Skill sorting, inter-industry skill wage premium, and production chains: evidence from India 1999-2000

INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 278

Skill Sorting, Inter-Industry Skill Wage Premium, and Production Chains: Evidence from India 1999-2000

Yoko ASUYAMA*

February 2011

Abstract

This paper proposes a mechanism that links industry's technological characteristics (i.e. quality of non-labor inputs, which is proxied by the length of industry production chains), industry-specific skill wage premium, and skill sorting across industries. It is hypothesized that high-skilled workers are sorted into industries where they can receive a higher skill wage premium, by working with better quality non-labor input. The quality of non-labor inputs is assumed to be worse in industries with longer production chains due to the increased involvement of low-skilled labor and poor infrastructure over the sequential production. By examining Indian wage and employment data for 1999-2000, empirical evidence to support this mechanism can be obtained: First, the skill wage premium is lower [higher] in industries with longer [shorter] production chains. Second, the skill wage premium is lower [higher] in industries with a higher [lower] proportion of low-skilled workers producing inputs outside their own industry. Third, the proportion of high-skilled workers is larger in industries with shorter production chains and lower ratio of low-skilled labor involved, i.e., a skill sorting trend can be observed.

Keywords: India, Industry wage, Production chains, Sequential production, Skill wage premium, Skill sorting

JEL classification: J24, J31

* Poverty Alleviation and Social Development Studies Group, Inter-disciplinary Studies Center, IDE (Yoko_Asuyama@ide.go.jp)

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

Skill Sorting, Inter-Industry Skill Wage Premium, and Production Chains: Evidence from India 1999-2000

Yoko Asuyama*

February 2011

(Preliminary)

Abstract

This paper proposes a mechanism that links industry's technological characteristics (i.e. quality of non-labor inputs, which is proxied by the length of industry production chains), industry-specific skill wage premium, and skill sorting across industries. It is hypothesized that high-skilled workers are sorted into industries where they can receive a higher skill wage premium, by working with better quality non-labor input. The quality of non-labor inputs is assumed to be worse in industries with longer production chains due to the increased involvement of low-skilled labor and poor infrastructure over the sequential production. By examining Indian wage and employment data for 1999-2000, empirical evidence to support this mechanism can be obtained: First, the skill wage premium is lower [higher] in industries with longer [shorter] production chains. Second, the skill wage premium is lower [higher] in industries with a higher [lower] proportion of low-skilled workers producing inputs outside their own industry. Third, the proportion of high-skilled workers is larger in industries with shorter production chains and lower ratio of low-skilled labor involved, i.e., a skill sorting trend can be observed.

Keywords: India, Industry wage, Production chains, Sequential production, Skill wage premium, Skill sorting

JEL codes: J24, J31

* Poverty Alleviation and Social Development Studies Group, Inter-disciplinary Studies Center, Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO) (Yoko_Asuyama@ide.go.jp)

1. Introduction

The pattern of skill distribution varies significantly across industries. For instance, Figure 1, which presents the educational attainment of workers across 42 industries in India in 1999-2000, supports this claim. Why do we observe different skill distribution patterns across industries? By utilizing Indian household survey data, this paper aims to empirically answer this question by examining the linkage among industry technological characteristics, industry-specific skill wage premium, and skill distribution patterns across industries. It is hypothesized that the length of industry production chains and the proportion of low-skilled workers involved across the chains are both negatively correlated with industry-specific skill wage premium. This is because as the length of production chains becomes longer, the negative effects on non-labor input quality caused by either low-skilled workers or poor infrastructure accumulate and become larger. The larger proportion of low-skilled workers involved in the chains further magnifies the damage accumulation. Then, it is assumed that the wages of high-skilled workers are dragged down more (i.e. skill wage premium is relatively low) in industries where non-labor input quality is worse due to defect accumulation. In consequence, high-skilled workers tend to be sorted into industries where they can enjoy higher skill wage premium by being matched with high-quality non-labor input.

The main contribution of this paper is to propose another factor (in particular, the length of industry production chains which is a proxy for the quality of non-labor inputs) that explains the inter-industry wage differentials (inter-industry differences in skill wage premium, in particular) and the different patterns of skill allocation across industries. To my knowledge, none of the previous studies have proposed this mechanism. Although a formal economic model is not presented in the current paper and constructing it is left for future research, this paper provides empirical evidence to support the proposed mechanism.

The idea of this paper is closely related with that of Sampson (2011), which develops an assignment model of skill across sectors. In his model, high-skilled workers are sorted into industries which utilize non-labor inputs with higher productivity so that they can best leverage their talent. He also empirically confirms that falls in the price of capital (i.e. productivity increase of non-labor input) utilized by the industry positively affect the growth in industry average wage, which is a proxy for industry skill level. The current paper differs from his paper in three ways: First, his model assumes that production function is log-submodular exhibiting substitutability between labor and non-labor inputs, and quantity of non-labor inputs is endogenously chosen so that high-skilled workers can best utilize their ability by working with larger quantities of non-labor inputs. In contrast, I assume that the production function exhibits complementarity between labor and non-labor inputs, and each worker must work with the same quantity of non-labor inputs of different quality levels (or quality-adjusted productivity levels).[1] Second, Sampson (2011) assumes that productivity of non-labor input changes due to exogenous technological progress. By contrast, I assume that quality-adjusted productivity of non-labor input is determined by the length of production chains and the degree of low-skilled labor’s involvement, assuming defect accumulation over sequential production. Finally, Sampson (2011) empirically examines the relationship between productivity of non-labor input and skill sorting by utilizing the industry average wage data assuming that they proxy industry skill level. This proximity is somewhat crude, as Sampson himself already recognizes. By utilizing within-industry data on workers’ skill distribution, the current paper links non-labor input quality and skill sorting across industries by empirically examining the association between non-labor input quality and industry-specific skill wage premium.

This paper also contributes to the literature on industry wage premium. As

[1] In his paper, Sampson indicates that if the production function is supermodular, which exhibits complementarity, and the quantity ratio between labor and non-labor inputs is fixed, positive assortative matching between high-skilled labor and high-quality non-labor inputs would occur, as hypothesized in the current paper.

Dickens and Katz (1987) compactly summarizes, previous studies try to explain the reason for the persistent inter-industry wage differences, which still remain after controlling observable individual characteristics, based on either competitive or non-competitive labor market models. From the competitive labor market model, the difference in industry wage premium is explained either by the variation in unobserved worker ability (Murphy and Topel 1987; Keane 1993; Abowd et al. 1999; Goux and Maurin 1999; Carruth et al. 2004), or by the differences in unobserved working conditions across industries.[2] From the non-competitive labor market model, the inter-industry wage differentials are explained by the efficiency wage model (Krueger and Summers 1988), the union-threat model (Dickens 1986), and the degree of trade protection/liberalization (Pavcnik et al. 2004; Goldberg and Pavcnik 2005; Dutta 2007; Lundin and Yun 2009). Part of the inter-industry wage differentials can also be explained by the different returns to skill across industries. For example, Robbins and Minowa (1996) and Pavcnik et al. (2004) find substantial variation in skill wage premium across industries in Brazil. Those papers claim that the returns to schooling can vary across industries either because workers with different education level might differ in the degree of labor market mobility, monitoring costs in the efficiency wage models, accumulation of industry-specific skills, or ability to bargain over wages. By contrast, my focus in this paper is on explaining the inter-industry skill wage differentials by the different quality of non-labor inputs which is proxied by the length of industry production chains.

This paper also contributes to providing a presumable mechanism that explains the empirical evidence presented by Asuyama (2011). Asuyama (2011) empirically finds that a country with higher [lower] skill dispersion such as India [China] has higher exports in industries with shorter [longer] production chains, although the mechanism behind this is not empirically examined. The empirical findings of this paper function as

[2] Murphy and Topel (1987) examine the association between the employment and earnings variability of the job and the inter-industry wage differentials, although they find little association.

favorable evidence for the skill sorting hypothesis rather than for the random matching hypothesis, both of which are considered possible in Asuyama (2011).

