INSTITUTE OF DEVELOPING ECONOMIES

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

IDE DISCUSSION PAPER No. 545

Skill Sorting and Production Chains: Evidence from India

Yoko Asuyama*

November 2015

Abstract

This study proposes a new mechanism that explains skill-sorting patterns and skill wage differentials across industries based on the length of the industry's production chain. A simple simultaneous production model shows that when the quality of intermediate inputs deteriorates rapidly along the production chains, high-skilled individuals choose to work in industries with shorter production chains because of higher returns to skill. I empirically confirm this skill-sorting pattern and these inter-industry skill wage differentials in India, where the quality of intermediate inputs is likely to degrade rapidly because of the high number of unskilled laborers, poor infrastructure, and less-advantaged technology. The results remain robust even when considering selection bias, alternative reasons for inter-industry skill wage differentials, and a different period. The results of this study have important implications when considering countries' industrial development patterns.

Keywords: India, Input quality, Production chains, Return to skill, Skill sorting, Skill wage premium

JEL classification: I25, I26, J24, J31, L23, O15

---

* Poverty Alleviation and Social Development Studies Group, Development 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 and Production Chains: Evidence from India

Yoko Asuyama[*†]

Abstract

This study proposes a new mechanism that explains skill-sorting patterns and skill wage differentials across industries based on the length of the industry's production chain. A simple simultaneous production model shows that when the quality of intermediate inputs deteriorates rapidly along the production chains, high-skilled individuals choose to work in industries with shorter production chains because of higher returns to skill. I empirically confirm this skill-sorting pattern and these inter-industry skill wage differentials in India, where the quality of intermediate inputs is likely to degrade rapidly because of the high number of unskilled laborers, poor infrastructure, and less-advantaged technology. The results remain robust even when considering selection bias, alternative reasons for inter-industry skill wage differentials, and a different period. The results of this study have important implications when considering countries' industrial development patterns.

JEL Classification: I25, I26, J24, J31, L23, O15

Keywords: India, Input quality, Production chains, Return to skill, Skill sorting, Skill wage premium

.

[*] Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO), Wakaba 3-2-2, Mihamaku, Chiba, 261-8545, Japan. E-mail: Yoko_Asuyama@ide.go.jp.

[†] I thank Yuji Genda, Hideaki Goto, Mizuki Harizuka, Masami Ishida, Satoshi Inomata, Md. Humayun Kabir, Tomohiro Machikita, Mayumi Murayama, Mari Nakamura, Hiroshi Sasaki, Jeffrey M. Wooldridge, and all the internal seminar participants at the Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO), and the University of Tokyo. This paper is the research output of the IDE-JETRO's FY2014–15 research project entitled "Transition from School to Work: Linkage of Education and Employment in India." I am grateful for the funding support from the IDE-JETRO

1. Introduction

It is widely known that India's recent economic growth has been fueled by service sectors such as business services (including software and information technology (IT)-enabled services), communications, and banking. More-traditional services, including hotels and restaurants, education, health, and trade and transport, have also undergone rapid growth (Eichengreen and Gupta, 2011). Since 1999 (up to 2014), service industries' share of the gross domestic product (GDP) has exceeded 50% in India (World Bank, 2015). This share seems large if India's development stage is considered. India's share of services in the GDP outweighed its predicted value for India's income by 6 percentage points on average from 1999 to 2014.[1]

One possible cause for India's service-led growth is its skill-sorting pattern. As Kamath (2011) and Sohoni and Kathuria (2014) showed, many highly talented graduates who studied engineering at the Indian Institutes of Technology (IIT), the most distinguished institutions of higher education in India, go on to choose non-engineering occupations, such as IT, finance, and consultancy services. For example, only 33% of students at IIT Bombay took engineering jobs in 2013 (Sohoni and Kathuria, 2014). Those highly intelligent students should foster India's manufacturing industries, but they do not; instead, they contribute to strengthening the competitiveness of India's service sector. This skill-sorting trend operates across all of India too. Because of individual-level skill sorting, the workforce's educational level tends to be higher in many service industries. In Figure 1, industries are sorted by the estimated number of completed years of education averaged over an industry's male regular wage/salaried (RWS) workers. The average education level of an industry's workforce falls in moving

[1] India's predicted value for the share of services in GDP is computed based on the cross-country OLS regression of share of services in GDP on a quartic polynomial in log of GDP per capita for the period 1960–2014 over 215 economies. Data are extracted from the World Bank (2015). The service industries correspond to divisions 50–99 of the International Standard Industrial Classification (ISIC), revision 3. Note that they do not include construction, electricity, water, and gas, which are classified as services in my study.

from left to right across the graph. Therefore, it is clear from the graph that many service industries, such as insurance (industry number 53), education and research (54), finance (banking, etc.; 52), medical and health services (55), post and telecommunication (51), public administration and defense (57), railway transport (49), electricity (44), and other services (56), are successful in attracting relatively higher-educated workers.

Why does this skill-sorting pattern occur? Kamath (2011) and Sohoni and Kathuria (2014) gave wage differentials as one possible cause. However, why are wages lower for engineering jobs compared to other service-sector jobs, such as business services and banking?

This study offers one possible answer to this question. I hypothesize that India's skill-sorting patterns and skill wage differentials are a result of interactions among India's unequal skill distribution, low input quality, and variations in industries' production chain lengths. First, producing manufacturing goods tends to require more intermediate inputs than needed to produce service goods. In other words, manufacturing industries tend to form longer production chains than service industries. This tendency is confirmed from the line graph in Figure 1 (or Figure 3 in Section 6). Second, similar to the O-ring theory by Kremer (1993), the quality of final good deteriorates more as more intermediate inputs are involved as a result of increased defect rates. For example, if the probability of a malfunction in each part is 1%, then that of a product composed of two units of the part becomes 1.99% (=1-0.99*0.99)*100). As will be discussed below, the magnitude of such quality deterioration is likely to be larger in a country such as India. In this case, wages in manufacturing industries that require many intermediate inputs are dragged down significantly because of substantial quality deterioration compared to wages in service industries. Consequently, high-skilled individuals choose not to work in manufacturing industries where they cannot earn wages worthy of their skills, instead choosing to work in service industries.

The quality deterioration concomitant with an increase in intermediate inputs is expected to be much more severe in developing countries such as India, which are characterized by a large pool of unskilled labor, poor infrastructure (e.g., unstable electricity supply and bumpy roads), and less-advanced technology. These factors all contribute to low-quality intermediate inputs and much higher defect rates when many inputs are combined. As of 2009–2010, 41% of India's working population are either illiterate or literate, but they have either never received formal education or failed to complete primary education (based on weekly status computed from NSSO (2009–2010)). According to the 2014–2015 Global Competitiveness Report published by the World Economic Forum, India ranks 103rd in terms of quality of electricity supply and 76th in terms of quality of roads among the 144 countries included in the report (World Economic Forum, 2014). In addition, India's rank in terms of local supplier quality, availability of latest technologies, and firm-level technology absorption is 78th, 110th, and 102nd, respectively, indicating the generally low quality of intermediate inputs in India. Although it is difficult to directly measure the quality of intermediate inputs in India, some examples can be offered here. For example, UNIDO (2010: p.7) cites a survey conducted by A.T. Kearny, who found that defect rates in the Indian auto component industry are in the range of 1000–2000 parts-per-million (ppm), whereas those of Japanese average around 100–200 ppm. World Bank (2004: p.52) also mentioned the low quality and quality inconsistency of India's textile and clothing products, which is likely to be a consequence of a fragmented production process spread over many small-scale units.

This study contributes to the literature in two main ways. First, it proposes a new mechanism that explains India's skill-sorting pattern, which is, in turn, likely to contribute to India's service-led growth. Second, this study theoretically and empirically presents a new mechanism that explains skill-sorting patterns and skill wage differentials across industries based on the length of an industry's production chains (or the amount of necessary intermediate inputs used to produce an industry's output). I present a simple model to show that when the quality of intermediate inputs deteriorates rapidly, exceeding the increasing speed of marginal revenue of skill, workers' skill is negatively associated with the length of an industry's production chains. In other words, higher-skilled individuals choose to work in industries with shorter production chains. In this study, I call this skill-sorting pattern negative sorting, and the opposite pattern is called positive sorting. Negative (positive) sorting occurs because wage returns to skill are higher in industries with shorter (longer) production chains. Using India's data for year 1999, I empirically confirm the existence of negative sorting because of seeking a higher return to skill in India. The results remain robust even when correcting for possible selection bias, controlling for alternative reasons for inter-industry skill wage differentials, and examining the 2009 data alone.

The rest of the paper is organized as follows. Section 2 overviews related studies and the ways in which this study contributes to the literature. Section 3 presents a simple simultaneous production model in which skill-sorting patterns depend on the length of industry production chains and intermediate input quality. Section 4 explains the empirical strategy. Section 5 describes the data sources and the construction of key variables. Section 6 presents the main estimation results for India in 1999. Section 7 provides robustness checks, and Section 8 concludes.

2. Related Literature and Contribution of This Study

First, this study is closely related to studies that examine how high-skilled workers are matched with other workers or intermediate inputs (Kremer 1993; Lucas 1978; Murphy, Shleifer, and Vishny 1991; Rosen 1982; Sampson 2013). All studies are theoretical. Thus, I can first contribute to the literature by providing empirical evidence.

Moreover, most studies present models in which high-skilled workers are matched with larger amounts of labor or intermediate inputs. For instance, most talented persons produce products that require more tasks (and thus, require more high-skilled workers) or work at later stages of sequential production (Kremer 1993), manage firms with larger numbers of employees (Lucas 1978; Murphy, Shleifer, and Vishny 1991; Rosen 1982), or work with larger amounts of high-quality intermediate inputs (Sampson 2013). My study shares common features with Kremer (1993) in terms of introducing quality deterioration as the number of inputs increases and with Sampson (2013) in terms of examining the matching of workers' skills with the quantity and quality of intermediate inputs. However, my study shows that an opposite matching pattern is possible: high-skilled workers can work with a smaller amount of intermediate inputs when the quality of intermediate inputs deteriorates substantially along the production chains.

Second, in terms of results, my study also shares certain features with Grossman (2004) and Asuyama and Goto (2015). Grossman (2004) built a two-sector model to show that most talented individuals choose the so-called "software" sector, in which they can work alone and get paid according to their own productivity. They are disinclined to work in a team production sector ("automobile" sector) in which the wages of high-skilled workers are dragged down by low-skilled team members because of imperfect labor contracts. I differ from Grossman (2004) in terms of introducing intermediate inputs, building a different multi-sector model, and providing an empirical analysis. Asuyama and Goto (2015) theoretically showed that high-skilled individuals choose to work at earlier production stages when the quality of intermediate input deteriorates rapidly or improves slowly with each production stage. Based on cross-country industry panel data, they also empirically confirmed their model's prediction. However, their model is based on sequential production, while mine considers simultaneous production. In addition, their data is less fine compared to mine. They use industry-level data, while I examine both industry- and individual-level data.

In addition, their classifications of industry and skill are much broader.[2] They do not provide any analysis on inter-industry skill wage differentials. Focusing on one country (i.e., India), I examine skill-sorting patterns and inter-industry skill wage differentials more rigorously in this paper.

