new private providers of insurance services, which constitutes a massive shake-up of the market. Competitive pressure on incumbent public insurers. FDI ceiling remains at 26%.	**1999/00** New Telecom Policy issued which defines the way ahead for a complete opening of national and international long distance market. Regulator strengthened, licensing fee arrangement made more favourable for private operators.
3-4	**2002/3** Foreign participation in Indian banks is made significantly easier. Clearance for up to 49% of equity is automatic and majority ownership is possible subject to case-wise approval.		**2002/3** National long distance market fully open with no restrictions on the number of operators. Public monopoly in international gateways abolished.

Source: Arnold, Javorcik, Lipscomb and Mattoo (2014).

The conceptual framework underlying the data analysis is presented in Figure 1. Services liberalization leads to an initial increase in the demand for skilled labor, since the liberalized services sectors require relatively high education levels. Since these sectors also pay more than the average, their expansion leads to an increase in the returns to education and, for those who take up these jobs, to an increase in household earnings. General productivity improvements, spurred by an offer expansion, price drops and quality improvements in these key business services may increase household income for workers in other sectors as well (see Arnold et al., 2014). Through the returns to education and household earnings channels, people are induced (or push their children) to undertake more schooling than they would otherwise, as they aspire to work in these services sectors once they have finished the required minimum education. In an agricultural society with a very low average level of schooling, this often translates into the ambition of completing middle or secondary education that amount to, respectively, 8 or 12 years of schooling in India. The increased average education also initiates a positive feedback loop: a rise in the supply of skilled labor attracts more firms, and, as the sectors expand, the demand for skilled labor further increases, reinforcing the cycle.

Figure 1. Conceptual framework



Source: authors' elaboration.

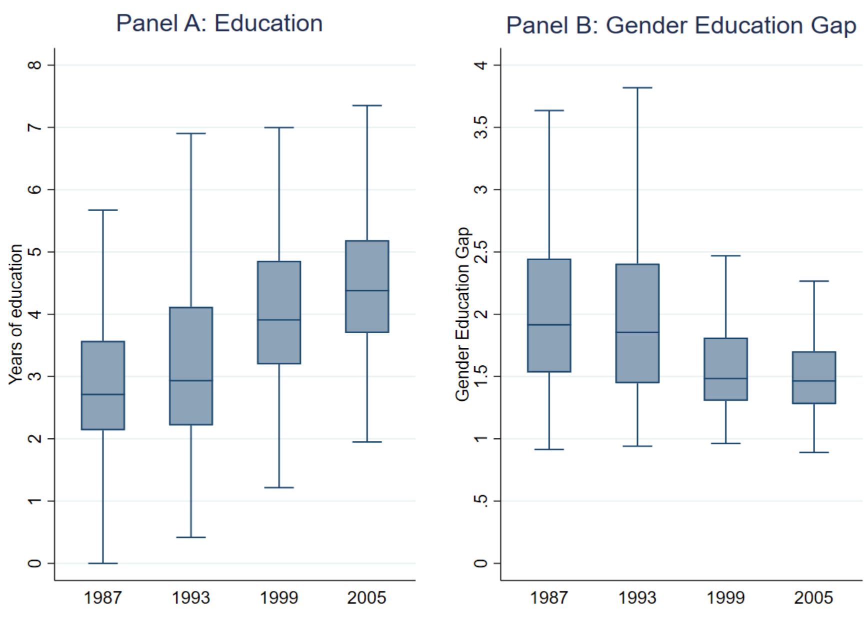
In addition to raising average schooling, services sector liberalization is also likely to increase women's education more than men's, thereby lowering the gender education gap. This is supported by several hypotheses tested in the literature. Munshi and Rosenzweig (2006) find that the Indian caste system tends to channel men to traditional blue-collar jobs, so that women will be more likely to take up new white-collar job opportunities that arise in the services sector. Another strand of literature supports the hypothesis that technological upgrading increases female employment as capital is more complementary to women's than men's labor (Galor and Weil, 1996; Saurè and Zaobi, 2014; Juhn, Ujhelyi and Villegas-Sanchez, 2015). This follows the general hypothesis that women have a comparative advantage in skilled tasks and enjoy higher returns to education (Pitt, Rosenzweig and Hassan, 2012). In the Indian context, it has been shown that declining poverty, through rising household earnings, disproportionately increases girls' schooling, because boys already had priority in school enrolment (Edmonds, Pavcnik and Topalova, 2010). In addition, there is evidence that mothers value children's (and particularly girls') education more than fathers: as a result, if mothers gain more bargaining power, they will improve girls' schooling more than boys' at the margin (Thomas, 1994; Qian, 2008).

III. Data and descriptive statistics

We use three sources of data for our analysis. The principal data are drawn from household/individual level employment surveys conducted by India's National Sample Surveys Organization (NSSO). These surveys contain detailed and nationally representative employment and education data, and we use four waves of their repeated cross-sections: 1987-88, 1993-94, 1999-2000 and 2005-06. For some robustness checks we also employ the 1983 wave, which however lacks the district identifier. To implement a local labor market approach (see Section IV), we create a panel from the repeated cross-sections by averaging the variables of interest at the district level. Owing to district boundary changes due to administrative mergers and splits, we build a harmonized set of around 400 parent-districts, whose area is weakly larger than the actual districts' boundaries in each year but is stable over time.

Second, we use the Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) database from the Development Data Lab, which contains detailed information on the number of schools across India's towns and villages. We sum the number of schools for each district and use this to proxy for the expansion of the schooling offer during the years under study. Given that these data are available only for the Population Census years (1991, 2001, 2011), we merge the 1991 data to our 1987-88 wave and the 2001 data to our 1999-2000 wave. The merge is imperfect given the different district harmonization processes and, when including the district number of schools as a covariate, we lose 18 districts out of 391 (less than 5%). Third, to control for the contemporaneous merchandise trade liberalization, we add average tariffs applied and faced as covariates based on WITS (World Integrated Trade Solutions) Trade Stats, which integrates data from both the UNCTAD TRAINS and the UN COMTRADE databases.

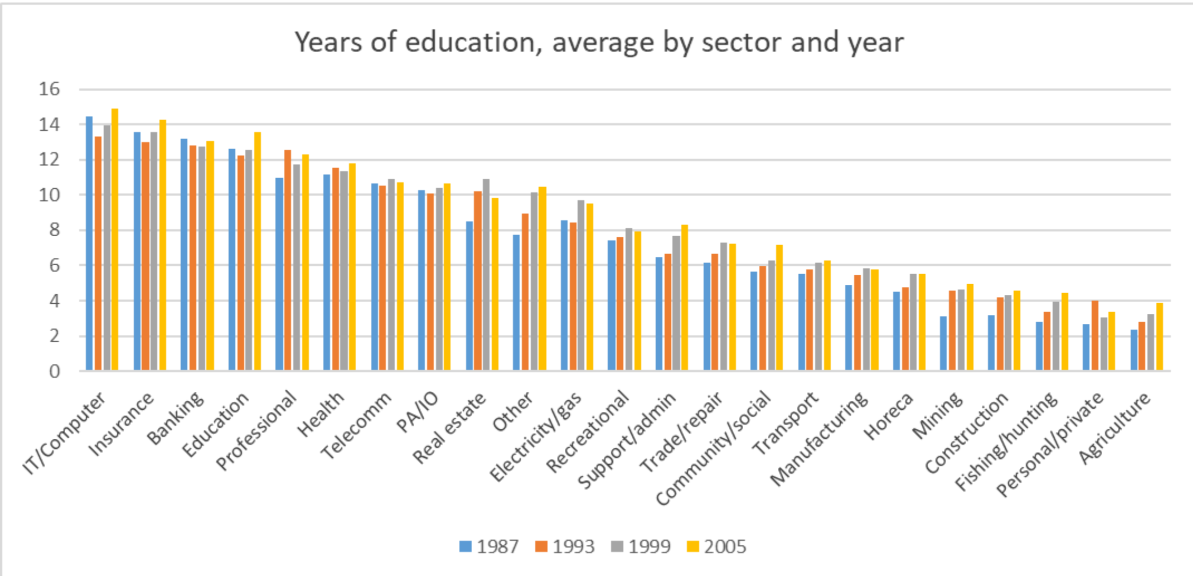
Figure 2. Box plots of average years of education and gender education gap in India



Note: the figure shows box plots with the median, minimum, maximum, first and fourth quartiles of the average years of schooling (Panel A) and gender education gap (Panel B) of sampled individuals in each NSS wave.

Figure 2 shows the change in our two outcome variables of interest – the average years of education and the gender education gap – over time. As shown in Panel A, average education in India increased its mean and median, particularly between 1993 and 1999, indicating an improvement of schooling levels right after the first liberalization wave. The period from 1999 to 2005 also shows positive developments. The gender education gap, defined as the ratio between average education of men over average education of women at the district level, has instead declined over time reaching progressively closer to 1, i.e. gender parity. Again, a major discontinuity is reported between 1993 and 1999 after which it slowly decreases until an average of 1.45 is reached in 2005. Figure 3 shows that the services sector in general requires higher average education levels, while the agriculture, fishing and hunting sectors lie at the end of the distribution. Importantly, the three liberalized services subsectors in our analysis – banking, insurance and telecommunications – all exhibit high average education levels.

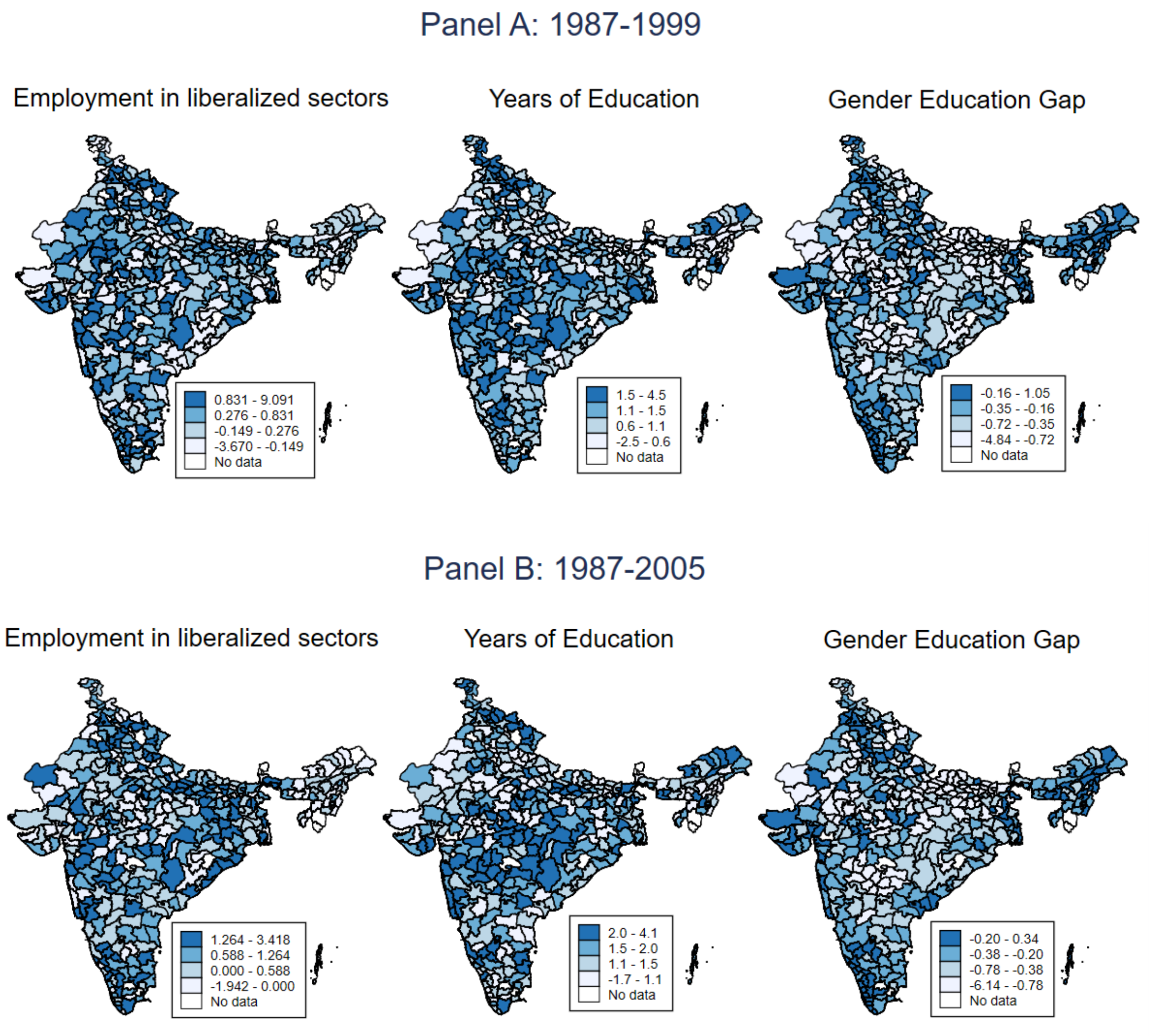
Figure 3. Average years of education by sector in India