The rest of the paper is organized as follows. Section 2 presents three hypotheses to be empirically tested in this paper and explains the skill sorting mechanism across industries. Section 3 explains the empirical strategy. Section 4 describes the data, and explains the construction of the key variables. Section 5 presents the estimation results. Section 6 concludes.

2. Hypotheses and Skill Sorting Mechanism

This paper empirically tests the following three hypotheses.

Hypothesis 1. *The skill wage premium is lower [higher] in industries with longer [shorter] production chains.*

Hypothesis 2. *The skill wage premium is lower [higher] in industries with higher [lower] proportion of low-skilled workers who are engaged in production across the chains.*

Hypothesis 3. *The proportion of high-skilled workers is larger in industries which pay higher skill premium, i.e. industries with shorter production chains and lower ratio of low-skilled workers involved across the chains (i.e., skill sorting trend).*

The above hypotheses are derived based on the assumption that as the length of industry production chains becomes longer (or the number of production stages involved in order to produce final industry output increases), the quality of semi-finished intermediate input utilized by the industry tends to be worse because the involvement of low-skilled workers or poor-quality infrastructure (e.g. power and transportation) increases with the number of production stages. As in the O-ring production function proposed by Kremer (1993), the negative impact on the quality of intermediate input at each stage accumulates more as the length of production chains becomes longer. Clearly, the more low-skilled workers are involved across the chains, the larger the accumulated negative effects on the quality of semi-finished intermediate input utilized by the industry become. It follows that high-skilled workers have more incentives to be sorted into industries with shorter production chains, where they can work with high-quality non-labor input. This is because I assume complementarity between labor and non-labor inputs, and thus marginal product and wage of high-skilled labor are dragged down more when working with low-quality non-labor inputs. In other words, since skill-wage premium is larger in industries which utilize higher-quality non-labor input, i.e. industries with shorter production chains and lower proportion of low-skilled workers involved, high-skilled workers are sorted into those industries.

This sorting mechanism is similar to that of Grossman (2004), where high-skilled workers are sorted into industries (e.g. software) in which an individual's contribution to the firm output can be measured perfectly and wages are paid according to their own productivity. In his model, high-skilled workers are disinclined to enter the industry characterized by team-production (e.g. automobile industry) in which each worker's contribution to the output is measured only imperfectly due to imperfect labor contracts. This is because the wages of high-skilled workers are dragged down by the low-skilled team members and become lower than the counterparts in the software industry. Instead of the imperfect labor contract, in the current paper the quality of non-labor semi-finished intermediate input depresses skill-wage premium.

Finally, it is assumed that inter-industry labor mobility is costly so that perfect skill sorting does not occur in reality. In consequence, one can simultaneously observe inter-industry skill wage differentials on one hand and skill-sorting trend on the other hand. In fact, as illustrated in Figure 1, in India each industry employs both low- and high-skilled workers, although their ratio differs substantially. The inter-industry labor mobility also seems low in the Indian sample in this study. Only 1.0% of the Indian workers (male, aged 15-65, full-time, regular wage/salaried workers examined in the empirical analysis) changed their industry (in terms of two-digit NIC-1998 level) during the two years before the date of survey.[3]

3. Empirical Strategy

In order to test Hypothesis 1, I estimate the following equation:

$$\ln Wage_{ij} = \alpha + \beta_1 Edu_i * Leontief_j + Edu_i * X\gamma + Z\delta + \varepsilon_{ij}, \text{ (1)}$$

where $\ln Wage_{ij}$ denotes the logarithm of wage of individual *i* in industry *j*; $Edu_i * Leontief_j$ is an interaction term between Edu_i, which is the skill level of individual *i* (estimated years of education in the current paper), and $Leontief_j$, which indicates the length of production chains in industry *j*; $Edu_i * X$ denotes interaction terms between Edu_i and other industry- or individual-specific variables; Z consists of a set of individual characteristics including industry affiliation dummies. The industry affiliation dummies absorb all the average impact of industry affiliation on wages, which is explicitly decomposed into various factors in the previous studies on the industry-wage premium. In this paper, my focus is on β_1, which captures the extent to which skill wage premium varies according to the length of the industry's production chains. If the estimated coefficient β_1 is significantly negative, it supports Hypothesis 1.

Second, in order to test Hypothesis 2, an interaction term, $Edu_i * LowEdu_j$ is added to equation (1) as follows:

$$\ln Wage_{ij} = \alpha' + \beta_1' Edu_i * Leontief_j + \beta_2' Edu_i * LowEdu_j + Edu_i * X\gamma' + Z\delta' + \varepsilon_{ij}', \text{ (2)}$$

where $LowEdu_j$ represents the proportion of low-skilled workers who are involved in production activities across the production chains of industry *j*. If both Hypotheses 1 and 2 are correct, we can expect that both β_1' and β_2' turn out to be negative.

Finally, Hypothesis 3 is simply tested by the following equation:

$$HighEdu_j = \alpha'' + \beta_1'' Leontief_j + \beta_2'' LowEdu_j + \varepsilon_j'', \text{ (3)}$$

where $HighEdu_j$ indicates the proportion of high-skilled workers working *within*

[3] There are 60 industry categories based on two-digit NIC-1998 code.

industry j. Hypothesis 3 is confirmed when the estimated coefficients β_1'' and β_2'' are both negative.

4. Data

This section summarizes the data used in the empirical analysis. More detailed explanations on data are provided in Appendix I.

It is expected that as the spread of workers' skills becomes greater, negative impacts on non-labor input quality caused by low-skilled workers become larger and high-skilled workers have more incentives to be sorted into industries with shorter production chains. Thus, stronger evidence to confirm the three hypotheses can be found by examining a country with a greater spread of skill. As such a country, I choose India. Skill distribution in India is characterized by a large number of illiterate populations and relatively large proportion of highly-educated individuals. As of 2004/05, the share of employed people who were illiterate or had only education below the primary level was 50%, while that of upper secondary and post-secondary education was 21% (15% and 6%, respectively). India's skill distribution is very unequal compared with other developing countries, such as China, where the proportion of workers who had received no schooling and those with post-lower secondary education was 8% and 19% respectively in 2005.[4]

Data on wage, education level, and other individual characteristics are extracted from the unit-level data of the Employment and Unemployment schedule of the National Sample Survey conducted in 1999-2000 (NSS 1999). The NSS 1999 covers 165,244 households with 819,013 persons across India. As Kijima (2006) states, the Employment and Unemployment schedule of the NSS is the only survey which collects information on individual's earnings and characteristics for the entire country

[4] Figures on workers' skill distribution of India and China are computed from NSSO, *Unit-level data of National Sample Survey (NSS), Employment and Unemployment schedule, 2004-05*, and China's *2005 National 1% Population Sample Survey* (SC and NBS, 2007). India's figures are computed based on the usual principal activity status.

through stratified random sampling procedure. The Employment and Unemployment schedule of the NSS was also conducted in 1972-73, 1977-78, 1983, 1987-88, 1993-94, and 2004-05. However, for the preliminary analysis conducted here, only 1999 data are used.[5]

I restrict the sample to male, prime-age (15-65 years), regular salaried/wage employees who have worked at least 5 days at their main economic activity during the reference week. $\ln Wage_{ij}$ is defined as the logarithm of weekly wage and salary earnings (either in cash or in kind, including bonus and perquisites) for the main economic activity.

$Leontief_j$, which indicates the length of production chains in industry *j*, is the column sum of the Leontief inverse coefficient of each industry computed from the input-output (IO) table of India 1998-99 (CSO 2005), as follows:

$$Leontief_j = \sum_k leon_{kj} \ ,$$

where $leon_{kj}$ is the Leontief inverse coefficient in cell *kj*. Subscripts *k* and *j* denote row and column of the IO table, respectively. $Leontief_j$ measures how many units of domestic inputs industry *j* requires, both directly and indirectly, to produce one unit of output in industry *j*. I use this $Leontief_j$ as a proxy for the length of the production chains of industry *j*.[6] It should be noted that only domestic inputs are used, as explained in Appendix I, since the quality of imported input is assumed to be relatively good and is not likely to be affected much especially by the domestic low-skilled workers.