Finally, this study is also related to studies based on Roy's model (Roy, 1951), which explains skill wage differentials and skill allocation across sectors. Different wage returns to observed or unobserved skills across sectors, such as industries or occupations, have been found in several empirical studies (Gibbons, Katz, Lemieux, and Parent, 2005; Heckman and Scheinkman, 1987; Keane and Wolpin, 1997; Pavcnik, Blom, Goldberg, and Schady, 2004). Pavcnik et al. (2004) speculated that returns to skill differ across sectors because (1) labor mobility, (2) the ability to bargain over wages, and (3) monitoring costs and the necessity of paying efficiency wages all differ between high- and low-skilled workers. Roy's self-selection framework offers another explanation. In this framework, workers are endowed with multiple sector-specific skills and can have only one job. In this situation, workers self-select into jobs based on their comparative advantage. That is, they choose occupations that offer higher returns to a skill with which they are relatively well endowed. Autor and Handel (2013) and Yamaguchi (2012) obtained some empirical support for this mechanism. Most Roy-type studies state that skills are differently rewarded across sectors because skill requirements vary among them. My study offers an alternative mechanism, that is, returns to skill can vary among sectors because of differences in production chain lengths.

3. Model

[2] The number of industries is around 35. In addition, only three skill categories based on education level are available.

Based on a simultaneous production model, this section analyzes how skill-sorting patterns are affected by intermediate input quality and production chain length. My model builds upon the O-ring theory by Kremer (1993) in terms of introducing quality deterioration with an increase in inputs and upon Sampson (2013) in terms of introducing the quantity and quality of intermediate inputs.

Consider a perfectly competitive economy with multiple industries. The output of a certain industry is produced by many identical production units, each of which comprises one individual with a skill (or productivity) of $\theta \in [0,1]$ who is working on *n* units of intermediate inputs. There exists only one type of intermediate input, which is the composite of various inputs.[3] Each industry is only distinguished by *n* (amount of necessary intermediate inputs), which I call the industry's length of production chains.

As with Sampson (2013), because of the zero-profit condition in a perfectly competitive market, each worker's wage equals the profit of his/her production unit, which is equal to revenue minus the cost of intermediate inputs. Then, by solving the wage maximization problem below, an individual will choose to work in industry *n** where he/she can receive the highest wage:

$$\underset{n}{Max}\, W(\theta,n,q) = Q(q,n)V(\theta,n) - nq\,, \tag{3.1}$$

where $W(\theta,n,q)$ is the wage of a worker with skill θ if he/she chooses industry *n* and works on *n* units of intermediate inputs of quality *q*. $q \in (0,1]$ is the quality of one unit of intermediate input, which is assumed to be exogenously determined by various factors, including levels of human capital, technology, and infrastructure of the economy. $Q(q,n)$ stands for the quality of aggregated intermediate inputs when *n* intermediate inputs with quality *q* are used to produce output. $Q_q > 0$ is assumed.[4] Importantly, and

[3] Here, the proportion of various inputs is assumed to be the same among composite inputs, although this is not a realistic assumption. Incorporating heterogeneity of inputs is left for future research.
[4] The notation "Y_x" stands for partial derivative of *Y* with respect to *x*.

similar to the O-ring theory by Kremer (1993), I assume that Q is decreasing in n ($Q_n < 0$) because the possibility of defects increases as more inputs are involved. For example, consider the quality of a car. If the failure rate of brakes and engines are 1% each, the probability of a car that incorporates both parts failing to work properly becomes 1.99% (= [1−0.99*0.99]*100). In terms of defect rate, the car's overall quality (final output quality) becomes worse than the quality of each part. Even if the quality of each input is perfect (q = 1), a negative Q_n can still occur if the assembly process itself entails quality deterioration that becomes more severe with an increase in inputs. Similarly to Kremer (1993), output price and quantity are combined in one function, $V(\theta, n)$, which is the value of output achieved if intermediate input quality exerts no influence. I assume $V_\theta > 0$, $V_n > 0$ and that both Q and V are twice-continuously differentiable. QV is the total revenue of the production unit. The price of one unit of intermediate input with perfect quality (that is, q = 1) is standardized to one. Thus, the cost of intermediate inputs is expressed by nq. Finally, I assume $Q_{nn} < 0$ and $V_{nn} < 0$.[5]

The first-order condition for the worker's maximization problem becomes

$$W_n = Q_n V + Q V_n - q = 0. \tag{3.2}$$

Then, by the implicit function theorem,

$$\frac{dn^*}{d\theta} = -\frac{Q_n V_\theta + Q V_{n\theta}}{Q_{nn} V + Q V_{nn} + 2 Q_n V_n}. \tag{3.3}$$

Because of the assumptions of $Q_{nn} < 0$, $V_{nn} < 0$, $Q_n < 0$, and $V_n > 0$, the denominator, which is the left-hand side of the second-order condition, is negative. If $dn^*/d\theta$ is negative, it implies that if an individual's skill level is higher, the amount of intermediate inputs he/she chooses to work with is lower (i.e., the industry's length of production chains is shorter). Such negative sorting ($dn^*/d\theta < 0$) occurs only when

[5] These two assumptions are not necessary but are sufficient conditions for the second-order condition to be satisfied.

$Q_n V_\theta + Q V_{n\theta} < 0$, or equivalently

$$\frac{V_{\theta n}}{V_\theta} < -\frac{Q_n}{Q}. \qquad \text{(3.4: negative sorting condition)}$$

By assumption, $Q_n < 0$, $V_\theta > 0$, and Q > 0. The sign of $V_{\theta n}$ can be either positive (when V is supermodular in workers' skill and amount of intermediate inputs) or negative (when V is submodular). When V is supermodular ($V_{\theta n} > 0$), this condition shows that when the speed of quality deterioration along the production chains exceeds the increasing speed of marginal revenue of workers' skill (V_θ), negative sorting occurs. On the other hand, if the magnitude of quality deterioration is not sufficiently large, positive sorting occurs. When V is submodular ($V_{\theta n} < 0$), equation (3.4) always holds regardless of the degree of quality deterioration. Even when no quality deterioration occurs ($Q_n = 0$), negative sorting can occur if V is submodular because the marginal revenue of a worker's skill falls as the production chain lengthens. Finally, it should be noted that regardless of the sign on $V_{\theta n}$, high-skilled individuals choose smaller-*n* industries because these industries offer the highest wages. For low-skilled individuals, however, a larger-*n* industry offers higher rewards. In this sense, returns to skill (or skill wage premiums) are larger in industries with shorter production chains, when negative sorting occurs under condition (3.4).

4. Empirical Strategy

This section tests (1) whether negative sorting occurs in India and (2) whether the returns to skill are larger in industries with shorter production chains. The model in the previous section indicates that when sorting depends on input quality (that is, when V is supermodular), negative sorting is more likely to be observed in developing countries such as India, where it can be expected that the input quality will fall rapidly as a result of many unskilled workers, low technology, and poor infrastructure.[6] In addition, as the skill distribution throughout the economy becomes more unequal, this skill-sorting trend emerges more sharply. In this sense, India is one of the most appropriate fields to examine skill-sorting patterns given its relatively large number of highly educated population as well as its large pool of unskilled workers. In 2009−2010, the proportion of workers who had received no schooling or failed to complete primary education was 41% (as mentioned above), compared to the 15% who had completed secondary education (computed based on NSSO (2009–2010)).

If V is submodular, negative sorting occurs regardless of the input quality in the economy. However, this submodular-V case can be excluded using the evidence from Asuyama and Goto (2015), which shares a similar hypothesis to that of this paper. Based on cross-country industry panel data, they empirically show that the economy's skill-sorting pattern depends on the input quality in the economy. Their results show that negative sorting occurs only in economies in which the rate of change in intermediate input quality along production chains is small (either quality deteriorates rapidly or improves slowly). Regardless of specifications, their results imply that negative sorting occurs in India, where input quality degrades rapidly (or improves slowly). However, the data from Asuyama and Goto (2015) is less fine compared to the data in the current paper (see Section 2). Thus, I examine India's skill-sorting pattern more rigorously in this paper. I also examine whether skill sorting is affected not only by production chain length but also by other supplemental quality indicators of intermediate inputs in order to test whether India's skill-sorting pattern depends on input quality.

[6] As explained in the car example in Section 3, it is natural to expect that low quality of an intermediate input (low q) leads to a large negative value of Q_n and that low q is caused by unskilled labor, low technology, and poor infrastructure, such as unstable electricity supply and bumpy roads.

4.1 Skill-Sorting Regression

To test whether negative sorting is observed in India, the following three equations are estimated using weighed least squares (WLS). The weight used is the survey weight of the dataset (equation (4.1)), the employment size of each industry (equation (4.2)), and the employment share of each industry averaged over the two periods (equation (4.3)), respectively:

$$ChainL_{ijt} = \alpha_{1t} + \beta_{1t} Skill_{ijt} + \gamma_{1t} X_{ijt} + \delta_{1t} ChainL\_sfamily_{ijt} + \varepsilon_{1ijt}\,, \qquad (4.1)$$

$$Skill_{jt} = \alpha_{2t} + \beta_{2t} ChainL_{jt} + \gamma_{2t} ChainQ_{jt} + \delta_{2t} Z_{jt} + \varepsilon_{2jt}\,, \qquad (4.2)$$

$$Skill_{jt} = \alpha_3 + \beta_3 ChainL_{jt} + \gamma_3 ChainQ_{jt} + \delta_3 Z_{jt} + F_t + F_j + \varepsilon_{3jt}\,, \qquad (4.3)$$

where subscripts *i*, *j*, and *t* denote worker, industry, and time period (mainly 1999 and 2009 as a robustness check), respectively. In every equation, ε stands for the error term. In equation (4.1), $ChainL_{ijt}$ measures the length of domestic production chains of industry *j* with which individual *i* is affiliated. $Skill_{ijt}$ stands for *i*'s skill level. X_{it} denotes a vector of individual characteristics, which includes estimated years of work experience and its square, dummies for being a Muslim, social groups, household head, marriage status, residence in rural area, and Indian states in which an individual lives.[7] $ChainL\_sfamily_{ijt}$ is the average $ChainL_{ijt}$ of other family members of the same gender. This is included because in India, an individual's job choice is assumed to be substantially affected by the jobs of same-gender family members.[8] Equation (4.1) is estimated separately for each period and at the individual level. If $\beta_{1t} < 0$ is observed, it indicates negative sorting; that is, high-skilled individuals choose industries with shorter production chains.

Because industry characteristics other than the length of production chains

---

[7] When the elements constructing my skill index (see Section 5.2.1) include work experience (i.e., in cases of *SK2* and *SK3*), experience and its square are not controlled for. Such a treatment also applies when testing inter-industry skill wage premium later.

[8] I also experiment with controlling for the average *ChainL* of other same-gender workers in the same district in addition to *ChainL_sfamily*, but the main results do not change.

cannot be controlled for in equation (4.1), equation (4.2) is also estimated separately for each period but at the industry level.[9] $Skill_{jt}$ and $ChainL_{jt}$ stand for the average skill level of workers and the length of domestic production chains for industry *j* at time *t*, respectively. If negative sorting occurs, workers' skill level should be higher in industries with shorter production chains. In this case, β_{2t} should be negative.

$ChainQ_{jt}$ represents a vector of supplemental quality indicators of intermediate inputs that are not captured by $ChainL_{jt}$. It includes an industry's dependence on imported inputs ($ChainQ\_Import_{jt}$) and the skill level embodied in inputs from other industries ($ChainQ\_Skill_{jt}$). $ChainQ_{jt}$ is controlled for because in reality, the input composition differs across industries. Consequently, the size of quality deterioration is determined by not only production chain length but also input composition and quality. It is expected that the quality of imported inputs is higher than that of domestic inputs. In addition, input quality can be measured by the skill level of workers who produce said inputs. A statistically significant value for γ_{2t} implies that a skill-sorting pattern depends on input quality (supermodular-*V* case). A positive value for γ_{2t} is expected because when comparing industries with the same $ChainL_{jt}$, input quality deterioration can be expected to be smaller in a higher-$ChainQ_{jt}$ industry, which consequently attracts more skilled workers.