Note: the figure shows the average years of schooling across sectors of employment in each NSS wave.

As our analysis relies on geographical variation within India, Figure 4 shows how changes in the two outcome variables relate to changes in our main independent variable, the share of people employed in the three liberalized services sectors (see next Section), at the district level. The changes in employment shares and average years of education are positively correlated across both the 1987-1999 (Panel A) and the 1987-2005 (Panel B) periods. In contrast, the gender education gap maps display a negative correlation with the other two variables: the color patterns seem exactly reversed. To study these relationships more thoroughly, we move to the econometric analysis in the next Section.

Figure 4. Distribution across Indian districts of selected variables



Note: the figure shows the average employment shares of liberalized services sectors, years of schooling and gender education gap across Indian districts, in differences, for 1987-1999 (Panel A) and 1987-2005 (Panel B).

IV. Empirical strategy

As we study the effect of services liberalization, which is a national policy, we cannot rely on a simple proxy measure such as the index in Table 1, since it would only have variation over time. Given India's dimension and its division into hundreds of districts, we employ a local labor market approach, as in Topalova (2007). We use districts as our unit of analysis since they are the relevant local labor market unit.[2] Considering each district as a local labor market, we can thus proxy the degree of exposure to services liberalization by measuring the proportion of people employed in to-be-liberalized sectors in a given district.

In order to take into account the potential bias stemming from unobserved time-invariant heterogeneity, we employ all variables in differences, thereby studying the effect of the change in the district proportion of people employed in liberalized sectors on the change in the district average years of education and the gender education gap. As the Indian liberalization wave kicked off in 1991 and according to Arnold, Javorcik, Lipscomb and Mattoo (2014) the three sectors (banking, insurance and telecommunications) started to be effectively liberalized in 1993, we take 1987 as a base year rather than 1993. This also takes into account the potential role of any anticipation effect. We thus study the impact from 1987 to 1999 and, to test if results are persistent, we also look at the long-term effect from 1987 to 2005. Our baseline specification can thus be written as:

$$(y_{d,t+k} - y_{d,t}) = \beta(S_{d,t+k} - S_{d,t}) + \gamma(X_{d,t+k} - X_{d,t}) + (\varepsilon_{d,t+k} - \varepsilon_{d,t}) \qquad (1)$$

where $y_{d,t}$ is one of the two dependent variables, i.e. average years of education and gender education gap, in district d and year t. k is equal to either 12 (for 1999) or 18 (for 2005). $S_{d,t}$ is our main independent variable of interest, which is the sum over all sectors of the share of workers employed in sector i and district d, $L_{i,d,t}$, interacted with a dummy variable for whether that sector was initially closed to private ownership and subsequently liberalized (the dummy is thus equal to 1 for banking, insurance and telecommunications). This allows us to interact the liberalization of services sectors at the national level with a district-specific labor market characteristic: districts with different prevalence of employment in different sectors have been heterogeneously impacted by the liberalization wave, thus affecting to a different extent our dependent variables.

$$S_{d,t} = \sum_i \frac{L_{i,d,t}}{L_{d,t}} * \mathbf{1}(Liberalized_i = 1) \qquad (2)$$

Among the district-level covariates in $X_{d,t}$ we include the following variables: proportion of women, married individuals, Hindus, members of backward (disadvantaged) classes[3], people living in rural areas and individuals that recently migrated[4], as well as average age, household size and household monthly

---

[2] A low documented level of migration across Indian districts prevents us from attributing the estimated effects to our variables of interest when instead they are due to migration (Rosenzweig and Evenson, 1977; Duflo and Pande, 2007; Topalova, 2010). We nevertheless control for the potential effect of migration in all baseline regressions.

[3] Backward classes include scheduled tribes and scheduled castes.

[4] Migrants are defined in the questionnaire as people whose place of enumeration differed from last usual place of residence, thus capturing also migration across districts. This variable is available only for the 1987 and 1999 waves.

expenditure per capita. We also include state fixed effects to capture potential factors that are common in districts within the same state. In addition, we control for the share of people employed in other high-skill services sectors[5], in other services (low- and medium-skill level) and in the manufacturing sector, to take into account the potential effect of the expansion of other sectors during the liberalization wave.

To consider the effect of the contemporaneous expansion of schools, we add the number of schools (of all levels from elementary schools to universities) per 100,000 people as a covariate. Moreover, to account for merchandise trade liberalization that took place during the same period, we compute the average district-level MFN tariffs, by weighting a sector's tariff[6] with the district employment share of that sector at baseline (in 1987), as in Topalova (2007) and in Edmonds, Pavcnik and Topalova (2010). Given that the 1990s were a period of trade liberalization in other countries as well, we also compute the average district-level tariffs faced by India, by applying the same district employment weights of the own tariffs to the average tariffs imposed on India by its trade partners, weighted by import volume at baseline, as in McCaig (2011).

a. Identification

To address a potential simultaneity bias, in particular with respect to the average education outcome, we employ an Instrumental Variable strategy. As an increase in education in a district might attract services firms to this district leading to reverse causality, we instrument the main independent variable, the change in the employment share of liberalized services sectors, with the pre-reform (1987) level of the employment share, $S_{d,1987}$. The exclusion restriction is that this instrument does not directly affect the change in average education or gender education gap in a district, after controlling for the covariates.

Our instrument is motivated by a very specific policy environment in India surrounding the three sectors that we study. All three were state monopolies and subject to a set of binding regulations that served policy purposes such as addressing the rural-urban divide rather than increasing efficiency. As a result, location choices by firms in these sectors prior to liberalization were heavily distorted and highly inefficient, leading to attractive locations being underserved. We rely, in particular, on two policies that resulted in almost arbitrary location decisions of banking, insurance and telecommunications services. The first policy was the introduction and removal, respectively, in 1977 and 1990 of a 1:4 branch licensing policy whereby for every branch a bank opened in a location that already had a bank branch, it was required to open four branches in unbanked areas (Cole, 2009). In fact, the expansion of rural bank branches was relatively higher in financially less developed states between 1977 and 1990 but the reverse was true before 1977 and after 1990 (Burgess and Pande, 2005). As a result, 1987 employment levels in finance and insurance are unrelated to economic fundamentals of districts, including their education levels. However, the removal of the policy led

[5] The other high-skill services sectors include IT and computer-related services, the Public Administration and International Organizations (PA/IO), health, education and professional services sectors. This group was defined based on the average number of years of schooling in the sector at baseline.

[6] The sector tariff is in turn a simple average at the HS2 level of the MFN tariff lines, including Ad Valorem Equivalents.

to a reversal of 1977 to 1990 location decisions such that 1987 employment levels are a significant predictor for subsequent employment growth.

Similarly, the initial expansion of telecommunication services was driven by public-private partnerships with the aim to set up special development zones such as Keonics City which was established in 1978. The decision where to locate these development zones was driven primarily by the availability of land and by land prices rather than by attractiveness of a location for the telecommunications industry according to anecdotal evidence (Prashanth, 2013). Hence, similar to banking and insurance, we expect to observe a negative relationship between 1987 employment levels and subsequent employment growth as employment shifted after the liberalization towards districts with better economic fundamentals.

In line with the expected effect of the policies, the first stage regressions show a significantly negative relationship between the initial levels and the change of the district employment shares in liberalized services sectors.[7] This means that districts with smaller initial shares experienced larger employment expansion in these sectors than districts with larger initial shares, regardless of the time lag considered. This also addresses a possible issue regarding the validity of the instrument which could arise if, because of aspirational motives, people in districts with high baseline employment shares in to-be-liberalized (and high-paying) sectors are induced to take up more education. This would result in agglomeration effects, for which districts with higher baseline employment shares in banking, insurance and telecommunications would also experience the largest increases of such shares over time. If such a mechanism drove our results, we would see a negative relationship between the change of district employment shares in liberalized sectors and the change of average years of education in the second stage, which is not the case. The first stage results also show that our instrument is strong, with the F-statistics of the Kleinbergen-Paap weak-identification test ranging between 29 and 60.

Although we believe that our instrument solves potential endogeneity concerns, we have run additional robustness checks employing an alternative instrument, based on the 1983 wave, that is arguably more exogeneous than the 1987 one. Unfortunately, this wave does not have a district identifier, so we interact the 1983 state-level employment shares in liberalized services sectors with the 1987 district population. We do not use directly this 1983-based instrument in our estimations since it lacks power in predicting our district-level independent variable. However, it can be helpful to test whether the main IV is exogenous, by employing the two instruments together and running several overidentifying restrictions tests. In fact, all tests lead to the conclusion that our identification strategy is valid: the description of our methodology and the results are available in Appendix A1.

[7] This stands in sharp contrast to evidence from less regulated countries such as the United States where industries tend to locate in local labour markets due to attractive characteristics of these localities which in turn triggers further sectoral agglomeration. Studies have relied on the persistence of such location decisions to obtain identification on questions related to, for instance, the routine task intensity of local labour markets (see, e.g., Autor et al., 2015). These studies instrument current task intensity with task intensity from previous decades, typically obtaining a strong positive relationship. Our negative first stage results thus underscore the stark and arbitrary distortions caused by inefficient policies in India before the 1990s.