As $LowEdu_j$, three variables ($bpleon_j$, $bpwin_j$, and $bpbtw_j$) are constructed. $bpleon_j$ is defined as the percentage of low-skilled workers (i.e. illiterate workers and literate workers without formal schooling or with below primary-level education), who

[5] The 1999 dataset has some advantages compared with the other rounds. It contains some important variables which were not asked in the former rounds (e.g. number of workers in the enterprise for which an individual is working). Using the data from NSS 2004 may create a problem when matching with variables computed from the IO table of 2003-04, since the quality of those variables is likely to be worse as a result of estimating the import flow matrix of 2003-04 from the 1993 IO table.

[6] See Asuyama (2011) for a more detailed explanation of $Leontief_j$.

are involved across all the production chains of industry j which include both chains within industry j and those in other industries. It is computed as follows:

$$bpleon_j = \sum_k bpwin_k * (leon_{kj} / Leontief_j),$$

where $bpwin_k$ is the percentage of low-skilled people working *within* industry k, which is computed from NSS 1999. $bpbtw_j$ is also constructed as follows:

$$bpbtw_j = \sum_{k \neq j} bpwin_k * [leon_{kj} / (Leontief_j - leon_{jj})]$$

The term $(Leontief_j - leon_{jj})$ indicates the length of production chains (or the amount of input required) outside industry j. Thus, $bpbtw_j$ represents the share of low-skilled workers who are involved in producing inputs (both directly and indirectly) outside industry j. It should be noted that all three $LowEdu_j$ variables are constructed based on all working individuals including female, non-prime-age, part-time, casual and self-employed workers.

In order to match the industry classifications of the IO table (115 industries) and those of NSS 1999 (5-digit NIC-1998 code), 42 industry categories are constructed as presented in Appendix II. The constructed key variables, $Leontief_j$, $bpleon_j$, $bpwin_j$, and $bpbtw_j$ for the 42 industries are reported in Table 1. $Leontief_j$ tends to be longer in the manufacturing industries and shorter in the agriculture, forestry, fishing, mining, and service industries. Since the Leontief inverse coefficient tends to be larger in cell jj, $bpleon_j$ and $bpwin_j$ show the similar trends. By contrast, $bpbtw_j$ exhibits a different trend. For instance, the share of low-skilled workers in agriculture, forestry, fishing, and mining industries is much smaller in terms of $bpbtw_j$ than in terms of $bpleon_j$ and $bpwin_j$. It indicates that those industries employ the large proportion of low-skilled workers within their industry, but the skill level embodied in the non-labor input from other industries is not so low.

A set of variables X includes the following industry-specific variables: the ratio of imported input to the total input; the ratio of imports of final goods to [output+import-export] which is an indicator for the degree of import competition; the ratio of export to output; a set of dummy variables which indicate whether the industry is delicensed, not reserved for the public sector, and open to foreign direct investment (FDI) up to 51% equity or more. X also includes individual-specific dummy variables indicating whether the individual is employed temporarily, a member of a union/association, working for a public or semi-public firm, working for a small firm employing less than 10 workers. Z includes those individual-specific dummy variables, dummies for Muslim religion, social groups, household headedness, marital status, occupation, rural residence, and a set of State and industry affiliation dummies. Table 2 presents the summary statistics for all the variables used in the regression analysis.

5. Estimation Results

Tables 3 and 4 report the regression results for estimation equations (1) and (2) which test Hypothesis 1 and 2. All estimations are obtained by weighted least squares regression by utilizing the survey weight of NSS 1999. The results in Table 4 restrict the sample to manufacturing and service industries by dropping four primary industries (agriculture; forestry and logging; fishing; mining and quarrying). This is because the quality-adjusted productivity of those primary industries is likely to be substantially affected by inputs such as land, weather, and natural resources, which are not included as inputs in the IO table and thus not captured by $Leontief_j$. Column (1) in both tables just adds an interaction term between Edu_i and $Leontief_j$ to the ordinary Mincer-type wage equation. Another interaction term between Edu_i and $bpleon_j$ is added in column (2), and more interaction terms are controlled in column (3). In all those specifications in both tables, the coefficient on the interaction term between Edu_i and $Leontief_j$ is negative and statistically significant, as predicted by Hypothesis 1.

The coefficient on the interaction term between Edu_i and $bpleon_j$ is not statistically significant in all specifications except for column (3) of Table 4, in which the coefficient is slightly positive and statistically significant. This contradicts

Hypothesis 2 which predicts a negative coefficient on $Edu_i * LowEdu_j$. In order to examine in more detail the effect of worker skill level embodied in the inputs on industry-specific education wage premium, column (4) decomposes the effect of the proportion of low-educated workers into that within industry *j* ($bpwin_j$) and that outside industry *j* ($bpbtw_j$). In this specification, the coefficient on $Edu_i * Leontief_j$ is still significantly negative. Furthermore, it turns out that the share of low-skilled workers involved in producing inputs outside industry *j* ($bpbtw_j$) negatively affects the education wage premium of industry *j*, as predicted by Hypothesis 2. By contrast, the effect of the share of low-skilled workers within industry *j* ($bpwin_j$) is either statistically insignificant (Table 3) or slightly but statistically significantly positive (Table 4). This positive association between within-industry higher percentages of low-skilled labor and higher skill wage premium is puzzling. It is necessary to check more carefully whether this result is robust by examining additional samples in future.

The negative impact of $Leontief_j$ on education premium is not negligible. If an individual moves to an industry with one larger $Leontief_j$ (for example, switch from Banking to Leather and leather product industry), the wage premium for one year schooling drops from 5.5% to 4.6%, even using the smallest negative estimate obtained in column (1) of Table 3.[7] If we use the estimate of the largest negative estimate (column (4) of Table 4), the education premium drops from 17.0% to 14.7%. This means that an individual with university-level education with 16 year schooling earns 28.6% point more at minimum (or 334.9% point more at maximum), *ceteris paribus*, if he works in the Banking industry rather than in the Leather and leather product industry.[8] The negative impact of $bpbtw_j$ is not as large as that of $Leontief_j$. The wage premium for one year schooling drops from 16.04% to 15.95% in Table 3, and from 17.02% to 16.87% in Table 4. As a result, the differences in wage premium for 16

[7] 5.5% = [exp(0.053)-1]*100. 4.6% = [exp(0.053-0.008)-1]*100

[8] The wage of an individual without schooling is set equal to 100 in both industries. 28.6%=[(1+5.5/100)^16-(1+4.6/100)^16]*100. 334.9%=[(1+17.0/100)^16-(1+14.7/100)^16]*100.

year schooling between the Banking and Leather industries become 13.1% point and 25.1% point, respectively.

Apart from $Leontief_j$ and $bpbtw_j$, whether individual i is temporarily employed, a member of a union/association, and working for a small firm; and whether this individual is working in an industry which faces higher import competition, utilizes a larger proportion of imported inputs, and is delicensed, all this is associated with lower education premium. In contrast, whether individual i is working for a public firm and working in an industry which is open to FDI is associated with higher education premium.

As mentioned before, Figure 1 illustrates the educational attainment of the employed population across industries using the NSS 1999 data. It can be seen that the pattern of skill distribution varies significantly across industries. In order to formally test whether skill sorting is taking place as predicted by Hypothesis 3, equation (3) is estimated. Table 5 reports the estimation results. As $HighEdu_j$, which indicates the proportion of high-skilled workers working *within* industry *j*, three variables are used: the percentage of workers with graduate and above education ($gradwin_j$), that with higher secondary and above education ($hswin_j$), and that with secondary and above education ($swin_j$).[9] The results in the upper panel utilize all industries, while those in the lower panel restrict samples to manufacturing and service industries. As expected, $Leontief_j$ is negatively associated with $HighEdu_j$ in all specifications, although some estimates are not statistically significant. If $Leontief_j$ is larger by one, the percentage of high-skilled workers is 10-30% smaller when we only focus on the statistically significant estimates. $bpleon_j$ is also negatively associated with $HighEdu_j$. However, it seems natural that the percentage of high-skilled workers ($HighEdu_j$) and that of low-skilled workers within industry *j* ($bpwin_j$), which is part of

[9] Note that upper-secondary (high school level) education in India is further divided into what is called “higher secondary” education (2 years) and “secondary” education (2 years). The lower-secondary (junior high school level) education is called “middle” education in India.