Z_{jt} denotes a vector of industry characteristics. It includes degrees of imports and exports of final goods. It also includes employment ratios of small firms, which approximate an industry's degree of informality.

Finally, industry-level skill-sorting equation (4.3) is estimated by pooling two-year samples and adding time and industry fixed effects (F_t and F_j, respectively). F_j controls for all time-invariant industry characteristics, some of which are not

[9] Here, I do not aim to identify causality but try to confirm the association between worker's skill level and production chain length. Thus, either *Skill* or *ChainL* can become the dependent or independent variable.

controlled for in equation (4.2). Again, a negative β_3 and positive γ_3 can be expected.

4.2 Skill Wage Premium Regression

To test whether the return to skill is larger in industries with shorter production chains, the following augmented Mincer-type wage equation (Mincer, 1974) is estimated by WLS with the survey weight:

$$\ln Wage_{ijt} = \alpha_{4t} + \beta_{4t} Skill_{ijt} * ChainL_{ijt} + \gamma_{4t} Skill_{ijt} * ChainQ_{ijt} + \delta_{4t} Skill_{ijt} * Z_{ijt} + \eta_{4t} Skill_{ijt} + \lambda_{4t} X_{ijt} + F_j + \varepsilon_{4ijt}, \qquad (4.4)$$

where $\ln Wage_{ijt}$ denotes the logarithm of a worker's wages in industry *j* at time *t*. If β_{4t}, the coefficient of the interaction term between $Skill_{ijt}$ and $ChainL_{ijt}$ is negative, it indicates that returns to skill are higher in industries with shorter production chains. As in the skill-sorting regressions, supplemental quality indicators of the chain ($ChainQ_{ijt}$) are also controlled for. A positive γ_{4t} is expected because it is expected that input quality deteriorates with a decrease in $ChainQ_{ijt}$. Similar to the effect of $ChainL_{ijt}$, this effect of quality deterioration can lead to a lower return to skill.

To control for other reasons causing inter-industry skill wage differentials, a vector Z_{ijt} is also interacted with $Skill_{ijt}$. It contains individual characteristics, such as union affiliation, employment nature (permanent or temporary), and affiliation with public and small firms. These factors are included because it can be expected that a skill wage premium applying to union members or public-sector workers would be smaller because the effect of market forces on the wage-setting mechanism might be weaker in these sectors. Based on empirical evidence from India (Azam, 2012; Dutta, 2006), it can also be expected that the skill wage premium is lower in informal sectors characterized by temporary employment or small firm size.

X_{it} is a vector of individual characteristics such as those included in equation (4.1) as well as occupation, union affiliation, employment nature, social security status, and affiliation with a public and small firm.[10] Controlling for these variables is particularly important in India, where a large informal sector exists. Industry affiliation dummies (F_j) are also included. F_j absorbs the industry-wage premium that is common for all workers regardless of their skill level. Finally, ε_{4ijt} denotes the error term.

The above identification strategy is possible because individuals with the same skill level choose different industries and receive different wages in India's dataset. This fact can be explained by the model developed by Dahl (2002), in which individuals maximize their utility, which is a function of earnings and tastes.[11] Thus, in reality, wages and individuals' preferences regarding job characteristics affect their industry choice. As a result, perfect negative sorting is not observed, and inter-industry (or more precisely, inter-*ChainL*) skill wage differentials can be identified.

5. Data

5.1 Main Data Sources and Sample Used

Data on individual-level variables, including wages, skill level, and other characteristics, as well as several industry-level variables, such as workers' skill level of industry, are primarily constructed from the unit-level data of the Employment and Unemployment schedules of the National Sample Survey (NSS). As Kijima (2006) states, the Employment and Unemployment schedule of the NSS is the only survey that collects information on individual's earnings, employment status, and other characteristics for all of India through a stratified random sampling procedure.

[10] Similar to the treatment in footnote 7, if the skill index is constructed from education, experience, and occupational information (i.e., in case of *SK3*), experience and occupations are not controlled for when estimating equation (4.4).
[11] Dahl (2002) theoretically and empirically examined individuals' patterns of self-selection and the difference in return to education across U.S. states.

Industry-level data, including the length of production chains, dependence on imported inputs, and import or export ratio, are constructed from the input-output (IO) tables for India.

I primarily examine the matched data on the 55th round of NSS data conducted in 1999–2000 and the 1998–1999 IO table (NSSO, 1999-2000; CSO, 1998-99). I call this period 1999. Year 2009 data, which is the matched data on the 2009–2010 (66th round) NSS and the 2007–2008 IO table (NSSO, 2009-2010; CSO, 2007-08), is also examined in the robustness analysis and used in the industry-panel skill-sorting regression (equation (4.3)). There are several reasons for primarily analyzing the 1999 data. First, the NSS sample size is much larger in this period, and consequently, the most finely grained industry classifications can be achieved.[12] Second, before this round, the NSS did not collect some important information, such as firm size and social security status. Third, after this round, it is not possible to construct accurate skill indices that incorporate occupational information because NSS occupational classifications become much broader. Finally, both *ChainL* and *ChainQ* become less accurate in subsequent years.[13]

There are 115 sectors in the 1998–1999 IO table. I match these sectors with the NSS's five-digit industry codes as closely as possible based on the descriptions of both sectors and industries. I also ensure that there is a sufficient number of observations (around 100 sample workers) with non-missing wage information for each industry because the within-industry wage gaps between skill groups is estimated in equation (4.4). As a result, 57 industries are created.[14] The concordance table on industry

[12] The numbers of households and individuals covered by the NSS are 115,409 and 564,740 (1993–1994), 165,244 and 819,013 (1999–2000), 124,680 and 602,833 (2004–2005), and 100,957 and 459,784 (2009–2010), respectively.

[13] This is because the construction of these two indices involves estimation based on the 1993–1994 IO table, for which the industry classification is the same as the 1998–1999 IO table but different from the 2007–2008 IO table (see Sections 5.2.2 and 5.2.3).

[14] When year 2009 data is examined as a robustness check in Section 7.3, 54 industries are created as a result of the same procedure. For the details of these 54 industry classifications, see

classification between IO tables and the NSS is provided in Appendix Table B.3(a).

I restrict the estimation sample to male, prime-age (15–65 years old), regular wage/salaried (RWS) employees who have worked full time and are not currently attending an educational institution. Following Kijima (2006), full-time workers are defined as those who have worked at least five days at their main economic activity during the reference week.[15] Because the actual hours worked are not asked in the NSS, wages are defined as the weekly wage and salary earnings (either in cash or in kind, including bonus and perquisites) for the main economic activity.

5.2 Construction of Key Variables

This subsection describes how I construct key variables, namely $Skill_{ijt}$, $ChainL_{jt}$, and $ChainQ_{jt}$. A brief description of other variables and the summary statistics are presented in Appendix A.

5.2.1 *Skill*: Workers' Skill

Workers' skill ($Skill_{ijt}$) is measured by the following three (or four) indices:

- **Skill index 1 (*SK1*), which equals estimated years of education**: Measuring individuals' skills by their levels of educational attainment is the most conventional method used in the labor and macroeconomics literature (Ingram and Neumann, 2006: p.37). Seven educational levels (illiterate; literate without formal schooling or

---

Appendix Table B.3(b).

[15] Wages, RWS working status, number of days worked, and industry and occupation affiliations are based on the status during the reference week. However, other individual characteristics, such as union affiliation, employment nature, social security status, and affiliation with a public firm and small firm, are only available for so-called "usual status," which is based on a reference period of one year. Thus, when controlling for these characteristics, I restrict the sample to individuals whose RWS working status, five-digit industry code, and three-digit occupation code are the same between weekly and usual (yearly) statuses. I also restrict this sample by only focusing on individuals who explicitly claim that they had no months without work during the reference year. Under this restriction, I assume that individuals' jobs based on weekly and yearly status are the same.

literate but have not completed primary education; primary; middle; secondary; higher secondary; graduate and above) can be identified in NSS 1999–2000. The years of education are estimated from these seven categories by allocating the corresponding number of years of schooling to each level respectively (0, 2.5, 5, 8, 10, 12, and 15 years).

- **Skill index 2 (*SK2*), which is constructed from education and experience**: An alternative skill index is constructed following Gibbons et al. (2005: p.698). First, a logarithm of wages is regressed on education category dummies, X_{it}, in equation (4.1) (or equation (4.4), depending on the specification of the regressions considered) and industry affiliation dummies.[16] Then, skill index 2 (*SK2*) is estimated by predicting the wage of each worker based solely on the worker's education and experience. Because the numbers of years of work experience are not available in NSS data, they are estimated by subtracting [estimated years of education plus 5] from age, following Kijima (2006). Finally, this skill index is normalized to have a zero mean.
- **Skill index 3 (*SK3(1)* and *SK3(2)*), which is constructed from education, experience, and task content of occupation**: Skill index 3 (*SK3*) is computed similarly as the predicted wage based on education, experience, and task content measure of occupations.[17] Autor and Handel (2013) showed that workers self-select into occupations that offer high returns to tasks in which they are relatively well endowed. Thus, I assume that occupation task content can be used as a proxy for a

[16] When examining *SK2* in skill-sorting regression, X_{it} of equation (4.1) is used. In case of a skill wage premium regression, X_{it} of equation (4.1) is used when occupations, union affiliation, employment nature, social security status, and affiliation with a public and/or small firm are not controlled for in the regression. When these variables are controlled for in the skill wage premium regression, X_{it} of equation (4.4) is used.

[17] Before prediction, the logarithm of wages is regressed on education category dummies, task content measures of occupations, X_{it} in equation (4.1) (or equation (4.4)) without occupation dummies, and industry affiliation dummies.

worker's skill in conducting these tasks. I experiment with two task content measures of occupation. The first measurement used to construct *SK3(1)* is cognitive and motor task complexity of occupations, which was extracted from Yamaguchi (2012). The second measurement used for *SK3(2)* is routine, abstract, and manual task intensity of occupations, which was extracted from Autor and Dorn (2013) and Dorn (2009). Because both measurements are constructed based on the task content of occupations in the United States (U.S.) around 1991 or 1977, they might be crude measures for the task content of India's occupations. However, as long as some commonality can be expected between occupation-specific skill requirements in the U.S. and India, *SK3* can serve as an appropriate proxy for $Skill_{ijt}$. Based on occupation content, I match occupation codes between India and the U.S. as closely as possible (Appendix Table B.4(b)). Then, occupation-specific task content measures for the U.S. are assigned to each occupation in India. Because the occupation codes in the 2009–2010 NSS are much broader than those in the 1999–2000 NSS, *SK3* is constructed only for 1999.

5.2.2 *ChainL*: Length of Production Chains

The length of production chains of industry *j* ($ChainL_{jt}$) is computed in a manner similar to that used in Asuyama (2012). In general, $ChainL_{jt}$ is the column sum of the Leontief inverse coefficient of industry *j* computed from the aggregated 57×57 sector IO table as follows:

$$ChainL_{jt} = \sum_k leon_{kjt} , \quad (5.1)$$

where $leon_{kjt}$ is the (*k*, *j*)th entry of the Leontief inverse coefficient matrix, *L*.[18] $ChainL_{jt}$ measures the amount of domestic intermediate inputs that industry *j* requires,

[18] $L = (I - A_d)^{-1}$, where *I* is the identity matrix and A_d is the input coefficient matrix for domestic input whose (*k*, *j*)th entry is a_{kj}, which measures the amount of domestic input from industry *k* directly used to produce one dollar's worth of industry *j*'s output.

both directly and indirectly, to produce one dollar's worth of that industry's output. It stands for the scope of production linkages with domestic intermediate input industries. Imported inputs are excluded from the calculation of $ChainL_{jt}$ because imported inputs are likely to be of higher quality than domestic inputs.