V. Results

The results below look at the effect of employment growth of liberalized services sectors in a district on (a) the average number of years of education and (b) the gender education gap, from 1987 to 1999. In all sets of regressions, we first report a simple correlation in first differences between the dependent and the main independent variable, and then gradually add covariates, estimating both OLS and instrumental variable specifications. For every variable we report (1) the simple regression coefficient, (2) the robust standard errors in parentheses and (3) the beta coefficient in brackets.[8] The latter allows for an easier interpretation than (1): a beta coefficient of x means that an increase by one standard deviation of the variable increases the dependent variable by x standard deviations. Furthermore, we present additional robustness checks, namely on long-term persistence, pre-trends, and the exclusion of districts with the largest cities. Finally, we document heterogeneities in our results across different time periods, rural and urban areas as well as the three services subsectors under consideration.

a. Average years of education

Table 2 presents our baseline results for the effect of services liberalization-induced employment growth on years of education from 1987 to 1999. The change in the employment share of liberalized services sectors is always highly significant, both in the simple OLS and the IV setup. While the first column, OLS0, is close to a simple correlation between the change in district's average years of schooling and the change in proportion of people employed in liberalized sectors (conditional on state fixed effects), adding demographic covariates in column OLS1 leads to a substantial drop in the magnitude of the coefficient. Instrumentation reveals a larger impact of the independent variable of interest, with OLS results always smaller than the IV results. However, the gap between OLS and IV estimates reduces when adding all relevant covariates, and becomes statistically insignificant in the full specification.

In columns OLS2 and IV2 we add other district-level covariates, i.e. the number of schools as well as the share of people employed in other high-skill services sectors and in all other services sectors (low- and medium-skill), which lead to a further decline in the main coefficient estimates. Finally, in columns OLS3 and IV3 we also control for the average tariffs imposed by India and the average tariffs faced by India, which further reduce the coefficients' magnitude but not their statistical significance.[9] In column FS3, the first stage regression for the full specification is presented. The instrument negatively predicts the

[8] Given that the normality and i.i.d. assumptions are often rejected in many IV applications and that, as Young (2019) shows, the bootstrap often reveals severely asymmetric IV confidence intervals, we re-ran all regressions bootstrapping the standard errors with 200 replications: the significance of the estimates does not substantially change. Alternatively, we clustered standard errors at the state level. This clustering increases some p-values, but the main findings do not change. Results are available upon request.

[9] For the sake of consistency, we also tested the OLS and IV specifications with the first and second sets of covariates using the sample of the full specification. Results (not reported) are virtually the same.

employment share in liberalized sectors and is statistically significant at the 1% confidence level. Furthermore, the Kleinbergen-Paap F-statistic is as high as 60.3 indicating that the instrument is strong.

Taking column IV3 as the preferred specification, we note that an increase by one standard deviation in the employment share of liberalized services sectors increases the average years of education in a district by about one sixth of a standard deviation. Given the strong increase in services employment in many districts over our sample period, this translates into a relative effect whereby moving a district from the 25th percentile of the employment share change distribution to the 75th percentile increases the impact on education by a factor of about 6. To gauge this effect in relation to India's overall human capital accumulation over that period, we performed some back-of-the-envelope calculations by computing the average effect of our independent variable and dividing this by the average change in the dependent variable. We find that, on average, the growth in employment shares of liberalized services sectors has contributed to 4.4% of the overall change in average years of education across districts. This is a sizeable contribution considering all other factors that potentially influenced human capital accumulation in those years.

Table 2. Dependent variable: years of education (1987 – 1999)

	OLS0	OLS1	IV1	OLS2	IV2	OLS3	IV3	FS3
Dependent Variable	Education	Education	Education	Education	Education	Education	Education	Empl liberalized sectors
Empl liberalized sectors	0.167***	0.100***	0.323***	0.084***	0.234***	0.068***	0.114**	
	(0.034)	(0.028)	(0.069)	(0.026)	(0.053)	(0.024)	(0.045)	
	[0.242]	[0.149]	[0.479]	[0.123]	[0.342]	[0.101]	[0.169]	
IV: 1987 empl share liber								-0.889***
								(0.108)
								[-0.666]
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES
Observations	392	391	391	373	373	371	371	371
R-squared	0.071	0.404		0.498		0.543		0.348
Kleibergen-Paap F-stat			29.2		41.5		60.3	60.3

Note: robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

In table A2 in the Appendix we report the full regression estimates. Looking at the demographic covariates, we find similar patterns as reported in the literature. The proportion of women and of married individuals are negatively correlated with the district average education, suggesting that in a couple it is often only one of the two (typically the man) increasing his or her education. As expected, the positive sign on household expenditure per capita and the negative sign on the proportion of people living in rural areas proxy the fact that districts with higher average wealth are also more educated. The same applies to districts with older populations and whose households have a larger size, as they are likely to have more educated individuals. Not surprisingly, a higher share of members of backward classes (scheduled castes and scheduled tribes) is negatively (but not significantly) associated with the average schooling in the district, while the proportion of Hindus (the most common religion in India) has a positive coefficient, although turning insignificant when adding the tariffs covariates.

Interestingly, the proportion of migrants (including those who relocated from a different Indian district) is not significant, which provides further support to the use of the local labor market approach since it corroborates the evidence claiming that migration across districts is limited in India (Rosenzweig and Evenson, 1977; Duflo and Pande, 2007; Topalova, 2010). It is also surprising that the number of schools per district is not significant in any regression. That said, the proportion of people employed in the education sector within the share of workers in other high-skill services is positively and significantly associated with district average schooling. Consistent with our expectations, the estimated impact of the share of workers in low- and medium-skill services sectors is close to zero and turns insignificant when adding the tariffs covariates. Own and faced average district tariffs are not significant in these regressions which suggests some degree of multicollinearity among these covariates.

b. Gender Education Gap

Does an increase in the employment share of liberalized services sectors also reduce the distance between the average education level of men and women? Table 3 confirms that it does, as both the OLS and the IV specifications show a significantly negative effect of the main independent variable on the gender education gap. Again, the simple OLS results seem to be biased towards zero, but when adding covariates the gap with the IV estimates becomes smaller, particularly in the last two columns. Nevertheless, looking at the full model, the IV3 coefficient is still more than twice as large as the OLS3 coefficient. Based on the preferred IV3 specification, an increase of one standard deviation in the employment share in liberalized services sectors lowers the gender education gap by one fifth of a standard deviation. The economic quantification following the back-of-the-envelope calculations indicates that our independent variable, on average, contributed to 8.3% of the overall change in the gender education gap across districts.

Table 3. Dependent variable: gender education gap (1987 – 1999)

	OLS0	OLS1	IV1	OLS2	IV2	OLS3	IV3	FS3
Dependent Variable	GEG	GEG	GEG	GEG	GEG	GEG	GEG	Empl liberalized sectors
Empl liberalized sectors	-0.051**	-0.048**	-0.276***	-0.046**	-0.236***	-0.040*	-0.099***	
	(0.025)	(0.023)	(0.067)	(0.023)	(0.056)	(0.022)	(0.038)	
	[-0.104]	[-0.098]	[-0.566]	[-0.091]	[-0.470]	[-0.080]	[-0.197]	
IV: 1987 empl share liber								-0.889***
								(0.108)
								[-0.666]
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES
Observations	392	391	391	373	373	371	371	371
R-squared	0.012	0.140		0.191		0.244		0.348
Kleibergen-Paap F-stat			29.2		41.5		60.3	60.3

Note: robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

As reported in Table A3 in the Appendix, a district's share of Hindus, of married people and of individuals living in rural areas are associated with a higher gender education gap. Conversely, a higher average household size (although not in all specifications) and household expenditure per capita are negatively correlated with the gender education gap. Again, the district proportion of migrants is not significantly linked to the outcome variable, while the number of schools per capita is associated with a reduction in the district's gender education gap. Finally, while in this set of regressions the employment shares of other services sectors are not a consistent predictor, the district tariff covariates are. Higher average own tariffs as well as faced tariffs (although less significantly) are associated with a higher gender education gap. This implies that their reduction, through the contemporaneous merchandise trade liberalization, also benefited gender parity in human capital accumulation.

c. Robustness checks

In this subsection we run three additional robustness checks related to the main analysis, namely on the persistence of our results to the 1987-2005 period, on the existence of pre-trends and on whether our results are affected by districts that contain the largest cities.

i. Long-term persistence

In Table 4, we confirm the results of the previous subsection by looking at a longer time frame, from 1987 to 2005, to test whether the effects are persistent in the long term. For the sake of brevity, we report just the IV results, for both the years of education and gender education gap outcome variables, as well as the first stage results for the full specification, condensed in one table. The only difference with the specifications in Tables 2 and 3 is that two covariates are missing, the district number of schools per capita and the share of migrants, which are not available for the 2005 wave. To control for the expansion of the schooling offer, however, we include a different proxy: the district share of people employed in the education sector (separately from the other high-skill services sectors). We do not have a substitute for the proportion of migrants, but, as mentioned before, it was never significant in the 1987-1999 regressions, confirming the existing literature on the topic.

Table 4. Dependent variables: years of education and gender education gap (1987 – 2005)

	IV1	IV2	IV3	IV1	IV2	IV3	FS3
Dependent Variable	Years of education			Gender Education Gap			Empl liberalized sectors
Empl liberalized sectors	0.441***	0.236**	0.066	-0.313***	-0.266***	-0.152**	
	(0.137)	(0.101)	(0.071)	(0.110)	(0.093)	(0.062)	
	[0.642]	[0.344]	[0.096]	[-0.524]	[-0.444]	[-0.255]	
IV: 1987 empl share liber							-0.543***
							(0.074)
							[-0.463]
State FE	YES	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	YES	YES	NO	YES	YES	YES
Tariffs covariates	NO	NO	YES	NO	NO	YES	YES
Observations	392	392	391	392	392	391	391
R-squared							0.369
Kleibergen-Paap F-stat	21.6	25.5	48.4	21.6	25.5	48.4	48.4

Note: robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Looking at the results with years of education as outcome variable (left side of Table 4), we observe that the magnitude of the coefficients on the district share of liberalized services employment is about the same as in Table 2 (i.e. for the 1987-1999 period), until we include the tariffs covariates. Then, the coefficient on the main independent variable of interest becomes insignificant, but so do the tariffs covariates themselves (full regression tables are available upon request). This suggests that, while the effect of services liberalization on average education is somewhat persistent over time, it cannot be disentangled from the merchandise trade liberalization's effect in the long term.