$bpleon_j$, are negatively correlated. Thus, instead of $bpleon_j$, $bpbtw_j$ is added in columns (7)-(9). $bpbtw_j$ is also negatively associated with $HighEdu_j$, although the association is not statistically significant when using $gradwin_j$ as the dependent variable. One percentage point increase in the share of low-skilled worker who are involved in producing inputs outside industry *j* ($bpbtw_j$) is associated with 1-2% decrease in the share of high-skilled workers in industry *j*. In sum, skill sorting hypothesis (Hypothesis 3) is confirmed in most of the specifications.

6. Concluding Remarks

This paper has proposed a mechanism that links industry's technological characteristics (i.e. quality of non-labor inputs), industry-specific skill wage premium, and skill sorting across industries. The quality of semi-finished non-labor input utilized in a certain industry is proxied by the length of production chains (or units of inputs required, either directly or indirectly, to produce an industry output). This is done by assuming that damage caused by low-skilled workers or poor infrastructure on input quality accumulates and becomes greater as the length of production chains becomes longer. The wages of high-skilled workers are dragged down more as the quality of non-labor input they work with becomes worse. In other words, skill wage premium becomes lower in industries which utilize low-quality non-labor input due to longer production chains and larger proportion of low-skilled workers involved across the chains. In consequence, skilled workers are sorted into those industries where they can receive higher skill-wage premium. In a real world where labor mobility is low, perfect skill sorting does not occur, and thus to some extent it is possible to observe both inter-industry skill wage differentials and skill sorting.

By examining Indian wage and employment data for 1999-2000, this paper also has provided empirical evidence to support the above mechanism. First, it has found that the skill wage premium is lower [higher] in industries with longer [shorter] production chains. Second, the skill wage premium is lower [higher] in industries with a higher [lower] proportion of low-skilled workers who are engaged in producing inputs outside their own industry. Third, the proportion of high-skilled workers is larger in industries with shorter production chains and lower ratio of low-skilled workers involved across the chains (i.e., a skill sorting trend is observable).

Several areas are left for future research. First, constructing a formal economic model for the mechanism proposed in this paper is essential. Second, increasing samples by adding other NSS rounds or using a larger dataset that includes other countries would be useful to make the empirical evidence of this paper more robust. Third, utilizing hourly wage data is also preferable. Last, but not least, finding a more appropriate variable which proxies individual “skill” or “ability” is necessary, although this is a very hard task. In this paper, I measured skill wage premium by private returns to education. However, measurement errors are likely to exist since years of schooling are estimated using the highest education level attained. It is also highly possible that persons with the same years of schooling may possess different levels of skills or abilities and thus differ in their productivities. In that case, higher education wage premium in a certain industry may reflect the situation where workers who have higher ability among the same educated workers have been sorted into the industry and earn higher wages due to their high productivity. This would mean that the empirical results to confirm Hypotheses 1 and 2 in this paper just reflect the result of skill sorting (Hypothesis 3). Such ability differences may partly explain the existence of inter-industry education wage differentials. However, due to the low labor mobility in India, assuming the co-existence of inter-industry skill wage differentials (in terms of returns to true ability) and imperfect skill sorting seems more realistic and reasonable.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251-335.

Aghion, Philippe, Robin Burgess, Stephen J. Redding, and Fabrizio Zilibotti. 2008. "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India." *American Economic Review*, 98(4): 1397-1412.

Asuyama, Yoko. 2011. "Skill Distribution and Comparative Advantage: A Comparison of China and India." IDE Discussion Paper No. 277.

Carruth, Alan, William Collier, and Andy Dickerson. 2004. "Inter-industry Wage Differences and Individual Heterogeneity." *Oxford Bulletin of Economics and Statistics*, 66(5): 811-846.

CSO (Central Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India). 2000. *Input-Output Transactions Table 1993-94*.

CSO. 2005. *Input-Output Transactions Table 1998-99*.

Dickens, William T. 1986. "Wages, Employment and the Threat of Collective Action by Workers." NBER Working Paper, No.1856.

Dickens, William T., and Lawrence F. Katz. 1987. "Inter-Industry Wage Differences and Industry Characteristics." In Lang, Kevin, and Jonathan S. Leonard, eds. *Unemployment and the Structure of Labor Markets*. New York and Oxford: Basil Blackwell.

Dutta, Puja Vasudeva. 2007. "Trade Protection and Industry Wages in India." *Industrial and Labor Relations Review*, 60(2): 268-286.

Goldberg, Pinelopi Koujianou, and Nina Pavcnik. 2005. "Trade, Wages, and the Political Economy of Trade Protection: Evidence from the Colombian Trade Reforms." *Journal of International Economics*, 66(1): 75-105.

Goux, Dominique, and Eric Maurin. 1999. "Persistence of Interindustry Wage

Differentials: A Reexamination Using Matched Worker-Firm Panel Data." *Journal of Labor Economics*, 17(3): 492-533.

Grossman, Gene M. 2004. "The Distribution of Talent and the Pattern and Consequences of International Trade." *Journal of Political Economy*, 112(1): 209-239.

Keane, Michael P. 1993. "Individual Heterogeneity and Interindustry Wage Differentials." *Journal of Human Resources*, 28(1): 134-161.

Kijima, Yoko. 2006. "Why Did Wage Inequality Increase? Evidence from Urban India 1983-99." *Journal of Development Economics*, 81(1): 97-117.

Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics*, 108(3): 551-575.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica*, 56(2): 259-293.

Lundin, Nannan, and Lihong Yun. 2009. "International Trade and Inter-Industry Wage Structure in Swedish Manufacturing: Evidence from Matched Employer-Employee Data." *Review of International Economics*, 17(1): 87-102.

Ministry of Commerce and Industry, Government of India. 2002. *Handbook of Industrial Policy and Statistics 2001*. New Delhi.

Ministry of Industry, Government of India. 1991. "Statement on Industrial Policy." July 24, 1991.

Murphy, Kevin M., and Robert H. Topel. 1987 "Unemployment, Risk, and Earnings: Testing for Equalizing Wage Differences in the Labor Market." In Lang, Kevin, and Jonathan S. Leonard, eds. *Unemployment and the Structure of Labor Markets*. New York and Oxford: Basil Blackwell.

NSSO (National Sample Survey Organisation, Ministry of Statistics and Programme Implementation, Government of India). *Unit-level data of National Sample Survey, Employment and Unemployment schedule, 1999-00 (55th round).*