To calculate $ChainL_{jt}$ for domestic inputs by equation (5.1), the total intermediate inputs need to be separated into domestic and imported categories and an IO table needs to be created on the basis of domestic input only. This is possible for the IO table from 1993–1994, when an import flow matrix is available. However, import flow matrices are not available for the 1998–1999 and 2007–2008 IO tables. Thus, $ChainL_{jt}$ for the 1998–1999 IO table is estimated as

$$\begin{aligned} ChainL_{j,t=98} &= ChainLT_{j,t=98} * [ChainL_{j,t=93} / ChainLT_{j,t=93}] \\ &* [ChainL_{j,t=98,WIOD} / ChainLT_{j,t=98,WIOD}] / [ChainL_{j,t=95,WIOD} / ChainLT_{j,t=95,WIOD}], \end{aligned} \tag{5.2}$$

or this can alternatively be written as

$$\begin{aligned} &ChainL_{jt} \text{ in 1998} = ChainLT_{jt} \text{ in 1998} * [ChainL_{jt} / ChainLT_{jt}] \text{ in 1993} \\ &\quad * \textit{change of } [ChainL_{jt} / ChainLT_{jt}] \textit{ between the two period}, \end{aligned}$$

where $ChainLT_{jt}$ stands for the length of the production chains computed based on total (domestic plus imported) inputs. The subscripts t = 93, 95, and 98 in equation (5.2) denote the corresponding year (1993–1994, 1995, and 1998(–99)) of the IO tables. The subscript "WIOD" means that the corresponding IO table is extracted from the World Input-Output Database (WIOD) (Timmer et al., 2015). Although its industry classification is much broader, the WIOD provides IO tables that separate domestic and imported inputs.[19] Because the WIOD's earliest year is 1995, I approximate the change of [$ChainL_{jt} / ChainLT_{jt}$] between 1993 and 1998 based on that between 1995 and 1998,

[19] In particular, I extract India's National Input-Output Tables, which were released in September 2012 by WIOD. WIOD adopts a 35-industry classification. Thus, as Appendix Table B.3(a) shows, [$ChainL_{jt} / ChainLT_{jt}$] of one WIOD industry is often applied to several industries based on my 57-industry classification.

assuming that the industrial structure does not change substantially within these few years. $ChainL_{jt}$ for the 2007–2008 IO table is computed in a similar manner by replacing the subscript "98" of equation (5.2) with "2007."

5.2.3 *ChainQ*: Supplemental Quality Indicators of Intermediate Inputs

First, an industry's dependence on imported inputs, $ChainQ\_Import_{jt}$, is computed for the years 1993–1994 as follows[20]:

$$ChainQ\_Import_{jt} = ML, \qquad (5.3)$$

where M is the 1×57 vectors whose jth entry is j's imported input ratio to output. L is the 57×57 Leontief inverse coefficient matrix computed from the 1993–1994 IO table for domestic inputs. Similar to the estimation method used to calculate $ChainL_{jt}$, the $ChainQ\_Import_{jt}$ for the 1998–1999 IO table is estimated as

$$\begin{aligned} &ChainQ\_Import_{j,t=98} \\ &= ChainQ\_Import_{j,t=93} * [ChainQ\_Import_{j,t=98,WIOD} / ChainQ\_Import_{j,t=95,WIOD}]. \quad (5.4) \end{aligned}$$

$ChainQ\_Import_{jt}$ for the 2007–2008 IO table is also computed in a similar manner.

Another quality indicator is $ChainQ\_Skill_{jt}$, the skill level of workers embodied in inputs from other industries. It is computed as follows:

$$ChainQ\_Skill_{jt} = \left(\sum\nolimits_{k \neq j} Eduy_{kjt} * leont_{kjt}\right) / \sum\nolimits_{k \neq j} leont_{kjt}, \qquad (5.5)$$

where $Eduy_{kjt}$ is the average years of education of all workers in industry k whose output is used as an input in industry j. Thus, $ChainQ\_Skill_{jt}$ is the average number of years of education embodied in inputs from other industries, weighted by k (input industry)'s share in the entire production chain lengths. $leont_{kjt}$ is the (k, j)th entry of the Leontief inverse coefficient matrix computed based on total inputs. Ideally, it should be computed based only on domestic inputs, but it was not available for the 1998–1999 and 2007–2008 IO tables. Thus, $leont_{kjt}$ is computed based on total inputs as the next

[20] I thank Satoshi Inomata for his advice on the construction of *ChainQ_Import*.

best method.

6. Estimation Results for 1999

6.1 Skill-Sorting Regression

Figure 2 (individual level) and Figure 3 (industry level) present the raw correlations between workers' skills expressed by the four skill indices and the industry's production chain lengths in 1999. These two figures generally show that high-skilled individuals work in industries with shorter production chains; that is, negative sorting seems to occur in India. It is also evident from Figure 3 that production chains tend to be shorter in service and primary industries than in manufacturing industries.[21]

The question of whether this simple correlation remains robust even when controlling for other factors is examined by estimating equations (4.1)–(4.3). First, Table 1 reports the estimation results for equation (4.1), that is, the individual-level skill sorting equation in 1999. Consistent with this paper's negative-sorting hypothesis, the coefficient on skill index is significantly negative in almost all specifications regardless of skill indices, industry coverage, and control variables. I examine the manufacturing and service industry samples, which exclude primary industries such as agriculture and mining, because final product quality in primary industries is substantially affected by land, weather, and natural resources, which IO tables do not include as inputs. I also examine skill sorting within the manufacturing sector to exclude the possibility that differences in *ChainL* do not represent variations in production chain length but rather only capture differences between service and manufacturing sectors. However, it should be noted that the sample size becomes much smaller when restricting the sample to

[21] Exact *ChainL* figures (and *Skill*, *ChainQ_Import*, and *ChainQ_Skill*) for each industry are provided in Appendix Table B.1.

manufacturing. A smaller sample size results in larger standard errors in the estimated coefficients.

Table 2 and Table 3 report the estimation results for the industry-level skill sorting equation for 1999 (equation (4.2)) and for the 1999–2009 panel (equation (4.3)), respectively. In 1999, except the manufacturing-industry sample, negative sorting can generally be observed regardless of skill indices. The negative-sorting trend is much less clear in the manufacturing sample (Table 2). However, when controlling for time-invariant industry characteristics using 1999–2009 panel data, negative sorting becomes more evident regardless of variations in industry coverage (Table 3). Moreover, a significant and unexpectedly negative sign of *ChainQ_Import* in column (6) of Table 2 now becomes insignificant.

ChainQ_Skill is insignificant or positively associated with an industry's worker skill level (*Skill*) in the 1999 manufacturing sample. However, contrary to expectation, the sign on the coefficient of *ChainQ_Skill* in the manufacturing sample turns out to be negative in the panel regression (Table 3). This is because between 1999 and 2009, industries with higher growth in *ChainQ_Skill* experienced lower growth in *Skill*. One possible reason for this negative coefficient of *ChainQ_Skill* is that *ChainQ_Skill* of industry *j* does not capture the quality (embodied skill level) of inputs sourced from within its own industry (i.e., industry *j*). From the perspective of each worker in industry *j*, the quality of inputs sourced from within industry *j* also matters because it affects his or her wages. Thus, I construct $ChainQ\_Skilltotal_{jt} = (\sum_k Eduy_{kjt} * leont_{kjt}) / \sum_k leont_{kjt}$, where subscript *k* includes *j*. When controlling for this *ChainQ_Skilltotal* instead of *ChainQ_Skill*, the coefficients on *ChainQ_Skilltotal* become positive in the panel for skill-sorting regression (Appendix Table B. 2). However, because the share of inputs sourced from within its own industry out of the total inputs is large in general,[22] it is natural to expect that $ChainQ\_Skilltotal_{jt}$ is positively associated with $Skill_{jt}$, which is the average skill level of workers in that same industry. Constructing a more sophisticated index to measure the input quality that is not captured by *ChainL* is left for future research.

6.2 Skill Wage Premium Regression

Table 4 reports the estimation results for the skill wage premium regression (equation (4.4)), which includes only one interaction term with *Skill*, that is, *Skill*ChainL*. A negative coefficient on this *Skill*ChainL* is consistent with my hypothesis that negative sorting occurs in India because the returns to skill are higher in industries with shorter production chains.

The first column of every sample includes various individual characteristics based on weekly status. The second column controls for industry wage premium by adding industry dummies. The third column also controls for various job characteristics based on yearly status.

The coefficients on *Skill*ChainL* are negative in some specifications but are not so robust. The signs on coefficients for other control variables are almost consistent with the literature and general expectations. Variables positively associated with wages in general are skill index, experience, being household head, being married, having an occupation other than farmer, working at a public enterprise, and being covered under the Provident Fund (India's social security fund). In contrast, experience squared, being a Muslim, belonging to a disadvantaged social group, being a farmer, working in a temporary job, working in a small firm, and living in a rural area tend to be negatively associated with wages.

[22] For instance, the length of production chains taken up by the same industry ($leont_{jjt}$) accounts for 55% of the total production chain length ($ChainLT_{jt}$) on average across 57 industries in 1999.

Next, other factors that may explain inter-industry skill wage differentials are also controlled for by adding interaction terms between these factors and *Skill*. Table 5 reports the estimation results for all the interaction terms with *Skill*. Importantly, the coefficients on *Skill*ChainL* become negative in most specifications; that is, returns to skill are higher in industries with shorter production chains. Consistent with the study's hypothesis, returns to skill tend to increase when supplemental production chain quality indicators (*ChainQ_Import*: dependence on imported inputs; *ChainQ_Skil*: average skill level embodied in the inputs from other industries) are higher. This implies that input quality affects inter-industry skill wage differentials. As expected, the skill wage premium tends to be smaller in the public sector and in the informal sector, which is characterized by temporary employment and small-sized firms.

7. Robustness Checks

This section provides various robustness checks, particularly for skill wage premium regression, by (1) correcting for possible selection bias, (2) controlling for alternative reasons for inter-industry skill wage differentials, and (3) examining a different period (year 2009).

7.1 Selection problems

Three types of selection problems are involved in the previous skill wage premium estimation, which was based on the sample of RWS employees. The first is the selection into either working or non-working. The second is the selection into working as an RWS employee or a self-employed/casual worker. The third is the selection into each industry. The first and second selection problems are less critical because this paper's focus is inter-industry variations in workers' skill levels and the skill wage premium. Thus, the estimation results are not biased as long as the study's population of interest is considered to be RWS employees. Using an RWS-employee sample is most appropriate to examine the skill-sorting pattern of highly skilled workers in particular (as illustrated by the job selection example offered in the Introduction about the most-promising IIT graduates) because most high-skilled individuals choose to work as RWS employees. In the 1999–2000 NSS, the ratios of RWS employees, self-employed, and casual workers among working individuals who had completed college/university education or more are 59.1%, 38.1%, and 1.8%, respectively. Being self-employed is also popular. However, the NSS does not provide wage data for self-employed persons. It is also much harder to control for the diversified job characteristics of self-employed workers.