On the right side of Table 4, the results for the gender education gap suggest an even stronger persistence of the patterns found in Table 3. The coefficient on the share of workers in liberalized services sectors is negative and significant in all specifications and, in the IV3 column, is even larger than the corresponding one for the 1987-1999 period. This supports the claim that the gender effect of services liberalization on education is persistent and may even become stronger over time. The absolute magnitude is also about the same as that of tariffs (which are statistically significant in this case). In conclusion, services sector liberalization strongly and persistently affects education, particularly for women, and its effect is at least as relevant as that of merchandise trade liberalization.

ii. Pre-trends

We proceed by testing whether there are pre-existing trends that could explain the relationship under study. To do so, we employ the 1983 wave, the only available wave before 1987, which however does not have the district identifier. As a result, we cannot run a district-level analysis, but only a state-level one, which reduces our sample size to less than 30 observations. We conduct this test relying on similar specifications as in the baseline regressions, but with the dependent variable and covariates expressed in changes between 1983 and 1987.[10] The key difference is that the share of workers in liberalized services sectors remains expressed in changes from 1987 to 1999. This is intended as a conditioning variable as we are interested in whether districts that saw an increase in services employment in the 1987-1999 period already saw an increase in the years of schooling or a decline in the gender education gap between 1983 and 1987. The results in Table 5 confirm that this is not case, since the coefficients in all specifications are not significant. We show in section V.d.i, where we re-run our main specifications at the state level, that this is not due to the small number of observations. In fact, results for both the 1987-1999 and the 1987-2005 periods, for both outcome variables and for both the OLS and the IV setup are statistically significant at the state level.

[10] Regressions include as covariates: proportion of women, married individuals, Hindus, members of backward classes and people living in rural areas, average age, household size and household monthly per capita expenditure, as well as the district shares of people employed in other high-skill services sectors, in all other services (with low- and medium-skill level) and in manufacturing sectors. Other covariates were not included as not available for all years.

Table 5. Pre-trends test. Dependent variables: years of education and gender education gap (1983 – 1987)

	OLS1	IV1	OLS2	IV2	OLS1	IV1	OLS2	IV2
Dependent Variable	Years of education (1983-87)				Gender Education Gap (1983-87)			
Empl liberalized sectors (1987-99)	0.048	-0.276	-0.008	-0.135	0.017	-0.011	-0.037	-0.025
	(0.094)	(0.232)	(0.092)	(0.207)	(0.039)	(0.125)	(0.065)	(0.162)
Demographics	YES	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	YES	YES	NO	NO	YES	YES
Observations	26	26	26	26	26	26	26	26
R-squared	0.927		0.950		0.549		0.583	

Note: robust standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

iii. Largest cities

The goal of this robustness test is to examine whether our baseline results are driven by districts which include the largest cities, for example because of internal migration towards them. Thanks to both the existing literature and the migration control variable being always insignificant, we have already provided ample evidence that migration across districts is not a concern. Nevertheless, our findings may still be affected by a few outlier districts for reasons other than migration. We thus test whether dropping these districts leads to major differences in our baseline results. We take population information from both 2001 and 2011 censuses and in Table 6 report the top 20 Indian cities ranked according to the 2001 information.

Table 6. Largest Indian cities by 2001 and 2011 population censuses

Rank	City	Population (2001)	Population (2011)
1	Mumbai	11,978,450	12,442,373
2	Delhi	9,879,172	11,007,835
3	Kolkata	4,572,876	4,486,679
4	Chennai	4,343,645	4,681,087
5	Bangalore	4,301,326	8,436,675
6	Hyderabad	3,637,483	6,809,970
7	Ahmedabad	3,520,085	5,570,585
8	Kanpur	2,551,337	2,767,031
9	Pune	2,538,473	3,115,431
10	Surat	2,433,835	4,467,797
11	Jaipur	2,322,575	3,046,163
12	Lucknow	2,185,927	2,815,601
13	Nagpur	2,052,066	2,405,665
14	Indore	1,474,968	1,960,631
15	Bhopal	1,437,354	1,795,648
16	Ludhiana	1,398,467	1,618,879
17	Patna	1,366,444	1,684,222
18	Visakhapatnam	1,345,938	1,730,320
19	Vadodara	1,306,227	1,670,806
20	Agra	1,275,134	1,585,704

We then re-run all our OLS and IV baseline (1987-1999) specifications, for both outcome variables, dropping from the sample districts with (i) the 7 largest cities, with 2001 population above 3 million and 2011 population above 5 million, and (ii) the 13 largest cities, with both 2001 and 2011 population above 2 million. The full results are available in Appendix A4: our main coefficient of interest is remarkably similar in both magnitude and significance across the three samples and for each specification. As a result, we can conclude that these districts are not outliers driving the results.

d. Heterogeneity

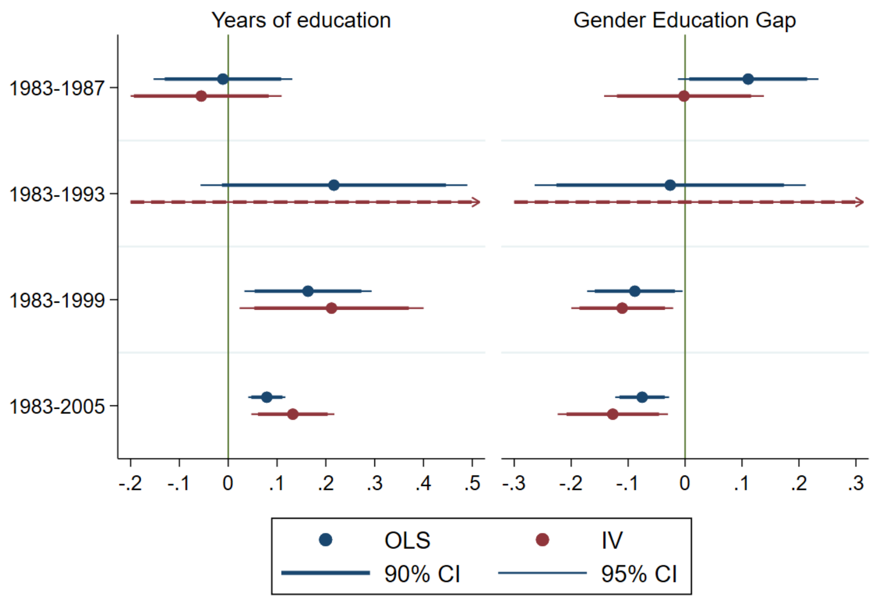
i. Across time: the role of liberalization

To see whether the measured relationship was already present before the liberalization wave , we test our main specifications in periods preceding the one under study. We do this to better understand the mechanisms at issue and what the liberalization context implies for the external validity of our results. That is, we want to understand if liberalization is a precondition for the positive effects we have found given that it spatially directs resources to their most productive use and prevents distortions that might hinder positive educational effects. For instance, if employment growth occurs in districts that do not have the demographics to capitalize on the incentives provided by services employment growth, liberalization is a pre-requirement for positive spillovers from structural change to education and gender equality. This information is important for our findings to better inform the literature on structural change and services-led development.

Again, we take as base year 1983 instead of 1987 and test whether: (i) the effect of increases in the employment shares of liberalized sectors on the change in average education and gender education gap is confirmed for the 1983-1999 and 1983-2005 periods, despite the small sample size and different base year; (ii) there is no significant effect between 1983 and 1987, the actual pre-liberalization test; (iii) there is any effect between 1983 and 1993. On this latter wave we are agnostic since, as already mentioned, it is a year exactly in the midst of the liberalization wave: on the one hand, it is likely that no impact could have already taken place on human capital accumulation; on the other hand, there may have been anticipation effects. Our instrumental variable for all specifications becomes the 1983 state level share of workers in liberalized services sectors, and we report both OLS and IV results to see whether conclusions differ with this new instrument.

Figure 5 reports coefficients and confidence intervals for the main independent variable of interest in each time period. The baseline results are confirmed, with both OLS and IV coefficients statistically significant at least at the 95% confidence level, for both the 1983-1999 and the 1983-2005 periods: as expected, the sign is positive for years of education as outcome variable and negative for the gender education gap. Furthermore, for the 1983-1987 period, the estimates are not significant, confirming that there are no effects before the 1990s. Interestingly, if anything, the point estimates have the opposite sign and the positive OLS coefficient for the gender education gap is actually significant at the 10% confidence level. Also, the 1983-1993 estimates are not statistically significant, but the OLS point estimates start having the expected sign, while the IV estimates are imprecise and are not even reported in the chart.

Figure 5. The role of liberalization: effect of main independent variable across time periods



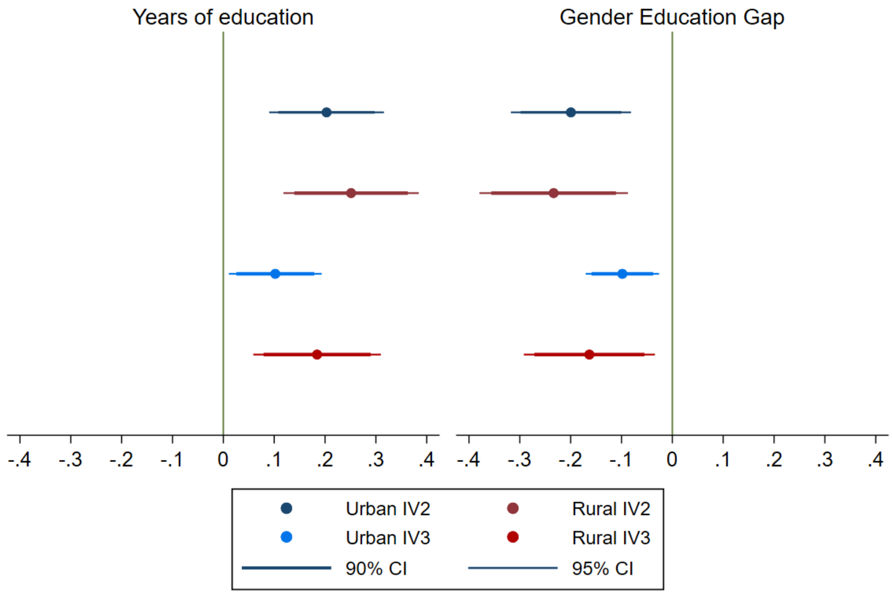
Note: the IV point estimates for the 1983-1993 period are 1.09 (for years of education) and 0.93 (for the gender education gap) and, because of the very imprecise estimation, are not reported in the chart to preserve its scale.

In sum, the effects of services employment expansion on education and its gender gap are attributable to and conditional on their liberalization, since they are detected only after 1993 and such dynamics were not present before then. This suggests that countries that want to benefit from the positive effects of services employment growth on education need to ensure that there are no barriers in place that distort the spatial allocation of services employment.

ii. Across space: urban vs rural

In this subsection we test whether our findings are mostly driven by urban or rural districts. As districts are usually composed by a mix of households in rural and urban areas, we split the sample of districts depending on the 1987 median of the share of households living in a rural area: districts with a proportion of rural households above 75% in 1987 are defined as rural, the others as urban districts. We then perform the same baseline regressions on these two subsamples. Figure 6 shows the coefficients and associated confidence intervals (at the 90% and 95% confidence levels) on our main independent variable for the urban-rural subsample regressions, reported for both IV2 (full set of covariates, except tariffs) and IV3 (with tariff covariates) specifications and for both outcome variables.

Figure 6. Baseline IV results by urban-rural subsamples



We find that: (i) all eight coefficients are statistically significant at least at the 95% confidence level; (ii) the point estimates for the rural subsample are slightly larger in absolute terms than those of the urban subsample; (iii) these differences are not statistically significant. In conclusion, urban and rural households on average enjoy similar benefits from services liberalization, both in terms of higher human capital accumulation and in terms of lower gender disparity in education.

iii. Across industries: disentangling sectors' effects

When taken together, our results for the three liberalized services sectors might suggest an equal contribution of banking, insurance and telecommunications to the increase in education and decrease in the gender education gap. In order to unpack these results, we run the same baseline IV regressions (for the 1987-1999 period) but separate the three services sectors to test whether their impact differs. In the first stage, we thus regress the district employment share of each sector in differences on all three instrumental variables, namely the levels of pre-reform employment shares in the banking, insurance and telecommunications sectors, with the full set of covariates. As expected, the strongest IV for each sector's employment share is the same sector's baseline share in levels.