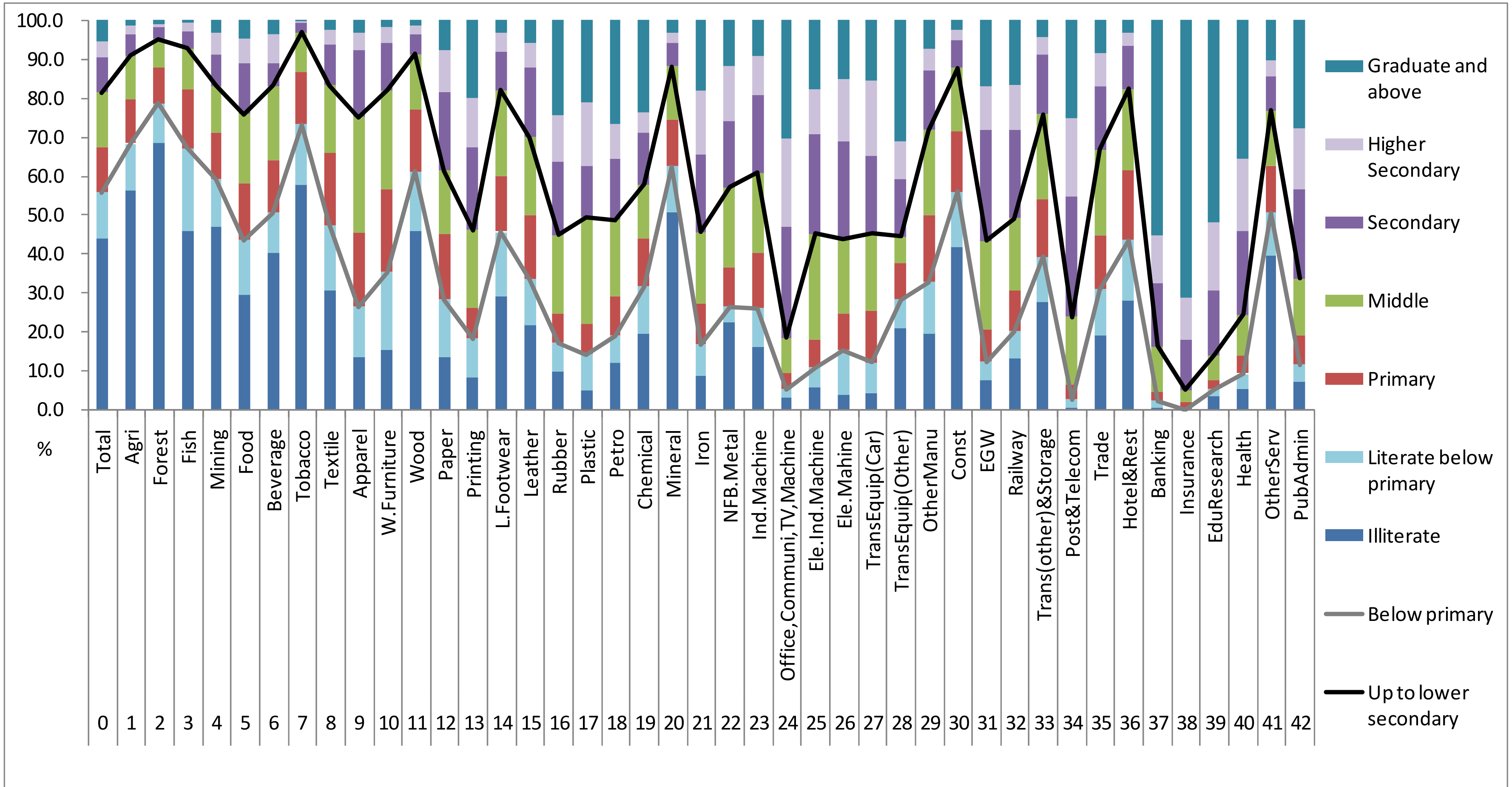
Pavcnik, Nina, Andreas Blom, Pinelopi Goldberg, and Norbert Schady. 2004. "Trade Liberalization and Industry Wage Structure: Evidence from Brazil." *World Bank Economic Review*, 18(3): 319-344.

Robbins, Donald, and Mari Minowa. 1996. "Do Returns to Schooling Vary across Industries?" In Birdsall, Nancy and Richard H. Sabot, eds. *Opportunity Forgone: Education in Brazil*. Washington, D.C.: Inter-American Development Bank.

Sampson, Thomas. 2011. "Assignment Reversals: Trade, Skill Allocation and Wage Inequality." Mimeo. Harvard University.

SC and NBS (State Council and National Bureau of Statistics of China). 2007. *2005 National 1% Population Sample Survey*. Beijing: China Statistics Press.

Figure 1 Educational Attainment of Employed Population by Industry



Notes: An individual is considered working based on the weekly activity status. For industry classification, see Appendix I.

Source: NSSO, *Unit-level data of National Sample Survey (NSS), Employment and Unemployment schedule, 1999-2000.*

Table 1 Industry-specific Indices of Production-Chain Length and Low-skilled Worker Ratios

		Leontief	(rank)	bpleon (%)	(rank)	bpwin (%)	(rank)	bpbtw (%)	(rank)
1	Agriculture	1.391	(34)	61.5	(2)	68.2	(3)	29.5	(27)
2	Forestry & logging	1.161	(41)	72.6	(1)	78.6	(1)	33.9	(17)
3	Fishing	1.227	(38)	61.4	(3)	67.1	(4)	34.1	(15)
4	Mining & quarrying	1.314	(36)	52.0	(6)	59.1	(7)	27.4	(32)
5	Food	1.914	(19)	45.4	(11)	43.4	(13)	48.1	(1)
6	Bevarage	1.946	(15)	45.5	(9)	50.4	(10)	40.1	(6)
7	Tobacco	1.884	(21)	59.2	(4)	73.2	(2)	41.3	(4)
8	Textile	2.157	(3)	41.9	(13)	47.4	(11)	35.2	(12)
9	Wearing apparel	2.247	(2)	32.5	(18)	26.5	(23)	37.4	(8)
10	Wooden furniture	1.772	(25)	37.6	(15)	35.3	(16)	41.1	(5)
11	Wood and wood products	1.710	(28)	53.8	(5)	61.0	(6)	42.2	(3)
12	Paper, paper products & newsprint	2.101	(5)	30.6	(20)	28.1	(22)	33.8	(19)
13	Printing & publishing	1.761	(27)	22.8	(29)	18.3	(28)	29.6	(26)
14	Leather footwear	2.073	(9)	39.5	(14)	45.5	(12)	33.8	(18)
15	Leather and leather products	2.312	(1)	34.3	(17)	33.4	(17)	35.5	(10)
16	Rubber products	2.128	(4)	26.0	(25)	17.1	(29)	34.3	(14)
17	Plastic products	1.960	(14)	21.3	(32)	14.0	(32)	29.8	(24)
18	Petroleum & coal tar products	1.501	(32)	24.6	(27)	18.8	(27)	36.7	(9)
19	Chemical products	1.942	(16)	31.8	(19)	31.6	(19)	32.2	(21)
20	Non-metallic mineral products	1.773	(23)	49.3	(7)	62.4	(5)	31.8	(22)
21	Iron and steel	2.090	(7)	21.8	(31)	16.9	(30)	29.7	(25)
22	Non-ferrous basic metals	1.679	(29)	27.4	(23)	26.5	(24)	28.7	(29)
23	Industrial machinery, machine tools, fabricated metal products	1.925	(17)	25.7	(26)	26.0	(25)	25.3	(37)
24	Office computing machines, communication equipment, electronic equipment (incl. TV)	1.764	(26)	13.3	(37)	5.2	(39)	24.7	(40)
25	Electrical industrial machinery	1.897	(20)	17.2	(36)	10.9	(36)	24.1	(41)
26	Electrical machinery and apparatus (wire, cable, batteries, electric appliances)	1.772	(24)	19.9	(33)	15.1	(31)	26.4	(33)
27	Motor vehicles	2.092	(6)	18.9	(35)	12.1	(34)	26.1	(34)
28	Other transport equipment (excl. aircraft)	2.069	(10)	27.2	(24)	28.3	(21)	26.0	(35)
29	Miscellaneous manufacturing (incl. watches and clocks, medical instruments, aircraft)	2.025	(12)	30.1	(21)	32.8	(18)	27.4	(31)
30	Construction	1.793	(22)	46.2	(8)	55.9	(8)	33.7	(20)
31	Electricity, gas and water	2.020	(13)	19.7	(34)	12.4	(33)	35.4	(11)
32	Transport (railway)	1.921	(18)	22.5	(30)	20.0	(26)	25.2	(39)
33	Transport (other), storage	1.614	(30)	34.8	(16)	39.2	(15)	28.0	(30)
34	Post & telecommunication	1.222	(39)	7.3	(42)	2.5	(40)	29.3	(28)
35	Trade (wholesale and retail)	1.235	(37)	29.9	(22)	30.9	(20)	25.3	(38)
36	Hotels and restaurants	2.025	(11)	45.5	(10)	43.4	(14)	47.6	(2)
37	Banking	1.315	(35)	8.7	(40)	2.3	(41)	35.0	(13)
38	Insurance	1.424	(33)	7.7	(41)	0.0	(42)	25.5	(36)
39	Education and research	1.183	(40)	9.6	(39)	5.3	(38)	34.0	(16)
40	Medical and health	2.081	(8)	24.0	(28)	9.1	(37)	37.7	(7)
41	Other services (RE, BusiServ, ComputerServ, Renting, Community, Other)	1.516	(31)	44.5	(12)	50.6	(9)	31.7	(23)
42	Public administration , defence	1.000	(42)	11.5	(38)	11.5	(35)		

Notes: For the definition of variables and industry classification, see Section 4 and Appendix.