Self-selection into industry is more critical. Individuals' wages are not observed for all industries; they are only observed for the single industry an individual chooses. In other words, the group of observed individuals working in a certain industry is not a random sample of the population. This can lead to a biased estimate for β_{4t} in equation (4.4), which is the inter-industry skill wage differentials caused by varied production chain lengths. Because β_{4t} is the focus, this possible selection bias needs to be tackled.

To correct for the selection bias, I utilize the control function approach of Wooldridge (2015: pp. 430-432), who extended the method of Garen (1984).[23] As mentioned above, the choice of $ChainL_{ij}$ is not randomly assigned to the population. Thus, the observed coefficient of $Skill_{ij} * ChainL_{ij}$ can be expressed as individual-specific inter-industry (or more precisely, inter- $ChainL_{ij}$) skill wage differentials, $g_i = \beta_4 + v_i$, where β_4 is the population-average inter-$ChainL_{ij}$ skill wage differentials needed to be identified and $E(v_i) = 0$. Then, the most basic version

[23] I thank Jeffrey M. Wooldridge for providing me with the Stata code for the Table 1 in Wooldridge (2015).

of equation (4.4) can be re-written as[24]

$$\ln Wage_{ij} = \alpha_4 + g_i Skill_{ij} * ChainL_{ij} + \eta_4 Skill_{ij} + \varphi_4 ChainL_{ij} + \lambda_4 X_{ij} + \varepsilon_{4ij}, \quad (7.1)$$

where X_{ij} is the same vector as in equation (4.1), which includes the estimated years of work experience and its square as well as dummies for being Muslim, social groups, household head, marriage status, residence in rural area, and Indian states of residence.[25]

By substituting $g_i = \beta_4 + v_i$ into equation (7.1), the following is obtained

$$\ln Wage_{ij} = \alpha_4 + \beta_4 Skill_{ij} * ChainL_{ij} + \eta_4 Skill_{ij} + \varphi_4 ChainL_{ij} + \lambda_4 X_{ij}$$
$$+ v_i * Skill_{ij} * ChainL_{ij} + \varepsilon_{4ij}. \quad (7.2)$$

I assume that only $\ln Wage_{ij}$ and $ChainL_{ij}$ are endogenous and that $ChainL_{ij}$ can be expressed by equation (4.1):

$$ChainL_{ij} = \alpha_1 + \beta_1 Skill_{ij} + \gamma_1 X_{ij} + \delta_1 ChainL\_sfamily_{ij} + \varepsilon_{1ij}, \quad (4.1)$$

where $E(\varepsilon_{1ij} \mid 1, Skill_{ij}, X_{ij}, ChainL\_sfamily_{ij}) = 0$. $ChainL\_sfamily_{ij}$ (average $ChainL_{ij}$ of other family members of the same gender) should be strongly correlated with $ChainL_{ij}$ and uncorrelated with ε_{4ij}. I assume that v_i and ε_{4ij} are linearly related to ε_{1ij}, that is, $E(v_i \mid \varepsilon_{1ij}) = \pi_1 \varepsilon_{1ij}$ and $E(\varepsilon_{4ij} \mid \varepsilon_{1ij}) = \pi_2 \varepsilon_{1ij}$. I also assume that v_i and ε_{4ij} are independent of $(1, Skill_{ij}, X_{ij}, ChainL\_sfamily_{ij})$. Then, the equation to estimate the skill wage premium becomes

$$E(\ln Wage_{ij} \mid 1, Skill_{ij}, X_{ij}, ChainL\_sfamily_{ij}, ChainL_{ij})$$
$$= \alpha_4 + \beta_4 Skill_{ij} * ChainL_{ij} + \eta_4 Skill_{ij} + \varphi_4 ChainL_{ij} + \lambda_4 X_{ij}$$
$$+ \pi_1 \varepsilon_{1ij} * Skill_{ij} * ChainL_{ij} + \pi_2 \varepsilon_{1ij}. \quad (7.3)$$

Thus, β_4 can be identified by regressing $\ln Wage_{ij}$ on 1, $Skill_{ij} * ChainL_{ij}$, $Skill_{ij}$, $ChainL_{ij}$, X_{ij}, $\hat{\varepsilon}_{1ij} * Skill_{ij} * ChainL_{ij}$, and $\hat{\varepsilon}_{1ij}$, where $\hat{\varepsilon}_{1ij}$ is the residual from the regression of equation (4.1).

---

24 Subscript *t* is omitted. In order to take the control function approach, $ChainL_{ij}$ is controlled for instead of industry dummies.

25 Experience and its square are only included when *SK1* is used as a skill index.

The estimation result of equation (4.1) is presented in the third column of every sample in Table 1. The results of the *F*-test on the null hypothesis $\delta_1 = 0$ shows that $ChainL\_sfamily_{ij}$ is strongly associated with $ChainL_{ij}$. Estimation results for equation (7.3) are reported in Table 6. First, the results of the *F*-test for the joint significance of ($\hat{\varepsilon}_{1ij} * Skill_{ij} * ChainL_{ij}$, $\hat{\varepsilon}_{1ij}$) show that selection bias exists only when using *SK1* as a skill index. In cases with other skill indices, it is not necessary to correct for selection bias. Second, even in case of *SK1*, the selection-corrected coefficient of $Skill_{ij} * ChainL_{ij}$ is still significantly negative. Thus, the results obtained in Section 6 are robust even when self-selection into industry is considered.

7.2 Alternative Reasons for Inter-industry Skill Wage Differentials

As mentioned in Section 2, returns to skill can vary across industries not only because different production chain lengths among industries but also because (1) labor mobility, (2) ability to bargain over wages, and (3) monitoring costs and necessity to pay efficiency wages vary between high-skilled and low-skilled workers across industries (Pavcnik et al., 2004).

- **Differences in labor mobility between skill groups (*Mob*)**: In a standard competitive labor-market model with perfect mobility, returns to skill are equalized over different industries. As the labor mobility of a certain skill group becomes lower in certain industries, the wages paid to that group in these industries deviate from market wages. Consequently, returns to skill vary across industries. Thus, the difference in labor mobility between high-skilled and low-skilled workers in each industry (Mob_{jt}) should be controlled for. As a measure for Mob_{jt}, I use the labor-mobility gap between high-skilled and low-skilled individuals in the sample.[26] Labor mobility, which is computed based on NSS data, is measured by

[26] High-skilled workers are defined as those with lower secondary or above education (10 or more years of education completed) and low-skilled workers are those with less than a

the ratio of individuals who changed industry (in terms of two-digit NIC-1998 levels) during the two years before the date of the NSS survey.[27]

- **Difference in bargaining power over wages between skill groups (*Power*)**: When the ratio of high-skilled workers is higher in industries with shorter production chains, bargaining power over wages might be greater for high-skilled workers in these industries. If this is the case, the higher skill premium in industries with shorter production chains is caused by not only input quality but also the relative bargaining power between skill groups. To exclude this possibility, the variable $Power_{jt}$ is constructed using the NSS data as the ratio of the number of union members among high-skilled workers to that of low-skilled workers in each industry.
- **Difference in monitoring costs (or efficiency wage) between skill groups (*EW*)**: Monitoring worker performance is easier for routinized tasks. In contrast, monitoring is costly for abstract or manual tasks that are harder to quantify. When monitoring is costly, firms can pay efficiency wages (which are higher than market wages) to prevent shirking (Shapiro and Stiglitz, 1984). If this is the case, in industries where the routine task intensity of high-skilled workers is higher than that of low-skilled workers, the need to pay efficiency wages to high-skilled workers is relatively lower. Consequently, the skill wage premium shrinks. Thus, the ratio of average routine task intensity between high-skilled and low-skilled workers of each industry (EW_{jt}) is controlled for.[28]

I control for the alternative reasons of Mob_{jt}, $Power_{jt}$, and EW_{jt} by

---

lower-secondary education when using *SK1* as the skill index. When using *SK2* or *SK3*, high-skilled (low-skilled) workers are defined as the third and fourth (first and second) quartiles of the corresponding skill distribution. This definition of high- and low-skilled workers is also used when constructing *Power* and *EW*.

[27] There are 60 industry categories based on two-digit NIC-1998 codes.

[28] As explained in the construction of *SK3(2)*, routine task intensity extracted from Autor and Dorn (2013) and Dorn (2009) is assigned to each individual based on individual occupation.

interacting them with $Skill_{ijt}$, and adding one of these interaction terms to equation (4.4). The estimation results presented in Table 7 imply that labor mobility and bargaining power do not generally explain inter-industry skill wage differentials. Somewhat unexpectedly, the skill wage premium tends to be higher in industries where routine task intensity is relatively higher for high-skilled workers. Importantly, the negative sign and statistical significance of the coefficient on $Skill_{ijt} * ChainL_{ijt}$ remain the same as in the results without controlling for these alternative explanations (third column of each sample in Table 5). In sum, this paper's baseline results remain robust.

7.3 Results for year 2009

As mentioned in Section 5.1, data for the year 1999 are superior to other years' data. However, to exclude the possibility that this study's results only capture a year-specific phenomenon, the year 2009's data is also examined in this section. As for the year-2009-specific analysis, a 54-industry classification is adopted due to the smaller sample size of the NSS data. In addition, *SK1* measures years of education more precisely because more detailed education categories are available in the 2009–2010 NSS: "higher secondary" in the 1999–2000 NSS is decomposed into "higher secondary" (12 years of schooling) and "diploma and certificate courses" (13 years). Similarly, "graduate and above" in the 1999–2000 NSS is decomposed into "graduate" (15 years) and "postgraduate and above" (17 years). Finally, it should be noted that *SK3* is not available for 2009 (see Section 5.2.1).

Table 8 reports the results of the individual-level skill-sorting regression in 2009. Similarly to the year 1999, negative sorting is observed in all-industry or manufacturing and service industry samples. In contrast, the trend of negative sorting is not clear when restricting the sample to manufacturing. Next, Table 9 shows the results of industry-level skill-sorting regressions in 2009. Again, negative sorting occurs in all-industry or manufacturing and service industry samples but not in the manufacturing sample.

Estimation results for the skill wage premium regression in 2009 are reported in Table 10. Overall, a similar trend as that seen in 1999 (Tables 4, 5) can be observed. Namely, returns to skill are higher in industries with shorter production chains. However, compared with 1999, the results are less robust when interaction terms between *Skill* and various control variables are included (columns (4)–(6)). In these cases, negative sorting is only clearly observed when using *SK2*. Higher values for *ChainQ_Import* and *ChainQ_Skill* tend to increase the skill wage premium, but the results are not so robust.

In sum, the results are less robust in 2009. However, trends consistent with this study's prediction are still observed. This might be because industrial structure does not change substantially between 1999 and 2009. An industry's production chain length, average worker skill level, dependence on imported inputs, and skill level embodied in inputs from other industries are highly correlated between the two periods.[29] The slightly less-robust results may partly reflect the less-precise measure for *ChainL* and *ChainQ*, as mentioned in Section 5.1.

8. Conclusion

In this paper, I have proposed a new mechanism to explain skill-sorting patterns and skill wage differentials across industries according to the length of industries' production chains. Using a simple model, I have shown that when the quality of intermediate inputs deteriorates rapidly along the production chains, high-skilled workers self-select into industries with shorter production chains ("negative sorting")

[29] When regressing the 2009 figure for each industry (*ChainL*, *Skill*, *ChainQ_Import*, or *ChainQ_Skill*) on the 1999 figure (and 1), the estimated coefficients range from 0.78 to 1.60. The corresponding *R*-squared ranges from 0.76 to 0.93.

because of higher returns to skill. I empirically confirm that such negative sorting because of seeking higher returns to skill is observed in India, where quality deterioration of inputs is likely to be substantial.