Table 7 reports the usual set of IV regressions, gradually increasing the number of covariates, as well as the first stage results of the full specification. The results show that it is the banking sector driving both the years of education and gender education gap results, with the latter being affected also by liberalization in the telecommunications sector, although to a lower extent. This is not surprising since these two sectors were heavily liberalized in this timeframe; their liberalization started earlier than that of the insurance sector and they are also significantly larger in terms of employment.

Table 7. Subsectors heterogeneity. DV: years of education and gender education gap (1987 – 1999)

	IV 1	IV2	IV 3	IV 1	IV 2	IV 3	FS 3B	FS 3I	FS 3T
Dependent Variable	Years of education			Gender Education Gap			Empl Banking	Empl Insurance	Empl Telecom
Empl Banking	0.314***	0.240***	0.131***	-0.257***	-0.206***	-0.084**			
	(0.070)	(0.054)	(0.049)	(0.080)	(0.060)	(0.041)			
	[0.360]	[0.277]	[0.152]	[-0.408]	[-0.322]	[-0.132]			
Empl Insurance	0.439	0.209	0.050	-0.246	-0.153	0.008			
	(0.270)	(0.276)	(0.236)	(0.220)	(0.212)	(0.156)			
	[0.143]	[0.067]	[0.016]	[-0.111]	[-0.067]	[0.003]			
Empl Telecom	0.301	0.250	0.043	-0.471**	-0.412**	-0.194			
	(0.203)	(0.182)	(0.146)	(0.204)	(0.172)	(0.120)			
	[0.210]	[0.172]	[0.030]	[-0.457]	[-0.385]	[-0.182]			
IV: 1987 empl share Banking							-0.976***	0.025	0.006
							(0.100)	(0.030)	(0.054)
							[-0.710]	[0.065]	[0.008]
IV: 1987 empl share Insurance							-0.420*	-0.844***	-0.045
							(0.220)	(0.102)	(0.223)
							[-0.068]	[-0.488]	[-0.012]
IV: 1987 empl share Telecom							0.076	0.022	-0.692***
							(0.126)	(0.046)	(0.130)
							[0.025]	[0.026]	[-0.388]
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	YES	YES	NO	YES	YES	NO	YES	YES
Tariffs covariates	NO	NO	YES	NO	NO	YES	NO	NO	YES
Observations	391	373	371	391	373	371	371	371	371
R-squared							0.385	0.263	0.214

Note: robust standard errors are in parentheses, beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1

In fact, the late liberalization of the insurance sector allows us to consider the results for this sector as a placebo test as the sector has the same characteristics as the other two sectors but was subject to the policy shock only at the very end of the examined period. Results for the 1987-2005 period, which encompasses several years after the liberalization of the insurance sector, reveal patterns of significant contributions also from this sector (not reported). The results underscore the importance of removing distortions from services sectors prior to employment growth in line with our findings on the role of liberalization in section V.d.i.

Finally, the fact that the banking sector drives our results may point towards the importance of channels other than the direct employment one. Banking sector is special thanks to its systemic role in the economy at large. Its expansion and increased productivity may benefit all other sectors and thus impact education outcomes also indirectly. We explore this more in details in the next section by looking at the channels driving our results.

VI. Mechanisms

We study the household earnings and the returns to education (RtE) channels as proposed by the literature. According to the earnings channel, if household income increases then the time children spend at school is likely to increase, given the lower need for them to contribute to the household income. As already explained, household earnings may increase thanks to the liberalization of services both directly, if members of the household get employed in such sectors, and indirectly, if the liberalization wave brings about general improvements in productivity and a reduction of operational costs in other sectors as well.

On a different note, the increased presence of high-paying jobs that require higher levels of education raises the overall returns to schooling and thus the likelihood that an individual invests in an additional year of education. This channel may work regardless of whether the individual will actually end up working in one of the liberalized services sectors or in another high-skill job (e.g. Public Administration, IT or professional services). In both cases, thanks to the increased education levels in a district, a higher skilled labor force will attract more firms demanding skilled labor and paying a wage premium, generating a positive feedback loop and reinforcing these mechanisms.

The way we test for the household earnings channel in Table 8 follows a three stage least squares (3SLS) approach. We regress each outcome variable (district average years of education on the left side of the panel and district gender education gap on the right side) on the logarithm of district average household earnings and the full set of covariates as in our preferred specification (see the first column of each panel). The earnings variable is instrumented with each liberalized services sector employment share (second column of each panel) and the latter are in turn instrumented with three employment shares in levels at baseline, as in the previous section. Given the double instrumentation, we report bootstrapped standard errors (with 200 replications). If the earnings variable is significant it means that the main effect is likely to work via this channel and significance of each employment share will indicate through which sector specifically.

As an additional robustness check, we also test the existence of each channel by using the *ivmediate* command in STATA. This approach to studying mechanisms is based on mediation analysis, which estimates both the overall effect of an instrumented treatment on an outcome variable and its two components: the indirect effect of the treatment through a mediator (channel) variable, also instrumented, and the direct effect after taking into account the role of the mediator. Unfortunately, it is less flexible than the 3SLS approach, since it does not allow for more than one instrumental variable and for more than one mediator at a time. As a result, with this methodology, we revert to the use of a single treatment and a single instrumental variable, i.e. the three liberalized sectors together. Nevertheless, it is informative to corroborate the main results obtained through the 3SLS methodology. For each mediator, we report in the Appendix a table with four columns: two for each outcome variable, both with and without our tariff covariates.

Table 8 suggests that, for both outcome variables, the baseline results work through the household earnings channel, which is always significant and with the expected signs: an increase in a district's earnings, driven by the change in employment in the three services sectors, raises the average years of education and lowers the gender education gap. In particular, the banking sector seems to be the main driver for the average education findings, while for the gender education gap the telecommunications sector brings a large contribution as well. Note that these results are identified only on the variance of earnings that impacts each dependent variable through the three instrumented services sectors employment shares.

Table 8. Household earnings channel. DV: years of education and gender education gap (1987 – 1999)

	Years of education					Gender education gap				
Dependent Variable	Education	Earnings	Banking	Insurance	Telecom	Gender gap	Earnings	Banking	Insurance	Telecom
Household earnings (log)	1.169**					-1.090**				
	(0.540)					(0.497)				
	[0.779]					[-1.039]				
Empl Banking		0.104***					0.179**			
		(0.036)					(0.090)			
		[0.188]					[0.192]			
Empl Insurance		0.046					0.063			
		(0.141)					(0.143)			
		[0.023]					[0.031]			
Empl Telecom		0.115					0.084**			
		(0.089)					(0.036)			
		[0.123]					[0.153]			
IV: 1987 empl share Banking			-1.000***	0.024	0.002			-0.998***	0.022	0.002
			(0.074)	(0.024)	(0.050)			(0.074)	(0.024)	(0.050)
			[-0.723]	[0.064]	[0.002]			[-0.722]	[0.057]	[0.003]
IV: 1987 empl share Insurance			-0.423	-0.847***	-0.044			-0.424	-0.842***	-0.041
			(0.272)	(0.087)	(0.186)			(0.272)	(0.087)	(0.186)
			[-0.068]	[-0.489]	[-0.012]			[-0.068]	[-0.486]	[-0.011]
IV: 1987 empl share Telecom			0.048	0.013	-0.698***			0.044	0.018	-0.700***
			(0.134)	(0.043)	(0.092)			(0.134)	(0.043)	(0.092)
			[0.016]	[0.016]	[-0.392]			[0.015]	[0.021]	[-0.393]
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
District-level covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Tariffs covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	364	364	364	364	364	364	364	364	364	364
R-squared			0.471	0.292	0.300			0.471	0.293	0.300

Note: 3SLS (three-stage least squares) regressions. Bootstrap standard errors (200 reps) are in parentheses, beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1

The mediation analysis confirms the presence of the household earnings channel. As shown in the first table of Appendix A5, most (that is, more than 80%) of the total effect of our treatment on both outcome variables takes place via the mediator. While the direct effect is very small and never significant, the indirect effect through the household earnings channel is large and highly significant when we exclude the tariffs covariates. Including them turns insignificant the indirect effect, but not the total effect. The Kleinbergen-Paap F-statistics for the first stages of the treatment on the instrument are very high as usual.

In Table 9 we employ again the 3SLS technique to test for the presence of the returns to education channel, using the same double instrumentation procedure. We obtain a measure of the district's average returns to schooling, separately for each year, by running a classic individual-level Mincerian regression of the logarithm of earnings on the years of schooling interacted with a district identifier, and other covariates[11]. We then extract the coefficient on each schooling-district interaction and sum it to the year-specific coefficient on the years of education variable common to all districts.

Looking at the first column of each panel, we notice that the returns to education channel is also present, positively (negatively) affecting the districts' average education (gender education gap) outcome variable in

[11] The other covariates include experience, experience squared, whether the individual is female, married, Hindus, belongs to a backward class, lives in a rural area and his/her household size.

a significant way. Again, as shown by the second column of each panel, the only services subsectors whose increase in employment share significantly affect this channel are the banking and telecommunications ones, but in this case both of them are relevant for both dependent variables. The existence of the returns to education channel is also confirmed by mediation analysis, as shown in the second table of Appendix A5.

Table 9. Returns to education channel. DV: years of education and gender education gap (1987 – 1999)

	Years of education					Gender education gap				
Dependent Variable	Education	Returns educ	Banking	Insurance	Telecom	Gender gap	Returns educ	Banking	Insurance	Telecom
Returns to education	0.097*					-0.123**				
	(0.058)					(0.062)				
	[0.517]					[-0.939]				
Empl Banking		1.647*					0.588*			
		(0.962)					(0.310)			
		[0.222]					[0.133]			
Empl Insurance		0.182					0.264			
		(1.478)					(1.304)			
		[0.011]					[0.017]			
Empl Telecom		0.770**					1.928**			
		(0.339)					(0.937)			
		[0.175]					[0.259]			
IV: 1987 empl share Banking			-1.001***	0.024	0.003			-1.002***	0.020	0.000
			(0.074)	(0.024)	(0.051)			(0.074)	(0.024)	(0.051)
			[-0.724]	[0.063]	[0.004]			[-0.725]	[0.053]	[0.000]
IV: 1987 empl share Insurance			-0.431	-0.843***	-0.046			-0.431	-0.842***	-0.045
			(0.273)	(0.087)	(0.186)			(0.273)	(0.088)	(0.186)
			[-0.069]	[-0.487]	[-0.012]			[-0.069]	[-0.486]	[-0.012]
IV: 1987 empl share Telecom			0.040	0.014	-0.701***			0.041	0.018	-0.698***
			(0.135)	(0.043)	(0.092)			(0.135)	(0.043)	(0.092)
			[0.013]	[0.016]	[-0.394]			[0.014]	[0.022]	[-0.392]
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
District-level covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Tariffs covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	362	362	362	362	362	362	362	362	362	362
R-squared			0.472	0.293	0.299			0.472	0.293	0.299

Note: 3SLS (three-stage least squares) regressions. Bootstrap standard errors (200 reps) are in parentheses, beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1

Although the two sets of results in the previous subsection designate both channels as working mechanisms for our main results, the two are correlated. To understand which channel is stronger, we thus run the same set of three stage least squares regressions with both channels together as independent variables and the three (instrumented) employment shares of liberalized services sectors as IVs for both, again using bootstrapped standard errors. Having three "instruments" (the employment shares) and two endogenous covariates (the household earnings and returns to education channels), the order condition for IV regressions is respected. Table 10 reports the results for years of education and Table 11 for the gender education gap outcome variables.