Table 2 Summary Statistics for the Variables Used in the Regression Analyses

Variable	Unit	Mean	Std. Dev.
wage	Indian Rupee	998.9	952.5
age	year	36.4	11.0
eduy	year	9.2	5.0
leontief		1.499	0.392
bpleon	%	27.8	15.6
bpbtw	%	31.3	5.7
bpwin	%	27.9	19.3
imported input	%	11.12	11.18
import	%	5.46	9.68
export	%	6.46	10.18
delicensed	dummy	0.999	0.036
private	dummy	0.795	0.404
FDI	dummy	0.507	0.500
muslim	dummy	0.099	0.298
SG (ST)	dummy	0.051	0.220
SG (SC)	dummy	0.147	0.354
SG (OBC)	dummy	0.294	0.456
SG (Other)	dummy	0.508	0.500
household head	dummy	0.713	0.453
married	dummy	0.796	0.403
temporary	dummy	0.281	0.450
union	dummy	0.432	0.495
occ1 (professionals)	dummy	0.051	0.219
occ2 (technicians)	dummy	0.097	0.296
occ3 (govt admin & executive officials)	dummy	0.011	0.103
occ4 (managers)	dummy	0.022	0.146
occ5 (clerical)	dummy	0.210	0.407
occ6 (sales)	dummy	0.069	0.254
occ7 (service)	dummy	0.106	0.308
occ8 (farmers etc.)	dummy	0.050	0.218
occ9 (production related: supervisors & foremen)	dummy	0.027	0.162
occ10 (production related: others)	dummy	0.336	0.472
occ11 (not classified)	dummy	0.022	0.147
public firm	dummy	0.345	0.476
small firm	dummy	0.326	0.469
rural	dummy	0.373	0.484

Notes: The number of observations is 24,955 for *bpbtw*, and 32,101 for the other variables. For the definition of variables, see Appendix I.

Table 3 Regression Results for Wage Equation (all industries)

	(1)		(2)		(3)		(4)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
age	0.042	(0.004) ***	0.042	(0.004) ***	0.044	(0.004) ***	0.046	(0.004) ***
age squared	-0.0004	(0.000) ***	-0.0004	(0.000) ***	-0.0004	(0.000) ***	-0.0004	(0.000) ***
eduy	0.053	(0.005) ***	0.053	(0.005) ***	0.120	(0.030) ***	0.149	(0.034) ***
eduy*leontief	-0.008	(0.003) ***	-0.009	(0.003) ***	-0.014	(0.004) ***	-0.013	(0.006) **
eduy*bpleon			0.0001	(0.000)	0.0001	(0.000)		
eduy*bpbtw							-0.001	(0.000) **
eduy*bpwin							0.00005	(0.000)
eduy*imported input					0.00002	(0.000)	-0.0003	(0.000)
eduy*import					-0.0003	(0.000) ***	-0.0004	(0.000) ***
eduy*export					0.0001	(0.000)	0.0001	(0.000)
eduy*temporary					-0.033	(0.003) ***	-0.030	(0.003) ***
eduy*union					-0.004	(0.003)	-0.001	(0.003)
eduy*public firm					0.003	(0.003)	-0.002	(0.003)
eduy*small firm					-0.008	(0.003) ***	-0.011	(0.003) ***
eduy*delicensed					-0.055	(0.028) *	-0.058	(0.029) **
eduy*private					-0.001	(0.005)	0.001	(0.008)
eduy*FDI					0.014	(0.004) ***	0.015	(0.004) ***
muslim	0.015	(0.015)	0.016	(0.015)	0.004	(0.015)	-0.0002	(0.017)
ST	-0.037	(0.025)	-0.037	(0.025)	-0.036	(0.024)	-0.050	(0.026) *
SC	-0.061	(0.016) ***	-0.061	(0.016) ***	-0.056	(0.016) ***	-0.040	(0.018) **
OBC	-0.065	(0.012) ***	-0.065	(0.012) ***	-0.063	(0.012) ***	-0.077	(0.012) ***
household head	0.104	(0.014) ***	0.104	(0.014) ***	0.097	(0.014) ***	0.095	(0.014) ***
married	0.061	(0.016) ***	0.061	(0.016) ***	0.052	(0.016) ***	0.045	(0.016) ***
temporary	-0.215	(0.014) ***	-0.215	(0.014) ***	0.046	(0.024) *	0.054	(0.025) **
union	0.192	(0.014) ***	0.192	(0.014) ***	0.225	(0.029) ***	0.244	(0.030) ***
occ1 (professionals)	0.576	(0.052) ***	0.571	(0.053) ***	0.504	(0.053) ***	0.534	(0.064) ***
occ2 (technicians)	0.380	(0.047) ***	0.374	(0.047) ***	0.326	(0.047) ***	0.332	(0.055) ***
occ3 (govt admin&exe)	0.556	(0.058) ***	0.552	(0.058) ***	0.480	(0.058) ***	0.596	(0.076) ***
occ4 (managers)	0.626	(0.067) ***	0.620	(0.067) ***	0.538	(0.068) ***	0.557	(0.076) ***
occ5 (clerical)	0.168	(0.040) ***	0.162	(0.042) ***	0.116	(0.042) ***	0.116	(0.048) **
occ6 (sales)	0.171	(0.047) ***	0.165	(0.048) ***	0.124	(0.049) **	0.125	(0.053) **
occ7 (service)	0.114	(0.041) ***	0.108	(0.042) ***	0.073	(0.041) *	0.009	(0.048)
occ9 (prod. supervisors)	0.265	(0.046) ***	0.258	(0.046) ***	0.227	(0.046) ***	0.222	(0.052) ***
occ10 (prod. others)	0.199	(0.039) ***	0.192	(0.040) ***	0.151	(0.040) ***	0.139	(0.046) ***
occ11 (not classified)	0.312	(0.054) ***	0.305	(0.055) ***	0.251	(0.055) ***	0.245	(0.061) ***
public firm	0.217	(0.015) ***	0.216	(0.015) ***	0.179	(0.034) ***	0.264	(0.037) ***
small firm	-0.186	(0.013) ***	-0.186	(0.013) ***	-0.112	(0.026) ***	-0.077	(0.026) ***
rural	-0.136	(0.013) ***	-0.137	(0.014) ***	-0.134	(0.013) ***	-0.131	(0.012) ***
State dummies	YES		YES		YES		YES	
Industry dummies	YES		YES		YES		YES	
Number of observations	32101		32101		32101		24955	
R-squared	0.590		0.590		0.598		0.598	
F-statistics	240.46		237.98		230.95		194.14	

Notes: The dependent variable is the logarithm of an individual's weekly wage. All estimations are obtained by weighted least squares regression using the survey weight of NSS 1999. Robust standard errors are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively. SG(Other) and occ8 (Farmers etc.) are omitted as reference categories. Public administration industry, for which *bpbtw* is not available, is not included in the estimation of column (4). For the definition of variables, see Appendix I.