Although the results are less robust, intermediate input quality that is not captured by production chain length also affects skill-sorting patterns and inter-industry skill wage differentials. After controlling for the effect of production chain length, it was found that returns to skill tend to be higher in industries with higher dependence on imported inputs and higher skill levels embodied in inputs from other industries. The effects of these two quality indicators on skill-sorting patterns are less clear. As mentioned in Section 6.1, measuring input quality more directly and precisely is essential for further analysis, but this is left for future research.

The results of this study have important implications for understanding countries' development patterns. As suggested by Grossman (2004) and Asuyama (2012), when the trend for negative sorting is strong, a country is likely to have a comparative advantage in industries with shorter production chains (e.g. many service industries). Negative sorting in India and its service-led growth is the most prominent example. Thus, if the governments of developing countries want to foster manufacturing industries, most of which are characterized by long production chains and high levels of job creation, upgrading the country's input quality by reforming education, technology, and infrastructure policies to mitigate negative sorting or induce positive sorting will be critical.

References

Asuyama, Yoko. 2012. "Skill Distribution and Comparative Advantage: A Comparison of China and India." *World Development*, 40(5): 956-969.

Asuyama, Yoko, and Hideaki Goto. 2015. "Negative Skill Sorting in Sequential

Production." Unpublished working paper.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, 103(5): 1553-1597.

Autor David H., and Michael J. Handel. 2013. "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages." *Journal of Labor Economics*, 31(2): S60-96.

Azam, Mehtabul. 2012. "Changes in Wage Structure in Urban India, 1983–2004: A Quantile Regression Decomposition." *World Development*, 40(6): 1135-1150.

CSO (Central Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India). *Input-Output Transactions Table 1993-94, 1998-99, and 2007-08*.

Dahl, Gordon. B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica*, 70(6): 2367-2420.

Dorn, David. 2009. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." Dissertation University of St. Gallen, No. 3613.

Dutta, Puja Vasudeva. 2006. "Returns to Education: New Evidence for India, 1983–1999." *Education Economics*, 14(4): 431-451.

Eichengreen, Barry, and Poonam Gupta. 2011. "The Service Sector as India's Road to Economic Growth." NBER Working Paper, 16757.

Garen, John. 1984. "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable." *Econometrica*, 52(5): 1199-1218.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. 2005. "Comparative Advantage, Learning, and Sectoral Wage Determination." *Journal of Labor Economics*, 23(4): 681-724.

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

Heckman, James, and Jose Scheinkman. 1987. "The Importance of Bundling in a Gorman-Lancaster Model of Earnings." *Review of Economic Studies*, 54(2): 243-255.

Ingram, Beth F., and George R. Neumann. 2006. "The Returns to Skill." *Labour Economics*, 13(1): 35-59.

Kamath, Anant. 2011. "Does Technical Education in India Contribute to Its Core-HRST? A Case Study of IIT Madras." *Science and Public Policy*, 38(4): 293-305.

Keane, Michael P., and Kenneth I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy*, 105(3): 473-522.

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.

Lucas, Jr., Robert E. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9 (2): 508-523.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics*, 106(2): 503-530.

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*, *1999-2000 (55th round) and 2009-2010 (66th 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.

Rosen, Sherwin. 1982. “Authority, Control, and the Distribution of Earnings.” *Bell Journal of Economics*, 13 (2): 311-323.

Roy, A.D. 1951. “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers*, *New Series*, 3(2): 135-146.

Sampson, Thomas. 2013. “Assignment Reversals: Trade, Skill Allocation and Wage Inequality.” Unpublished working paper. London School of Economics.

Shapiro, Carl, and Joseph E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74(3): 433-444.

Sohoni, Milind, and Vinish Kathuria. 2014. The Elite University: Are We Too Selective?” In IIT Bombay Alumni Association ed. *Fundmatics*, No. 12, December 2014: 2-22.

Timmer, Marcel P., Dietzenbacher, Erik, Los, Bart, Stehrer, Robert, and de Vries, Gaaitzen J. 2015. “An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production.” *Review of International Economics*, 23(3): 575-605.

UNIDO (United Nations Industrial Development Organization). 2010. *Final Report: UNIDO-ACMA Partnership Programme*.

Wooldridge, Jeffrey M. 2015. “Control Function Methods in Applied Econometrics.” *Journal of Human Resources*, 50(2): 420-445.

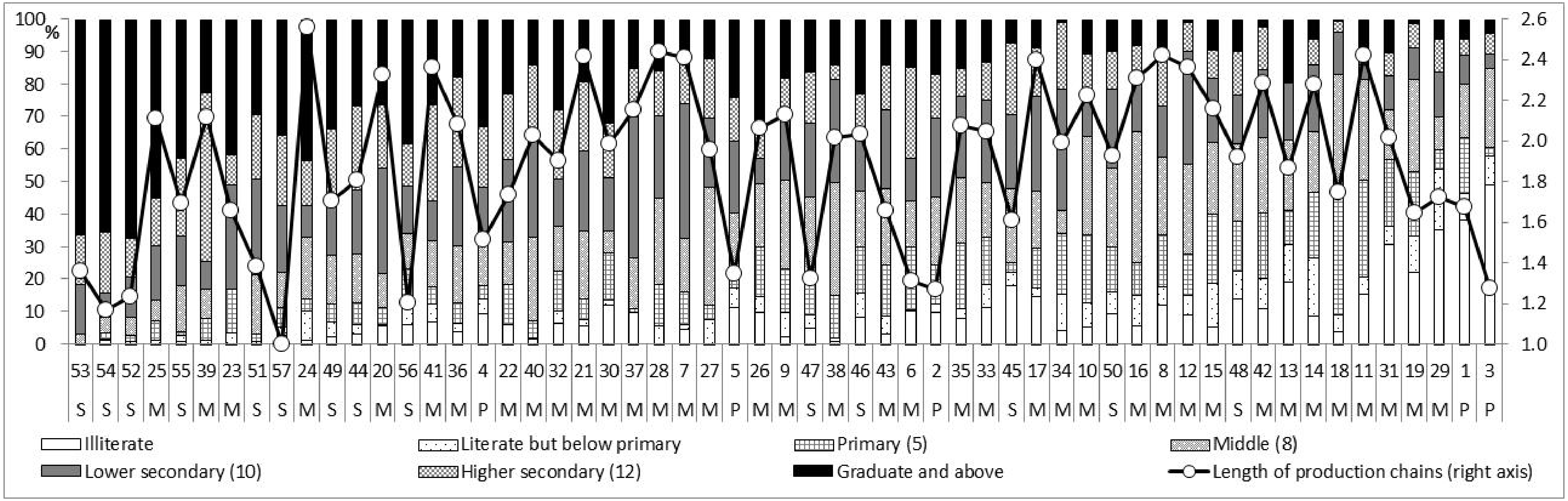
World Bank. 2004. *Trade Policies in South Asia: An Overview (Vol. 3): Some Key Sectors*. Washington, DC: World Bank.

World Bank. 2015. World Development Indicators. 28 July, 2015 version.

World Economic Forum. 2014. *The Global Competitiveness Report 2014-2015*.

Yamaguchi, Shintaro. 2012. “Tasks and Heterogeneous Human Capital.” *Journal of Labor Economics*, 30(1): 1-53.

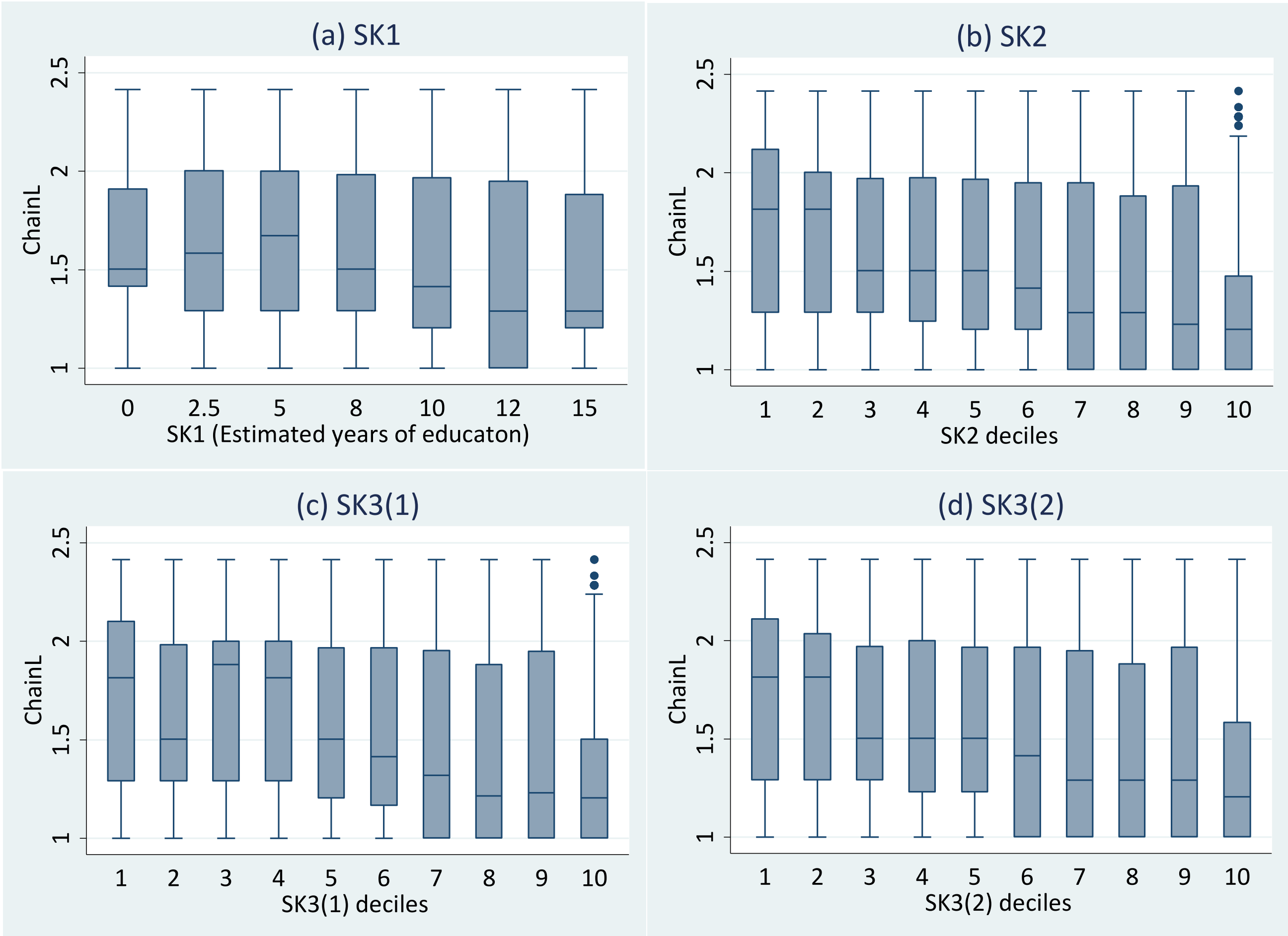
Figure 1. Educational Attainment of Indian Male Regular Wage/Salaried Workers by Industry in 2009
(Sorted from Left to Right by Industry Workforce's Average Years of Education)



Notes: The bar graph represents the percentage of workers completing the corresponding educational level. Workers include male, prime-age (15–65 years old), regular wage/salaried (RWS) employees who have worked full time and are not currently attending an educational institution (the same sample used in this paper's empirical analysis). The figures in parentheses after some educational levels indicate the corresponding regular schooling years completed. The industry's production chain length is computed as explained in Section 5.2.2. The x-axis stands for the industry's classification number (see Appendix Table B.3 for the industry description). P, M, and S indicate primary, manufacturing, and service sector, respectively. Industries are sorted by the estimated completed years of education (*SK1* in Section 5.2.1) averaged over the industry's workforce (i.e., average years of education become lower when moving from left to right).
Source: Computed by author from NSSO (2009–2010) based on weekly status.