The household earnings channel clearly dominates the returns to education one, which turns insignificant once we consider the two channels together. This is true for both outcome variables. The effect is also remarkably strong: an increase in average household earnings by one standard deviation raises districts' average education by more than three quarters of a standard deviation, and lowers the gender education gap by about the same fraction of a standard deviation. Looking at column 2 in both Tables 10 and 11, we note again that banking and telecommunications services are driving these results for the household earnings channel, although the latter with a lower level of statistical significance.

Table 10. Household earnings and returns to education channels. DV: years of education (1987 – 1999)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Education	Earnings	Returns educ	Banking	Insurance	Telecom
Household earnings (log)	1.287***					
	(0.494)					
	[0.787]					
Returns to education	-0.123					
	(0.128)					
	[-0.649]					
Empl Banking		0.200***	0.230			
		(0.055)	(0.439)			
		[0.392]	[0.052]			
Empl Insurance		0.195	0.830			
		(0.226)	(2.657)			
		[0.106]	[0.052]			
Empl Telecom		0.367*	2.104			
		(0.198)	(1.765)			
		[0.426]	[0.283]			
IV: 1987 empl share Banking				-0.828***	0.048	0.054
				(0.078)	(0.030)	(0.059)
				[-0.599]	[0.125]	[0.066]
IV: 1987 empl share Insurance				-0.197	-0.815***	0.027
				(0.302)	(0.105)	(0.252)
				[-0.032]	[-0.470]	[0.007]
IV: 1987 empl share Telecom				0.226	0.045	-0.625***
				(0.148)	(0.049)	(0.131)
				[0.075]	[0.054]	[-0.351]
State FE	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES
District-level covariates	YES	YES	YES	YES	YES	YES
Tariffs covariates	YES	YES	YES	YES	YES	YES
Observations	364	364	364	364	364	364
R-squared				0.428	0.277	0.279

Note: three-stage least squares regressions. Bootstrap standard errors (200 reps) are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

Table 11. Household earnings and returns to education channels. DV: gender education gap (1987 – 1999)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Gender gap	Earnings	Returns educ	Banking	Insurance	Telecom
Household earnings (log)	-0.860**					
	(0.393)					
	[-0.757]					
Returns to education	-0.033					
	(0.096)					
	[-0.251]					
Empl Banking		0.200***	0.269			
		(0.059)	(0.453)			
		[0.391]	[0.061]			
Empl Insurance		0.202	0.239			
		(0.252)	(2.357)			
		[0.110]	[0.015]			
Empl Telecom		0.367**	2.081			
		(0.171)	(1.742)			
		[0.427]	[0.280]			
IV: 1987 empl share Banking				-0.829***	0.047	0.053
				(0.093)	(0.031)	(0.053)
				[-0.599]	[0.123]	[0.064]
IV: 1987 empl share Insurance				-0.192	-0.808***	0.035
				(0.299)	(0.105)	(0.250)
				[-0.031]	[-0.467]	[0.010]
IV: 1987 empl share Telecom				0.226	0.045	-0.625***
				(0.151)	(0.054)	(0.149)
				[0.075]	[0.054]	[-0.351]
State FE	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES
District-level covariates	YES	YES	YES	YES	YES	YES
Tariffs covariates	YES	YES	YES	YES	YES	YES
Observations	364	364	364	364	364	364
R-squared				0.428	0.277	0.279

Note: three-stage least squares regressions. Bootstrap standard errors (200 reps) are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

In order to further explore our main mechanism, household earnings, we also test whether it mostly works through the mothers' or fathers' earnings, possibly generating different dynamics within the household. There is indeed evidence from different settings that mothers value children's - and particularly girls' - education more than fathers (Thomas, 1994; Qian, 2008; Edmonds, Pavcnik and Topalova, 2010). As a result, if mothers gain more bargaining power, at the margin they are likely to improve girls' schooling more than boys' because boys already had priority in getting schooling, thereby decreasing the gender education gap.

We thus run a set of three stage least squares regressions in a setup similar to that of the last subsection, with district-level average mothers' and fathers' earnings as the channels to be tested. The only difference is that here we do not include the merchandise trade liberalization measures. Including them turns insignificant both channels and the merchandise trade liberalization variables themselves, which is likely due to the further slicing of the earnings data whose availability is already scant in our dataset. These results should thus be taken as merely indicative of a potential link between our main channel and the existing literature.

Table 12 shows the 3SLS results with the gender education gap as outcome variable, for which we find a significant difference between the parents' earnings, while for years of education there is no such evidence (not reported). What we find is in fact in line with the literature mentioned above: mothers' earnings contribute more than fathers' earnings to reducing the gender education gap, thereby pushing girls' human capital accumulation relatively more than boys', although this is only marginally significant.

Table 12. Mothers' and fathers' earnings channels. DV: gender education gap (1987 – 1999)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Gender gap	Earnings mothers	Earnings fathers	Banking	Insurance	Telecom
Earnings mothers (log)	-1.246*					
	(0.752)					
Earnings fathers (log)	0.120					
	(0.903)					
Empl Banking		0.156**	0.072			
		(0.061)	(0.053)			
Empl Insurance		0.069	0.423**			
		(0.269)	(0.214)			
Empl Telecom		0.352**	0.323**			
		(0.171)	(0.153)			
IV: 1987 empl share Banking				-0.789***	0.063***	0.051
				(0.062)	(0.020)	(0.051)
IV: 1987 empl share Insurance				-0.288	-0.794***	0.050
				(0.232)	(0.076)	(0.191)
IV: 1987 empl share Telecom				0.198*	0.033	-0.563***
				(0.120)	(0.039)	(0.098)
State FE	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES
District-level covariates	YES	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	NO
Observations	337	337	337	337	337	337
R-squared				0.455	0.321	0.263

Note: 3SLS (three-stage least squares) regressions. Robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

VII. Conclusion

The services sector has become dominant in India's economy since the late 1990s and this paper suggests that its liberalization has been key to raising the country's human capital accumulation and related gender parity. We employ a local labor market perspective and study the effects of the 1990s liberalization wave in the banking, insurance and telecommunications subsectors on education and the gender education gap at the district level. Our baseline regressions are run over the 1987-1999 timeframe, but we also test the long-term effects from 1987 to 2005. We employ all variables in differences to eliminate district-specific time-invariant effects and always include state fixed effects as well as different sets of covariates, from demographics to district-level characteristics to variables that capture the contemporaneous merchandise trade liberalization. To prevent other potential sources of bias we also employ an instrumental variable strategy, which is robust to an alternative instrument definition.

We find that the employment share of liberalized services sectors is a consistently significant determinant of both the average number of years of education (positively) and the gender education gap (negatively) at the district level. We estimate that an increment by one standard deviation of the share of workers in our three key services sectors - banking, insurance and telecommunications - increases (decreases) the district's average years of schooling (gender education gap) by about one sixth (one fifth) of a standard deviation. This is equivalent to 4.4% (8.3%) of the average change of the dependent variable reported during the 1987-1999 period. The economic and statistical significance of these findings hold using both OLS and IV procedures in first differences, with all sets of covariates, both in the medium (12 years lag) and in the longer term (18 years lag).

We rule out the possibility that the contemporaneous expansion of the education offer or of other services sectors drive these results. Moreover, our findings are quite similar between predominantly rural and urban districts, remaining robust after splitting the sample and when dropping districts with the largest Indian cities. We also do not detect any presence of pre-trends and we show that the beneficial effects of increased employment in the three subsectors materialize only after their liberalization. Interestingly, these effects are at least as relevant as those of merchandise trade liberalization, which received most of the attention in the literature so far. This leads us to the conclusion that liberalizing key services sectors is indeed a crucial policy to boost human capital accumulation and lower the gender education gap in a developing country like India. Our results thus highlight a channel, that has so far received little attention, through which services liberalization can boost economic development.

The sectors that particularly contribute to our findings are banking and telecommunication services, likely through both their direct (employment) and indirect effects (higher productivity in downstream sectors as well as lower operational costs thanks to greater competition, lower prices and better quality of the service offer). In fact, although both the returns to education and the household earnings channels seem to be at work, the latter is the main driver of the results. Mothers' earnings seem to be particularly effective in reducing the gender education gap, a result in line with the literature showing mothers' preference for girls' education. There is scope for future research to study more in detail the two main hypothesized channels, as well as the transmission mechanism underlying the increase in skilled labor supply.

References

1. Aghion, P., Burgess, R., Redding, S. J., & Zilibotti, F. (2008). The unequal effects of liberalization: Evidence from dismantling the License Raj in India. *American Economic Review*, *98*(4), 1397-1412.
2. Arnold, J. M., Javorcik, B., Lipscomb, M., & Mattoo, A. (2015). Services reform and manufacturing performance: Evidence from India. *The Economic Journal*, *126*(590), 1-39.
3. Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in Mexico. *American Economic Review*, *106*(8), 2046-85.
4. Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, *103*(6), 2121-68.
5. Autor, D. H., Dorn, D., & Hanson, G .H. (2015). Untangling trade and technology: evidence from local labour markets. *Economic Journal*, 125, 621–646.
6. Baldwin, R., & Forslid, R. (2020). *Globotics and development: When manufacturing is jobless and services are tradable* (No. w26731). National Bureau of Economic Research.
7. Basu, K. (2015). *An economist in the real world: the art of policymaking in India*. MIT Press.
8. Becker, G. S. (1957). The economics of discrimination Chicago. *University of Chicago*.
9. Burgess, R. & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *The American Economic Review*, 95(3), 780-795.
10. Cole, S. (2009). Financial Development, Bank Ownership, and Growth: Or, Does Quantity Imply Quality? *The Review of Economics and Statistics*, 91(1), 33-51.
11. Duflo, E., & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, *122*(2), 601-646.
12. Edmonds, E. V., Pavcnik, N., & Topalova, P. (2010). Trade adjustment and human capital investments: Evidence from Indian tariff reform. *American Economic Journal: Applied Economics*, *2*(4), 42-75.
13. Galor, O., & Weil, D. N. (1996). The gender gap, fertility, and growth. *American Economic Review*, *86*(3), 374.
14. Greenland, A., & Lopresti, J. (2016). Import exposure and human capital adjustment: Evidence from the US. *Journal of International Economics*, *100*, 50-60.
15. Hallward-Driemeier, M., & Nayyar, G. (2017). *Trouble in the Making? The Future of Manufacturing-Led Development*. World Bank Publications, Washington D.C.
16. Harrison, A. E., Martin, L. A., & Nataraj, S. (2012). Learning versus stealing: how important are market-share reallocations to India's productivity growth?. *The World Bank Economic Review*, *27*(2), 202-228.
17. Heath, R., & Mobarak, A. M. (2015). Manufacturing growth and the lives of Bangladeshi women. *Journal of Development Economics*, *115*, 1-15.
18. Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *The Quarterly Journal of Economics*, *127*(2), 753-792.
19. Juhn, C., Ujhelyi, G., & Villegas-Sanchez, C. (2014). Men, women, and machines: How trade impacts gender inequality. *Journal of Development Economics*, *106*, 179-193.