Table 4 Regression Results for Wage Equation (manufacturing and service industries)

	(1)			(2)			(3)			(4)		
	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
age	0.042	(0.004)	***	0.042	(0.004)	***	0.044	(0.004)	***	0.045	(0.004)	***
age squared	-0.0004	(0.000)	***	-0.0004	(0.000)	***	-0.0004	(0.000)	***	-0.0004	(0.000)	***
eduy	0.054	(0.005)	***	0.054	(0.005)	***	0.109	(0.030)	***	0.157	(0.035)	***
eduy*leontief	-0.009	(0.003)	***	-0.009	(0.004)	**	-0.018	(0.005)	***	-0.020	(0.007)	***
eduy*bpleon				-0.00001	(0.000)		0.0004	(0.000)	***			
eduy*bpbtw										-0.001	(0.000)	***
eduy*bpwin										0.0004	(0.000)	***
eduy*imported input							-0.0003	(0.000)		-0.001	(0.000)	***
eduy*import							0.0003	(0.000)		0.0003	(0.000)	
eduy*export							-0.0002	(0.000)		-0.0003	(0.000)	
eduy*temporary							-0.032	(0.003)	***	-0.029	(0.003)	***
eduy*union							-0.008	(0.004)	**	-0.006	(0.004)	
eduy*public firm							0.006	(0.003)	*	0.001	(0.003)	
eduy*small firm							-0.009	(0.003)	***	-0.012	(0.003)	***
eduy*delicensed							-0.041	(0.028)		-0.042	(0.029)	
eduy*private							0.004	(0.005)		0.010	(0.008)	
eduy*FDI							0.006	(0.005)		0.007	(0.005)	
muslim	0.009	(0.016)		0.009	(0.016)		-0.003	(0.016)		-0.009	(0.017)	
ST	-0.022	(0.027)		-0.022	(0.027)		-0.027	(0.026)		-0.039	(0.030)	
SC	-0.075	(0.017)	***	-0.075	(0.017)	***	-0.072	(0.017)	***	-0.056	(0.019)	***
OBC	-0.062	(0.012)	***	-0.062	(0.012)	***	-0.061	(0.012)	***	-0.075	(0.013)	***
household head	0.110	(0.014)	***	0.110	(0.014)	***	0.103	(0.014)	***	0.100	(0.015)	***
married	0.064	(0.017)	***	0.064	(0.017)	***	0.056	(0.017)	***	0.052	(0.017)	***
temporary	-0.225	(0.014)	***	-0.225	(0.014)	***	0.039	(0.027)		0.056	(0.027)	**
union	0.196	(0.015)	***	0.196	(0.015)	***	0.278	(0.034)	***	0.307	(0.036)	***
occ1 (professionals)	0.400	(0.066)	***	0.400	(0.066)	***	0.351	(0.065)	***	0.324	(0.089)	***
occ2 (technicians)	0.198	(0.061)	***	0.198	(0.061)	***	0.169	(0.060)	***	0.116	(0.083)	
occ3 (govt admin&exe)	0.375	(0.070)	***	0.375	(0.070)	***	0.325	(0.070)	***	0.392	(0.100)	***
occ4 (managers)	0.437	(0.079)	***	0.437	(0.079)	***	0.379	(0.079)	***	0.343	(0.100)	***
occ5 (clerical)	-0.009	(0.057)		-0.009	(0.058)		-0.033	(0.057)		-0.089	(0.079)	
occ6 (sales)	-0.0002	(0.063)		-0.0001	(0.063)		-0.022	(0.063)		-0.075	(0.083)	
occ7 (service)	-0.053	(0.058)		-0.053	(0.058)		-0.065	(0.057)		-0.178	(0.080)	**
occ9 (prod. supervisors)	0.078	(0.062)		0.078	(0.062)		0.066	(0.060)		0.001	(0.082)	
occ10 (prod. others)	0.017	(0.057)		0.017	(0.057)		-0.005	(0.056)		-0.072	(0.078)	
occ11 (not classified)	0.092	(0.067)		0.092	(0.067)		0.065	(0.066)		0.005	(0.087)	
public firm	0.213	(0.016)	***	0.213	(0.016)	***	0.147	(0.037)	***	0.238	(0.041)	***
small firm	-0.182	(0.013)	***	-0.182	(0.013)	***	-0.099	(0.026)	***	-0.054	(0.026)	**
rural	-0.138	(0.014)	***	-0.138	(0.014)	***	-0.135	(0.014)	***	-0.129	(0.013)	***
State dummies	YES			YES			YES			YES		
Industry dummies	YES			YES			YES			YES		
Number of observations	30070			30070			30070			22924		
R-squared	0.579			0.579			0.587			0.589		
F-statistics	227.21			225.31			213.54			180.21		

Note: Same as in Table 3.

Table 5 Regression Results for Skill Sorting

sample	A. All industries								
dependent variable	**gradwin**	**hswin**	**swin**	**gradwin**	**hswin**	**swin**	**gradwin**	**hswin**	**swin**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
leontief	-11.888	-12.189	-9.273	-12.793 ***	-13.461 ***	-10.994 **	-9.014	-6.628	0.074
	(7.169)	(9.143)	(11.492)	(4.479)	(4.609)	(4.167)	(7.736)	(9.607)	(11.670)
bpleon				-0.756 ***	-1.062 ***	-1.437 ***			
				(0.095)	(0.098)	(0.088)			
bpbtw							-0.657	-1.092 **	-1.654 **
							(0.414)	(0.514)	(0.624)
constant	36.388 ***	46.111 ***	56.258 ***	62.403 ***	82.673 ***	105.726 ***	52.583 ***	71.513 ***	92.968 ***
	(12.852)	(16.391)	(20.601)	(8.664)	(8.916)	(8.060)	(17.547)	(21.790)	(26.469)
Number of observations	42	42	42	42	42	42	41	41	41
R-squared	0.064	0.043	0.016	0.644	0.763	0.874	0.111	0.131	0.160
F-statistics	2.75	1.78	0.65	35.3	62.77	135.23	2.36	2.86	3.63
Prov >F	0.105	0.190	0.425	0.000	0.000	0.000	0.108	0.069	0.036

sample	B. Manufacturing and service industries								
dependent variable	**gradwin**	**hswin**	**swin**	**gradwin**	**hswin**	**swin**	**gradwin**	**hswin**	**swin**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
leontief	-23.330 ***	-27.530 ***	-29.633 **	-9.881 *	-8.109	-3.071	-23.037 ***	-24.837 **	-23.354 *
	(7.565)	(9.499)	(11.710)	(5.680)	(5.690)	(4.754)	(8.403)	(10.280)	(12.208)
bpleon				-0.844 ***	-1.219 ***	-1.667 ***			
				(0.134)	(0.135)	(0.113)			
bpbtw							-0.618	-1.039 **	-1.587 ***
							(0.388)	(0.475)	(0.564)
constant	59.183 ***	76.665 ***	96.805 ***	59.470 ***	77.080 ***	97.371 ***	79.061 ***	105.829 ***	137.194 ***
	(13.909)	(17.465)	(21.530)	(9.672)	(9.690)	(8.096)	(18.178)	(22.238)	(26.411)
Number of observations	38	38	38	38	38	38	37	37	37
R-squared	0.209	0.189	0.151	0.628	0.757	0.883	0.263	0.278	0.293
F-statistics	9.51	8.4	6.4	29.56	54.62	132.44	6.06	6.54	7.05
Prov >F	0.004	0.006	0.016	0.000	0.000	0.000	0.006	0.004	0.003

Notes: The dependent variable is the percentage of high-skilled workers (gradwin =graduate and above, hswin = higher secondary and above, swin = secondary and above) in industry *j*. Standard errors are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively

Appendix I. Data Sources and Construction of Variables Used in the Analysis

Data / Variable	Sources / Construction method
wage	Source: NSSO, *Unit-level data of National Sample Survey, Employment and Unemployment schedule, 1999-00 [NSS 1999]*. Weekly wage and salary earnings (received or receivable, including bonus and perquisites, expressed in terms of Indian Rupees) are for full-time economic activities, which are those done for at least 5 days, during the reference week. They include both in-cash and in-kind earnings. Wages and salary in kind are valuated at the current retail price by the NSSO. In the NSS, it is considered working for a half day if an individual has worked for 1 hour or more but less than 4 hours in a day. If an individual has worked for 4 hours or more, it is considered working for a full day. This information is used to compute the number of working days. Only samples of male, prime-age (15-65), regular salaried/wage employees who have worked at least 5 days at their main economic activity during the reference week are used in the regression analyses.
age	Source: *NSS 1999*. Individual *i*'s age.
eduy	Source: *NSS 1999*. Years of education, which are estimated from the highest general education level attained by individual *i*, are determined by allocating the following number of years of schooling to each level: illiterate (0 years), literate without formal schooling or literate but below primary (2.5 years), primary (5 years), middle (8 years), secondary (10 years), higher secondary (12 years), graduate and above (16 years).
leontief	Sources: *Input-Output Transaction Table* of 1998-99 and 1993-94 (CSO, 2000, 2005). *[India IO 1998, 1993]* $Leontief_j$, which is the column sum of the Leontief inverse coefficient of industry *j*, is computed as follows: $$Leontief_j = \sum_k leon_{kj} ,$$ where $leon_{kj}$ is the Leontief inverse coefficient in cell *kj*. Subscripts *k* and *j* denote row and column of the IO table, respectively. The Leontief inverse coefficient matrix *L* comprised of *k* * *j* $leon_{kj}$ s is computed as $$L = (I - A_d)^{-1} ,$$ where *I* is the identity matrix; A_d is the input coefficient matrix for domestic inputs, in which the coefficient in cell *kj* is the domestic input in cell *kj* divided by the output of industry *j*. Since the India IO 1998 does not contain an import flow matrix, the values of domestic and imported inputs are estimated by using IO 1993 which contains an import flow matrix. It is assumed that the share of imported input for each column *j* in the total import is unchanged from 1993 to 1998. For more details on computation, see the Appendix I of Asuyama (2011).
bpleon	Sources: *NSS 1999*, and *India IO 1998, 1993.* For the definition and construction of the index, see Section 4 in the main text.

bpbtw	Sources: *NSS 1999*, and *India IO 1998, 1993.* For the definition and construction of index, see Section 4 in the main text.
bpwin	Source: *NSS 1999*. For the definition and construction of index, see Section 4 in the main text.
imported input	Sources: *India IO 1998, 1993.* The ratio of imported input is calculated for each industry as the percentage of the value of imported inputs to the value of total inputs, using the IO tables of India. Import values are estimated as in the above explanation of Leontief.
Import	Sources: *India IO 1998.* An industry's ratio of final good import is defined as [import/(output+import-export)]*100(%).
export	Sources: *India IO 1998.* An industry's ratio of final good export is defined as [export/output]*100(%).
delicensed	Sources: Aghion et al., (2008) and *Handbook of Industrial Policy and Statistics 2001* (Ministry of Commerce and Industry, 2002). A binary variable that equals one if industry *j* is still covered under compulsory industrial licensing which was stipulated in the Industries (Development and Regulation) Act of 1951, and zero otherwise. Only industry 7 (tobacco) is coded as one.
private	Sources: "*Statement on Industrial Policy, July 24, 1991*" by the Ministry of Industry, Government of India and *Handbook of Industrial Policy and Statistics 2001* (Ministry of Commerce and Industry, 2002). A binary variable that equals one if industry *j* is *not* reserved for the public sector, and zero if reserved for that sector. Only industries 32 (railway transport) and 42 (public administration) are coded as one.
FDI	Sources: Aghion et al., (2008) and *Handbook of Industrial Policy and Statistics 2001* (Ministry of Commerce and Industry, 2002). A binary variable that equals one if industry *j* is partly (in terms of some sub-industry level) or entirely opened to automatic approval of FDI for up to 51 percent equity or more, and zero otherwise.
muslim	Source: *NSS 1999*. A binary variable that equals one if individual *i*'s religion is Islam, and zero otherwise.
SG (ST, SC, OBC, Other)	Source: *NSS 1999*. Dummy variables that indicate to which social group individual *i* belongs. ST is scheduled tribe, SC is scheduled caste, OBC is other backward class, and Other is other social groups.
household head	Source: *NSS 1999*. A binary variable that equals one if individual *i* is the head of the household, and zero otherwise.

married	Source: *NSS 1999*. A binary variable that equals one if individual *i* is currently married, and zero otherwise.
temporary	Source: *NSS 1999*. A binary variable that equals one if individual *i*'s nature of employment is temporary, and zero if permanent.
union	Source: *NSS 1999*. A binary variable that equals one if individual *i* is a member of a union/association, and zero otherwise.
occ (1-11)	Source: *NSS 1999*. Dummy variables that indicate individual *i*'s occupation. Figures in the parenthesis below indicate NCO-1968 code. occ1: Professionals. (00, 02, 05, 07, 10-14, 150, 16-19) occ2: Technicians etc. (01, 03, 04, 06, 08, 09, 151-156, 159) occ3: Government administrative & executive officials (20, 21, 31) occ4: Managers (22-30) occ5: Clerical and related workers (3) occ6: Sales workers (4) occ7: Service workers (5) occ8: Farmers, fishermen, hunters, loggers, and related workers (6) occ9: Production and related workers, transport equipment operators and laborers: supervisors & foremen (Among 71-98, all three-digit codes ending with zero (e.g. 710, 720, 730, ... 980) occ10: Production and related workers, transport equipment operators and laborers: other than supervisors & foremen (7-9 except for those recorded as occ9) occ11: Not classified (X)
public firm	Source: *NSS 1999*. A binary variable that equals one if the enterprise for which individual *i* is working is either a public or semi-public type, and zero otherwise.
small firm	Source: *NSS 1999*. A binary variable that equals one if the number of workers of the enterprise for which individual *i* is working is less than 10, and zero otherwise.
rural	Source: *NSS 1999*. A binary variable that equals one if individual *i*'s area of residence is rural, and zero if urban.

Appendix II. Industry Classification

		IO 1998 (IO code)	NSS 1999 (NIC-1998 code)
1	Agriculture	1-17, 19, 20	01
2	Forestry & logging	21	02
3	Fishing	23	05
4	Mining & quarrying	23-32	10-14
5	Food	18, 33-38	151-154
6	Bevarage	39	155
7	Tobacco	40	16
8	Textile	41-47, 49	17
9	Wearing apparel	48	181
10	Wooden furniture	50	36101
11	Wood and wood products	51	20
12	Paper, paper products & newsprint	52	21
13	Printing & publishing	53	221, 222
14	Leather footwear	54	192
15	Leather and leather products	55	191, 182
16	Rubber products	56	251
17	Plastic products	57	252
18	Petroleum & coal tar products	58, 59	23
19	Chemical products	60-68	24
20	Non-metallic mineral products	69-71	26
21	Iron and steel	72-74	271, 273
22	Non-ferrous basic metals	75	272
23	Industrial machinery, machine tools, fabricated metal products	76-81, 83	28, 291, 292
24	Office computing machines, communication equipment, electronic equipment (incl. TV)	82, 88, 90	30, 32
25	Electrical industrial machinery	84	311, 312
26	Electrical machinery and apparatus (wire, cable, batteries, electric appliances)	85-87, 89	293, 313-315, 319
27	Motor vehicles	93	34
28	Other transport equipment (excl. aircraft)	91, 92, 94-96	351, 352, 359
29	Miscellaneous manufacturing (incl. watches and clocks, medical instruments, aircraft)	97, 98	331-333, 353, 361(excluding 36101), 369
30	Construction	99	45
31	Electricity, gas and water	100-102	40, 41
32	Transport (railway)	103	601
33	Transport (other), storage	104, 105	602, 603, 61-63
34	Post & telecommunication	106	64
35	Trade (wholesale and retail)	107	50-52
36	Hotels and restaurants	108	55
37	Banking	109	65, 67
38	Insurance	110	66
39	Education and research	112	73, 80
40	Medical and health	113	851, 852
41	Other services (RE, BusiServ, ComputerServ, Renting, Community, Other)	111, 114	70-72, 74, 853, 90-93, 95, 99
42	Public administration , defence	115	75

Notes: The above classification of 42 industries is designed so that the content of the industries between the IO table data and NSS 1999 data is matched. In order to obtain reliable data on within-industry skill distribution, the number of samples for employed persons extracted from NSS 1999 is kept to at least around 100 in each industry.