Figure 2. Correlation between Skill Level and Industry's Production Chain Length in 1999

(Individual-level box plots based on all-industry sample)



Notes: The horizontal line in the middle of the box denotes the median *ChainL*.

Figure 3. Correlation between Skill Level and Industry's Production Chain Length in 1999

(Industry-level unweighted association based on all-industry sample)

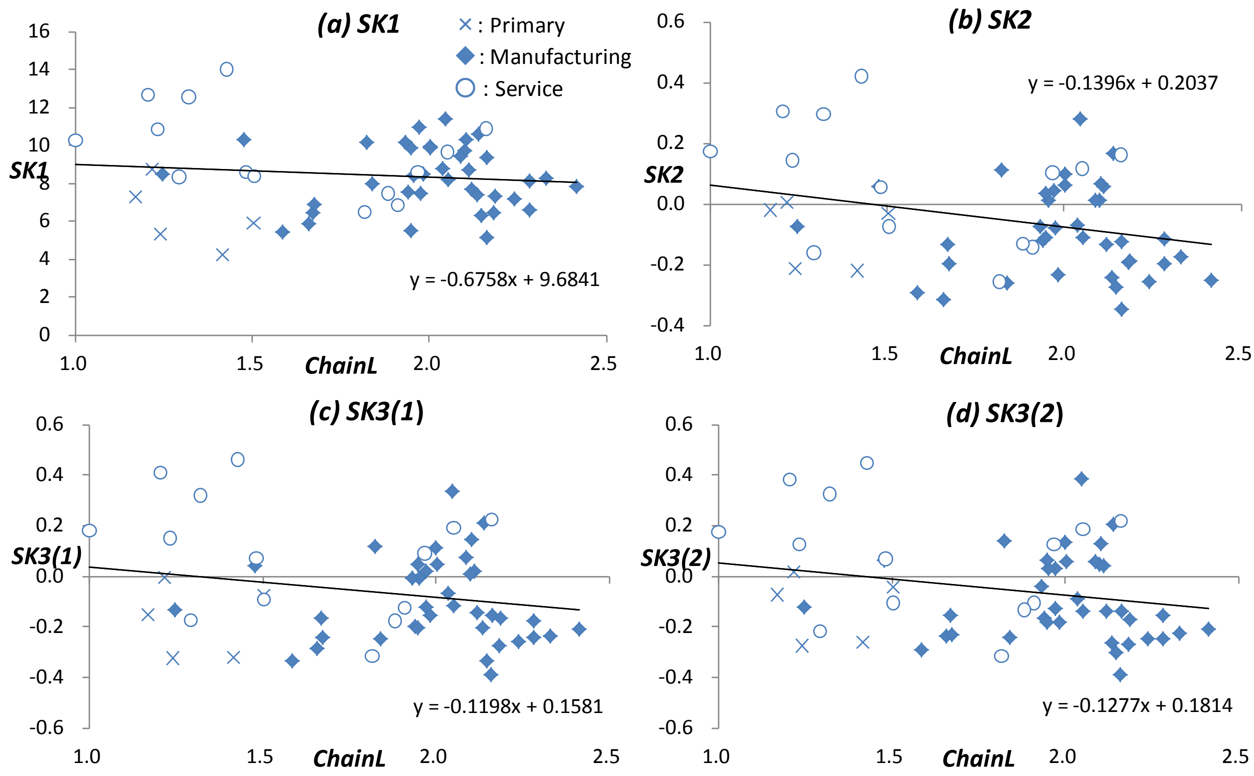


Table 1. Individual-Level Skill-Sorting Regression (Eq. (4.1)) in 1999

	All-industry sample			Manufacturing/Service sample			Manufacturing sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	-0.016 ***	-0.024 ***	-0.017 ***	-0.023 ***	-0.029 ***	-0.021 ***	-0.002 ***	-0.003 ***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ChainL_sfamily			0.402 ***			0.398 ***			0.098 ***
			(0.016)			(0.017)			(0.014)
R-squared	0.034	0.105	0.208	0.054	0.118	0.216	0.002	0.034	0.071
N	45,861	33,812	12,645	42,680	31,473	11,616	9,164	6,761	2,832
F-test for ChainL_sfamily			615.369			566.509			49.078
SK2	-0.299 ***	-0.332 ***	-0.246 ***	-0.342 ***	-0.368 ***	-0.278 ***	-0.038 ***	-0.038 ***	-0.035 *
	(0.012)	(0.012)	(0.018)	(0.012)	(0.013)	(0.019)	(0.010)	(0.013)	(0.020)
ChainL_sfamily			0.405 ***			0.404 ***			0.100 ***
			(0.017)			(0.017)			(0.014)
R-squared	0.065	0.107	0.214	0.080	0.115	0.218	0.004	0.035	0.074
N	32,612	32,612	12,043	30,505	30,505	11,144	6,605	6,605	2,732
F-test for ChainL_sfamily			600.000			565.958			50.095
SK3(1)	-0.253 ***	-0.271 ***	-0.210 ***	-0.296 ***	-0.305 ***	-0.241 ***	-0.029 ***	-0.028 ***	-0.021
	(0.010)	(0.011)	(0.015)	(0.011)	(0.012)	(0.016)	(0.009)	(0.011)	(0.016)
ChainL_sfamily			0.403 ***			0.401 ***			0.103 ***
			(0.017)			(0.017)			(0.014)
R-squared	0.062	0.106	0.215	0.079	0.114	0.221	0.004	0.034	0.075
N	31,805	31,805	11,706	29,870	29,870	10,899	6,536	6,536	2,702
F-test for ChainL_sfamily			582.993			545.971			51.321
SK3(2)	-0.233 ***	-0.242 ***	-0.175 ***	-0.265 ***	-0.265 ***	-0.193 ***	-0.037 ***	-0.038 ***	-0.032 *
	(0.010)	(0.011)	(0.015)	(0.011)	(0.011)	(0.016)	(0.009)	(0.010)	(0.016)
ChainL_sfamily			0.409 ***			0.410 ***			0.101 ***
			(0.017)			(0.017)			(0.014)
R-squared	0.053	0.095	0.206	0.064	0.100	0.208	0.006	0.036	0.077
N	31,805	31,805	11,706	29,870	29,870	10,899	6,536	6,536	2,702
F-test for ChainL_sfamily			605.524			580.147			49.607

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is *ChainL*. The explanatory variables in columns (1), (4), and (7) only include the corresponding skill index and a constant. In columns (2), (5), and (8), *Muslim*, *SG1-3*, *Hhead*, *Married*, *Rural*, and state dummies (plus *EXP* and *EXP* squared in case of column (2)) are also controlled for. In columns (3), (6), and (9), *ChainL_sfamily* is additionally controlled for. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 2. Industry-Level Skill-Sorting Regression (Eq. (4.2)) in 1999

Dep. Var.		All-industry sample			Manufacturing/Service sample			Manufacturing sample		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	*ChainL*	-2.036***	-2.043	-2.735**	-2.371***	-2.444*	-3.134***	-1.205	-0.091	-1.196
		(0.753)	(1.386)	(1.128)	(0.695)	(1.399)	(0.796)	(1.436)	(1.529)	(0.916)
	ChainQ_Import		1.722	-0.663		-3.982	-8.408**		6.740	-3.778
			(6.423)	(5.792)		(4.836)	(4.047)		(4.803)	(3.662)
	ChainQ_Skill		-0.725	-0.785		0.356	0.145		0.425	0.748***
			(0.951)	(0.763)		(0.419)	(0.370)		(0.454)	(0.234)
	R-squared	0.148	0.111	0.415	0.305	0.287	0.717	0.022	0.220	0.669
	N	57	56	56	52	51	51	38	38	38
SK2	*ChainL*	-0.215***	-0.129	-0.190**	-0.232***	-0.148	-0.214***	-0.130	-0.025	-0.123*
		(0.073)	(0.132)	(0.079)	(0.069)	(0.144)	(0.067)	(0.107)	(0.120)	(0.069)
	ChainQ_Import		-0.389	-0.569		-0.645	-1.003***		0.593	-0.406
			(0.470)	(0.359)		(0.437)	(0.309)		(0.373)	(0.276)
	ChainQ_Skill		-0.031	-0.038		0.020	-0.002		-0.001	0.030
			(0.059)	(0.036)		(0.045)	(0.028)		(0.042)	(0.020)
	R-squared	0.237	0.119	0.673	0.303	0.200	0.804	0.033	0.100	0.778
	N	57	56	56	52	51	51	38	38	38
SK3(1)	*ChainL*	-0.240***	-0.163	-0.238**	-0.264***	-0.192	-0.272***	-0.145	-0.021	-0.128
		(0.090)	(0.168)	(0.113)	(0.085)	(0.181)	(0.096)	(0.130)	(0.125)	(0.085)
	ChainQ_Import		-0.326	-0.556		-0.705	-1.159**		0.726	-0.366
			(0.609)	(0.497)		(0.552)	(0.434)		(0.462)	(0.372)
	ChainQ_Skill		-0.041	-0.049		0.033	0.007		0.023	0.057**
			(0.080)	(0.054)		(0.054)	(0.037)		(0.049)	(0.027)
	R-squared	0.199	0.101	0.594	0.278	0.202	0.757	0.031	0.172	0.728
	N	57	56	56	52	51	51	38	38	38
SK3(2)	*ChainL*	-0.219**	-0.124	-0.197*	-0.238***	-0.147	-0.227**	-0.173	-0.045	-0.165*
		(0.090)	(0.170)	(0.108)	(0.087)	(0.185)	(0.095)	(0.128)	(0.120)	(0.084)
	ChainQ_Import		-0.457	-0.681		-0.764	-1.196***		0.737	-0.372
			(0.574)	(0.457)		(0.537)	(0.416)		(0.453)	(0.349)
	ChainQ_Skill		-0.030	-0.039		0.030	0.004		0.013	0.050**
			(0.072)	(0.045)		(0.055)	(0.037)		(0.050)	(0.025)
	R-squared	0.179	0.086	0.633	0.233	0.160	0.759	0.042	0.150	0.750
	N	57	56	56	52	51	51	38	38	38

Notes: Figures are WLS estimates with weight = employment size of each industry. “Dep. Var.” denotes the dependent variable. The explanatory variables in columns (1), (4), and (7) only include *ChainL* and a constant. In columns (2), (5), and (8), *ChainQ_Import*, *ChainQ_Skill* are additionally included as regressors. In columns (3), (6), and (9), *Import, Export*, and *Smallfirm* are also additionally controlled for. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 3. Industry-Level Skill-Sorting Regression: 1999 and 2009 Panels (Eq. (4.3))