20. McCaig, B. (2011). Exporting out of poverty: Provincial poverty in Vietnam and US market access. *Journal of International Economics*, *85*(1), 102-113.
21. Mehta, A., & Hasan, R. (2012). The effects of trade and services liberalization on wage inequality in India. *International Review of Economics & Finance*, *23*, 75-90.
22. Menon, N., & Van der Meulen Rodgers, Y. (2009). International trade and the gender wage gap: New evidence from India's manufacturing sector. *World Development*, *37*(5), 965-981.
23. Munshi, K., & Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, *96*(4), 1225-1252.
24. Nayyar, G. (2012). The quality of employment in India's services sector: exploring the heterogeneity. *Applied Economics*, 44(36), 4701-4719.
25. Ngai, L. R., & Petrongolo, B. (2017). Gender Gaps and the Rise of the Service Economy. *American Economic Journal: Macroeconomics*, 9 (4): 1-44.
26. Oster, E., & Steinberg, B. M. (2013). Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics*, *104*, 123-135.
27. Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2012). Human capital investment and the gender division of labor in a brawn-based economy. *American Economic Review*, *102*(7), 3531-60.
28. Prashanth, G. N. (2013). The brick 'n' mortar story of Silicon City. Deccan Herald. https://www.deccanherald.com/content/364708/brick-n-mortar-story-silicon.html. Accessed on 20/01/2021.
29. Qian, N. (2008). Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, *123*(3), 1251-1285.
30. Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth*, 21(1), 1-33.
31. Rosenzweig, M. R., & Evenson, R. (1977). Fertility, schooling, and the economic contribution of children of rural India: An econometric analysis. *Econometrica,* 1065-1079.
32. Shastry, G. K. (2012). Human capital response to globalization education and information technology in India. *Journal of Human Resources*, *47*(2), 287-330.
33. Sauré, P., & Zoabi, H. (2014). International trade, the gender wage gap and female labor force participation. *Journal of Development Economics*, *111*, 17-33.
34. Thomas, D. (1994). Like father, like son; like mother, like daughter: Parental resources and child height. *Journal of Human Resources*, 950-988.
35. Topalova, P. (2007). Trade liberalization, poverty and inequality: Evidence from Indian districts. In *Globalization and Poverty* (pp. 291-336). University of Chicago Press.
36. Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics*, *2*(4), 1-41.
37. Young, A. (2019). Consistency without Inference: Instrumental Variables in Practical Applications. London School of Economics.

Appendix

A1. Instrument validity: robustness checks

Although Section IV explains why our instrumental variable should not cause endogeneity concerns, we further probe its validity by exploiting the 1983 NSS wave and constructing an alternative instrument. With respect to the liberalization period this wave is more behind in time, thereby even less prone to endogeneity concerns. However, it does not have a district identifier and the smallest geographical unit available is the state. To construct an IV that has variation at the district level, we then interact the proportion of workers in liberalized services sectors at the state level in 1983 with the district population in 1987:

$$IV_b = S_{s,1983} * P_{d,1987}$$

We first try employing this new instrumental variable ("IV_b") in substitution to the original one ("IV_a") and see whether our results hold. We find that the coefficients have the same sign and similar magnitude, but they are only significant with the demographic covariates set and at the 10% level. When we add the district-level covariates they are no more statistically significant, which is due to the low IV strength: also the first stage becomes insignificant. Expectedly, there is not a significant relationship between the change over time of a district phenomenon and its state counterpart in levels.

DV: years of education and gender education gap (1987-1999)

	(3)	(6)	(8)	(3)	(6)	(8)	(4)	(9)
	IV1	IV2	IV3	IV1	IV2	IV3	FS1	FS3
Dependent Variable		Education			Gender Education Gap		Empl liberalized sectors	
Empl liberalized sectors	0.436*	0.347	0.052	-0.419*	-0.482	-0.155		
	(0.246)	(0.326)	(0.249)	(0.219)	(0.377)	(0.228)		
IV: 1983_liber share*1987_popul							-0.014***	-0.009*
							(0.004)	(0.005)
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	YES	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	YES	YES	NO	NO	YES	NO	YES
Tariffs covariates	NO	NO	YES	NO	NO	NO	NO	YES
Observations	391	373	371	391	373	371	391	371

Note: standard errors clustered at the state level are in parentheses, beta coefficients in brackets.

We nevertheless use IV_b to test for the validity of the original IV_a, using the overidentifying restrictions test in its Wooldridge heteroskedasticity-robust version. We use six alternative formulations of IV_b: we interact the 1983 state-level share of people working in liberalized services sectors with either the 1987 district-level population or labor force, including or excluding migrants across states or districts. The table below reports the p-values of the test for specifications with different sets of controls, outcome variables and time spans: in all cases they are well above the standard 0.05 threshold. This implies that there is no sign of endogeneity for our original IV, thereby reinforcing our identification strategy.

Overidentifying restrictions test (Wooldridge heteroskedasticity-robust), p-value

IVa	IVb	1st (Educ_87-99)	2nd (Educ_87-05)	3rd (Educ_87-99)	3rd (GEG_87-99)
		Set of controls			
1987d_liber share	1983s_liber share*1987d_population	0.890	0.871	0.359	0.545
	1983s_liber share*1987d_population (excl migrants across states)	0.956	0.945	0.306	0.559
	1983s_liber share*1987d_population (excl migrants across districts)	0.730	0.729	0.239	0.680
	1983s_liber share*1987d_labor force	0.998	0.936	0.382	0.746
	1983s_liber share*1987d_labor force (excl migrants across states)	0.748	0.687	0.292	0.785
	1983s_liber share*1987d_labor force (excl migrants across districts)	0.473	0.447	0.210	0.891

A2. Full baseline regression results - years of education (1987-1999)

DV: years of education (1987-1999)

	OLS0	OLS1	IV1	OLS2	IV2	OLS3	IV3	FS3
Dependent Variable	Education	Education	Education	Education	Education	Education	Education	Empl liberalized sectors
Empl liberalized sectors	0.167***	0.100***	0.323***	0.083***	0.241***	0.067***	0.112**	
	(0.034)	(0.028)	(0.069)	(0.026)	(0.054)	(0.024)	(0.045)	
	[0.242]	[0.149]	[0.479]	[0.122]	[0.353]	[0.099]	[0.166]	
IV: 1987 empl share liber								-0.890***
								(0.108)
								[-0.667]
Female		-2.316	-2.382	-3.039*	-3.161*	-3.638**	-3.651**	3.417
		(1.810)	(1.741)	(1.689)	(1.657)	(1.551)	(1.542)	(2.417)
		[-0.069]	[-0.071]	[-0.091]	[-0.094]	[-0.106]	[-0.107]	[0.068]
Age		0.174***	0.166***	0.186***	0.178***	0.162***	0.160***	0.059
		(0.033)	(0.033)	(0.032)	(0.032)	(0.035)	(0.035)	(0.039)
		[0.358]	[0.340]	[0.389]	[0.370]	[0.332]	[0.328]	[0.081]
Married		-2.304*	-2.786**	-2.511**	-2.875**	-1.335	-1.488	1.802
		(1.258)	(1.324)	(1.169)	(1.227)	(1.107)	(1.127)	(2.258)
		[-0.122]	[-0.148]	[-0.135]	[-0.155]	[-0.070]	[-0.078]	[0.064]
Backward class		-0.493	-0.372	-0.382	-0.314	-0.312	-0.296	-0.885
		(0.342)	(0.402)	(0.307)	(0.345)	(0.277)	(0.283)	(0.708)
		[-0.061]	[-0.046]	[-0.049]	[-0.040]	[-0.040]	[-0.038]	[-0.077]
Hindus		0.632**	0.738**	0.596*	0.704**	0.335	0.375	-0.162
		(0.310)	(0.345)	(0.332)	(0.348)	(0.331)	(0.335)	(0.519)
		[0.081]	[0.094]	[0.074]	[0.088]	[0.042]	[0.047]	[-0.014]
Household size		0.089*	0.085	0.111**	0.110**	0.124***	0.122**	-0.071
		(0.054)	(0.055)	(0.051)	(0.052)	(0.048)	(0.048)	(0.083)
		[0.088]	[0.084]	[0.109]	[0.109]	[0.122]	[0.121]	[-0.047]
Rural		-1.931***	-1.577***	-1.542***	-1.417***	-1.496***	-1.463***	-0.475
		(0.271)	(0.287)	(0.273)	(0.271)	(0.285)	(0.277)	(0.470)
		[-0.401]	[-0.328]	[-0.319]	[-0.293]	[-0.294]	[-0.287]	[-0.063]
Household expenditure		0.832***	0.899***	0.824***	0.894***	0.843***	0.863***	-0.384
		(0.195)	(0.227)	(0.186)	(0.206)	(0.178)	(0.182)	(0.443)
		[0.211]	[0.228]	[0.206]	[0.224]	[0.212]	[0.217]	[-0.065]
Migrant (incl across districts)		0.518	-0.270	0.403	-0.157	0.443	0.285	2.905**
		(0.596)	(0.737)	(0.556)	(0.627)	(0.530)	(0.537)	(1.153)
		[0.041]	[-0.021]	[0.031]	[-0.012]	[0.034]	[0.022]	[0.151]
Number of schools per capita				0.001	0.001	0.002	0.002	-0.001
				(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
				[0.036]	[0.035]	[0.084]	[0.082]	[-0.023]
Empl other high-skill services				0.021**	0.019**	0.034***	0.033***	0.030***
				(0.008)	(0.008)	(0.010)	(0.010)	(0.012)
				[0.234]	[0.206]	[0.295]	[0.285]	[0.179]
Empl medium-low skill services				0.006**	0.005*	0.002	0.002	0.009*
				(0.003)	(0.003)	(0.002)	(0.003)	(0.005)
				[0.104]	[0.085]	[0.034]	[0.031]	[0.103]
Own tariffs						-0.031	-0.032	0.115***
						(0.023)	(0.023)	(0.036)
						[-0.296]	[-0.299]	[0.732]
Faced tariffs						-0.252	-0.263	0.761*
						(0.352)	(0.349)	(0.449)
						[-0.175]	[-0.182]	[0.358]
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES
Observations	392	391	391	373	373	371	371	371
R-squared	0.071	0.404		0.494		0.542		0.348
Kleibergen-Paap F-stat			29.2		41.5		60.3	60.3

Note: robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

A3. Full baseline regression results - gender education gap (1987-1999)

	DV: gender education gap (1987-1999)							
	OLS0	OLS1	IV1	OLS2	IV2	OLS3	IV3	FS3
Dependent Variable	GEG	GEG	GEG	GEG	GEG	GEG	GEG	Empl liberalized sectors
Empl liberalized sectors	-0.051**	-0.048**	-0.276***	-0.045**	-0.237***	-0.040*	-0.100***	
	(0.025)	(0.023)	(0.067)	(0.023)	(0.056)	(0.022)	(0.038)	
	[-0.104]	[-0.098]	[-0.566]	[-0.091]	[-0.472]	[-0.081]	[-0.199]	
IV: 1987 empl share liber								-0.890***
								(0.108)
								[-0.667]
Female		0.245	0.312	0.700	0.848	0.720	0.737	3.417
		(1.076)	(1.249)	(1.138)	(1.269)	(1.114)	(1.132)	(2.417)
		[0.010]	[0.013]	[0.028]	[0.034]	[0.029]	[0.029]	[0.068]
Age		-0.009	-0.000	-0.021	-0.011	-0.013	-0.011	0.059
		(0.020)	(0.022)	(0.020)	(0.021)	(0.020)	(0.021)	(0.039)
		[-0.027]	[-0.001]	[-0.060]	[-0.030]	[-0.037]	[-0.030]	[0.081]
Married		4.448***	4.940***	4.553***	4.995***	4.206***	4.408***	1.802
		(1.558)	(1.443)	(1.500)	(1.420)	(1.493)	(1.441)	(2.258)
		[0.327]	[0.363]	[0.334]	[0.366]	[0.300]	[0.314]	[0.064]
Backward class		-0.120	-0.244	-0.100	-0.183	-0.156	-0.177	-0.885
		(0.287)	(0.347)	(0.279)	(0.331)	(0.271)	(0.274)	(0.708)
		[-0.021]	[-0.042]	[-0.017]	[-0.032]	[-0.027]	[-0.031]	[-0.077]
Hindus		0.620**	0.512	0.760**	0.629*	0.844***	0.792**	-0.162
		(0.271)	(0.312)	(0.315)	(0.341)	(0.309)	(0.315)	(0.519)
		[0.110]	[0.091]	[0.129]	[0.107]	[0.144]	[0.135]	[-0.014]
Household size		-0.052	-0.048	-0.099**	-0.098**	-0.067*	-0.065	-0.071
		(0.043)	(0.047)	(0.044)	(0.048)	(0.041)	(0.041)	(0.083)
		[-0.072]	[-0.066]	[-0.133]	[-0.132]	[-0.089]	[-0.087]	[-0.047]
Rural		0.648***	0.287	0.518**	0.365	0.662***	0.619**	-0.475
		(0.227)	(0.247)	(0.242)	(0.249)	(0.252)	(0.251)	(0.470)
		[0.187]	[0.083]	[0.146]	[0.103]	[0.176]	[0.165]	[-0.063]
Household expenditure		-0.307	-0.375	-0.364	-0.449*	-0.552**	-0.577**	-0.384
		(0.223)	(0.251)	(0.226)	(0.250)	(0.254)	(0.251)	(0.443)
		[-0.108]	[-0.132]	[-0.124]	[-0.153]	[-0.188]	[-0.196]	[-0.065]
Migrant (incl across districts)		0.967	1.772*	1.020	1.700	0.861	1.068	2.905**
		(0.966)	(1.045)	(0.992)	(1.050)	(0.903)	(0.861)	(1.153)
		[0.106]	[0.195]	[0.106]	[0.176]	[0.090]	[0.111]	[0.151]
Number of schools per capita				-0.002**	-0.002**	-0.003***	-0.002**	-0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
				[-0.121]	[-0.118]	[-0.147]	[-0.144]	[-0.023]
Empl other high-skill services				-0.013***	-0.010***	-0.006	-0.005	0.030***
				(0.004)	(0.003)	(0.006)	(0.006)	(0.012)
				[-0.191]	[-0.145]	[-0.076]	[-0.058]	[0.179]
Empl medium-low skill services				0.002	0.003	0.005**	0.005**	0.009*
				(0.002)	(0.003)	(0.003)	(0.003)	(0.005)
				[0.049]	[0.080]	[0.122]	[0.126]	[0.103]
Own tariffs						0.044***	0.045***	0.115***
						(0.011)	(0.011)	(0.036)
						[0.562]	[0.568]	[0.732]
Faced tariffs						0.292	0.306*	0.761*
						(0.178)	(0.186)	(0.449)
						[0.274]	[0.287]	[0.358]
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES
Observations	392	391	391	373	373	371	371	371
R-squared	0.012	0.140		0.191		0.242		0.348
Kleibergen-Paap F-stat			29.2		41.5		60.3	60.3

Note: robust standard errors are in parentheses, beta coefficients in brackets.
*** p<0.01, ** p<0.05, * p<0.1

A4. District heterogeneity

DV: years of education (1987-1999)

Dependent Variable	OLS 0	OLS 1	IV 1	OLS 2	IV 2	OLS 3	IV 3	FS 3
	Baseline results							
	Years of education							Empl liberalized sectors
Empl liberalized sectors	0.167***	0.100***	0.323***	0.084***	0.234***	0.068***	0.114**	
	(0.034)	(0.028)	(0.069)	(0.026)	(0.053)	(0.024)	(0.045)	
	[0.242]	[0.149]	[0.479]	[0.122]	[0.353]	[0.099]	[0.166]	
IV: 1987 empl share liber								-0.889***
								(0.108)
								[-0.667]
Observations	392	391	391	373	373	371	371	371
R-squared	0.071	0.404		0.498		0.543		0.348
	Without 7 largest cities (population > 3M)							
Empl liberalized sectors	0.168***	0.102***	0.251***	0.087***	0.217***	0.070***	0.110**	
	(0.035)	(0.028)	(0.054)	(0.026)	(0.049)	(0.024)	(0.044)	
	[0.240]	[0.150]	[0.368]	[0.128]	[0.317]	[0.104]	[0.163]	
IV: 1987 empl share liber								-0.904***
								(0.109)
								[-0.663]
Observations	385	384	384	371	371	369	369	369
R-squared	0.070	0.410		0.502		0.544		0.357
	Without 13 largest cities (population > 2M)							
Empl liberalized sectors	0.165***	0.101***	0.241***	0.087***	0.210***	0.071***	0.114**	
	(0.035)	(0.028)	(0.054)	(0.026)	(0.049)	(0.024)	(0.044)	
	[0.237]	[0.148]	[0.353]	[0.128]	[0.307]	[0.105]	[0.168]	
IV: 1987 empl share liber								-0.908***
								(0.113)
								[-0.656]
Observations	379	378	378	365	365	363	363	363
R-squared	0.068	0.408		0.499		0.541		0.357
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES

Note: robust standard errors are in parentheses, beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1

Note: the drop of 7 and 13 observations in column OLS0 reduces to 2 and 8 observations in columns OLS2 and OLS3 because 5 of the districts dropped would have been lost anyway by including the district number of schools control variable. However, dropping this covariate does not lead to any substantial change neither in terms of statistical significance nor in terms of economic magnitude of the estimates.

DV: gender education gap (1987-1999)

	Baseline results							
Dependent Variable	Gender Education Gap							Empl liberalized sectors
	OLS 0	OLS 1	IV 1	OLS 2	IV 2	OLS 3	IV 3	FS 3
Empl liberalized sectors	-0.051**	-0.048**	-0.276***	-0.046**	-0.236***	-0.040*	-0.099***	
	(0.025)	(0.023)	(0.067)	(0.023)	(0.056)	(0.022)	(0.038)	
	[-0.104]	[-0.098]	[-0.566]	[-0.091]	[-0.470]	[-0.080]	[-0.197]	
IV: 1987 empl share liber								-0.890***
								(0.108)
								[-0.667]
Observations	392	391	391	373	373	371	371	371
R-squared	0.012	0.140		0.191		0.244		0.348
	Without 7 largest cities (population > 3M)							
Empl liberalized sectors	-0.052**	-0.053**	-0.242***	-0.047**	-0.224***	-0.041*	-0.099***	
	(0.026)	(0.024)	(0.055)	(0.023)	(0.053)	(0.022)	(0.037)	
	[-0.106]	[-0.108]	[-0.490]	[-0.094]	[-0.446]	[-0.082]	[-0.199]	
IV: 1987 empl share liber								-0.904***
								(0.109)
								[-0.663]
Observations	385	384	384	371	371	369	369	369
R-squared	0.012	0.145		0.193		0.243		0.357
	Without 13 largest cities (population > 2M)							
Empl liberalized sectors	-0.052**	-0.054**	-0.229***	-0.049**	-0.214***	-0.042*	-0.101***	
	(0.026)	(0.024)	(0.054)	(0.023)	(0.052)	(0.022)	(0.038)	
	[-0.105]	[-0.110]	[-0.463]	[-0.097]	[-0.425]	[-0.085]	[-0.201]	
IV: 1987 empl share liber								-0.908***
								(0.113)
								[-0.656]
Observations	379	378	378	365	365	363	363	363
R-squared	0.012	0.148		0.195		0.240		0.357
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Demographics	NO	YES	YES	YES	YES	YES	YES	YES
District-level covariates	NO	NO	NO	YES	YES	YES	YES	YES
Tariffs covariates	NO	NO	NO	NO	NO	YES	YES	YES

Note: robust standard errors are in parentheses, beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1

Note: the drop of 7 and 13 observations in column OLS0 reduces to 2 and 8 observations in columns OLS2 and OLS3 because 5 of the districts dropped would have been lost anyway by including the district number of schools control variable. However, dropping this covariate does not lead to any substantial change neither in terms of statistical significance nor in terms of economic magnitude of the estimates.

A5. Mediation analysis: household earnings and returns to education channels

Channel: Household Earnings (1987-1999)

	(1) Education	(2) Educ w/ tar	(3) Gender Gap	(4) GEG w/ tar
total effect	0.226***	0.117***	-0.218***	-0.107***
	(0.0525)	(0.0441)	(0.0519)	(0.0367)
direct effect	0.0270	0.0151	0.0325	0.0413
	(0.0315)	(0.0577)	(0.0322)	(0.0619)
indirect effect	0.199***	0.102	-0.251***	-0.149
	(0.0758)	(0.103)	(0.0866)	(0.115)
Observations	366	364	366	364
K-P F-stat	41.81	60.49	41.81	60.49

Standard errors robust to heteroscedasticity are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

Channel: Returns to Education (1987-1999)

	(1) Education	(2) Educ w/ tar	(3) Gender Gap	(4) GEG w/ tar
total effect	0.226***	0.113**	-0.216***	-0.107***
	(0.0525)	(0.0440)	(0.0516)	(0.0367)
direct effect	-0.137	0.0272	0.234	0.0262
	(0.367)	(0.0489)	(0.473)	(0.0698)
indirect effect	0.363	0.0857	-0.450	-0.133
	(0.557)	(0.0897)	(0.707)	(0.124)
Observations	364	362	364	362
K-P F-stat	41.78	60.45	41.78	60.45

Standard errors robust to heteroscedasticity are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01