Dep. Var.		All-industry sample			Manufacturing/Service sample			Manufacturing sample		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	*ChainL*	-1.624***	-1.834***	-1.495***	-1.683***	-1.856***	-1.755***	-1.148*	-1.028	-1.395*
		(0.526)	(0.526)	(0.507)	(0.610)	(0.603)	(0.462)	(0.595)	(0.698)	(0.703)
	ChainQ_Import		-2.505	-1.908		-2.031	-1.509		-2.518	-2.158
			(1.543)	(1.670)		(1.410)	(1.187)		(2.528)	(2.376)
	ChainQ_Skill		-0.227	-0.260		-0.266	-0.388		-1.254***	-1.197**
			(0.246)	(0.254)		(0.247)	(0.273)		(0.451)	(0.452)
	R-squared	0.683	0.660	0.681	0.733	0.710	0.757	0.419	0.530	0.562
	N	114	112	112	104	102	102	76	76	76
SK2	*ChainL*	-0.120***	-0.137***	-0.139***	-0.116**	-0.131***	-0.147***	-0.169**	-0.176**	-0.190***
		(0.040)	(0.041)	(0.048)	(0.047)	(0.046)	(0.053)	(0.063)	(0.069)	(0.066)
	ChainQ_Import		-0.111	-0.132		-0.093	-0.118		-0.255	-0.211
			(0.105)	(0.103)		(0.102)	(0.088)		(0.155)	(0.159)
	ChainQ_Skill		-0.023	-0.021		-0.027	-0.030		-0.078**	-0.076**
			(0.019)	(0.019)		(0.019)	(0.020)		(0.032)	(0.032)
	R-squared	0.170	0.200	0.223	0.168	0.209	0.243	0.199	0.308	0.346
	N	114	112	112	104	102	102	76	76	76

Notes: Figures are WLS estimates with weight = NSS survey weight. “Dep. Var.” denotes the dependent variable. The explanatory variables of each column include those in the corresponding column of Table 2 plus dummies for year 2009 and industries. Standard errors clustered by industry are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 4. Skill Wage Premium Regression (Eq. (4.4)) in 1999

	All-industry sample			Manufacturing/Service sample			Manufacturing sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Skill index = SK1									
SK1	0.093***	0.076***	0.070***	0.092***	0.076***	0.069***	0.077***	0.097***	0.149
	(0.003)	(0.005)	(0.010)	(0.003)	(0.005)	(0.010)	(0.017)	(0.028)	(0.105)
SK1*ChainL	-0.010***	-0.005	-0.013**	-0.009***	-0.005	-0.013**	-0.007	-0.020	-0.056
	(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.006)	(0.008)	(0.013)	(0.050)
Exp	0.050***	0.044***	0.014	0.049***	0.042***	0.014	0.045***	0.042***	0.034***
	(0.003)	(0.003)	(0.010)	(0.003)	(0.003)	(0.010)	(0.004)	(0.004)	(0.010)
Exp^2	-0.001***	-0.001***	-0.00004	-0.001***	-0.0005***	-0.00004	-0.001***	-0.001***	-0.0004**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Muslim	-0.057***	-0.040**	-0.072**	-0.058***	-0.045***	-0.076**	-0.028	-0.005	-0.165*
	(0.016)	(0.016)	(0.036)	(0.016)	(0.016)	(0.037)	(0.032)	(0.031)	(0.088)
SG1	0.021	-0.019	-0.041	0.027	-0.008	-0.046	-0.041	-0.047	-0.151
	(0.028)	(0.027)	(0.052)	(0.031)	(0.031)	(0.052)	(0.067)	(0.066)	(0.121)
SG2	-0.027	-0.061***	-0.039	-0.043**	-0.074***	-0.040	-0.051	-0.041	-0.081
	(0.018)	(0.017)	(0.031)	(0.019)	(0.018)	(0.031)	(0.034)	(0.032)	(0.076)
SG3	-0.058***	-0.068***	-0.028	-0.057***	-0.065***	-0.028	-0.023	-0.019	-0.056
	(0.013)	(0.013)	(0.026)	(0.014)	(0.013)	(0.026)	(0.025)	(0.023)	(0.054)
Hhead	0.151***	0.130***	0.093***	0.157***	0.137***	0.097***	0.128***	0.124***	0.076
	(0.016)	(0.015)	(0.035)	(0.016)	(0.016)	(0.035)	(0.026)	(0.026)	(0.069)
Married	0.087***	0.075***	0.049	0.097***	0.085***	0.046	0.072**	0.071**	-0.100
	(0.019)	(0.018)	(0.051)	(0.019)	(0.019)	(0.051)	(0.030)	(0.029)	(0.084)
Occ1	0.258***	0.389***	0.419***	0.243***	0.378***	0.420***	0.823***	0.789***	0.671***
	(0.028)	(0.032)	(0.046)	(0.028)	(0.032)	(0.046)	(0.148)	(0.136)	(0.153)
Occ2	0.291***	0.275***	0.327***	0.287***	0.269***	0.331***	0.480***	0.459***	0.513***
	(0.036)	(0.034)	(0.082)	(0.036)	(0.034)	(0.082)	(0.071)	(0.064)	(0.134)
Occ3	0.480***	0.429***	0.418***	0.469***	0.423***	0.419***	0.986***	0.993***	0.076
	(0.048)	(0.049)	(0.091)	(0.049)	(0.049)	(0.092)	(0.163)	(0.188)	(0.201)
Occ4	0.455***	0.481***	0.701***	0.452***	0.474***	0.695***	0.848***	0.844***	1.112***
	(0.050)	(0.049)	(0.065)	(0.052)	(0.051)	(0.065)	(0.076)	(0.074)	(0.158)
Occ5	0.059***	0.058***	0.118***	0.046**	0.048**	0.121***	0.189***	0.231***	0.157
	(0.021)	(0.021)	(0.038)	(0.021)	(0.021)	(0.038)	(0.060)	(0.060)	(0.148)
Occ6	-0.268***	-0.056*	0.057	-0.267***	-0.051*	0.059	0.147	0.222**	0.380**
	(0.025)	(0.030)	(0.075)	(0.026)	(0.030)	(0.075)	(0.128)	(0.101)	(0.157)
Occ8	-0.301***	-0.151***	-0.043	0.014	0.051	-0.042	0.137	0.129	-0.436**
	(0.029)	(0.041)	(0.084)	(0.064)	(0.063)	(0.084)	(0.151)	(0.140)	(0.173)
Occ9	0.108***	0.112***	0.093	0.073**	0.091**	0.085	0.242***	0.213***	0.205*
	(0.035)	(0.043)	(0.065)	(0.036)	(0.045)	(0.067)	(0.055)	(0.064)	(0.119)
Occ10	0.037**	0.048***	0.089***	0.017	0.041**	0.091***	0.122***	0.120**	0.168
	(0.017)	(0.018)	(0.033)	(0.017)	(0.019)	(0.034)	(0.047)	(0.049)	(0.103)
Occ11	0.071*	0.227***	0.088	0.071*	0.210***	0.095	0.132	0.118	0.908***
	(0.040)	(0.042)	(0.168)	(0.039)	(0.041)	(0.168)	(0.112)	(0.110)	(0.180)
Temporary			-0.258***			-0.268***			-0.162**
			(0.073)			(0.075)			(0.068)
Union			-0.020			-0.022			0.050
			(0.048)			(0.049)			(0.067)
Publicfirm			0.103***			0.106***			0.353***
			(0.034)			(0.035)			(0.077)

Smallfirm			-0.225***			-0.226***			-0.095
			(0.037)			(0.037)			(0.073)
SS			0.148*			0.143*			0.086
			(0.077)			(0.079)			(0.066)
Rural	-0.161***	-0.158***	-0.008	-0.145***	-0.158***	-0.010	-0.142***	-0.116***	0.006
	(0.014)	(0.014)	(0.034)	(0.015)	(0.014)	(0.034)	(0.024)	(0.025)	(0.053)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies		Yes	Yes		Yes	Yes		Yes	Yes
R-squared	0.495	0.540	0.541	0.479	0.522	0.542	0.464	0.498	0.589
N	32,612	32,612	4,897	30,505	30,505	4,814	6,605	6,605	861
Skill index = SK2									
SK2	1.139***	0.919***	1.030***	1.137***	0.905***	1.004***	1.264***	1.080***	4.276***
	(0.068)	(0.068)	(0.186)	(0.069)	(0.069)	(0.186)	(0.338)	(0.336)	(1.214)
SK2*ChainL	0.005	0.052	-0.021	0.000	0.053	-0.011	-0.132	-0.090	-1.676***
	(0.043)	(0.041)	(0.117)	(0.043)	(0.041)	(0.117)	(0.162)	(0.160)	(0.591)
R-squared	0.501	0.547	0.544	0.487	0.530	0.545	0.476	0.507	0.590
N	32,612	32,612	4,897	30,505	30,505	4,814	6,605	6,605	861
Skill index = SK3(1)									
SK3(1)	1.357***	1.019***	1.062***	1.270***	0.996***	1.051***	1.450***	1.321***	3.582***
	(0.059)	(0.056)	(0.129)	(0.059)	(0.056)	(0.130)	(0.309)	(0.319)	(1.024)
SK3(1)*ChainL	-0.133***	-0.013	-0.043	-0.097**	-0.009	-0.039	-0.192	-0.169	-1.256**
	(0.041)	(0.038)	(0.081)	(0.041)	(0.038)	(0.081)	(0.153)	(0.157)	(0.491)
R-squared	0.484	0.544	0.535	0.475	0.527	0.535	0.452	0.484	0.543
N	31,805	31,805	4,883	29,870	29,870	4,800	6,536	6,536	860
Skill index = SK3(2)									
SK3(2)	1.359***	1.000***	1.050***	1.256***	0.975***	1.034***	1.535***	1.359***	3.832***
	(0.056)	(0.055)	(0.128)	(0.056)	(0.054)	(0.128)	(0.305)	(0.305)	(0.990)
SK3(2)*ChainL	-0.123***	0.000	-0.034	-0.079**	0.005	-0.029	-0.224	-0.176	-1.347***
	(0.039)	(0.037)	(0.082)	(0.039)	(0.037)	(0.082)	(0.151)	(0.151)	(0.478)
R-squared	0.488	0.546	0.528	0.483	0.531	0.529	0.462	0.492	0.548
N	31,805	31,805	4,883	29,870	29,870	4,800	6,536	6,536	860

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is the logarithm of weekly wages. When using *SK2*, explanatory variables of each column include those in case of *SK1* minus *Exp* and *Exp*^2 (*Exp* squared). When using *SK3*, they include those in case of *SK2* minus occupation dummies. The reference category for social group and occupation is *SG4* (other) and *Occ7* (service worker), respectively. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 5. Skill Wage Premium Regression (Eq. (4.4)) in 1999 with Additional Interaction Terms with Skill Index

	All-industry sample			Manufacturing/Service sample			Manufacturing sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	0.051***	-0.007	0.010	0.067***	-0.010	0.007	0.025	-0.025	0.032
	(0.018)	(0.036)	(0.036)	(0.018)	(0.036)	(0.036)	(0.032)	(0.095)	(0.095)
SK1*ChainL	-0.014***	-0.021**	-0.025**	-0.014***	-0.019*	-0.022**	-0.006	0.022	0.006
	(0.005)	(0.010)	(0.010)	(0.005)	(0.010)	(0.010)	(0.016)	(0.045)	(0.043)
*SK1*ChainQ_Import*	0.096***	0.215***	0.167**	0.092***	0.211***	0.162**	0.069	0.358**	0.376**
	(0.028)	(0.064)	(0.067)	(0.028)	(0.065)	(0.068)	(0.069)	(0.175)	(0.177)
*SK1*ChainQ_Skill*	0.005**	0.011**	0.013***	0.003	0.010**	0.012***	0.005	-0.004	-0.006
	(0.002)	(0.005)	(0.005)	(0.002)	(0.005)	(0.005)	(0.005)	(0.011)	(0.012)
*SK1*Temporary*			-0.027**			-0.028**			-0.013
			(0.011)			(0.012)			(0.015)
*SK1*Union*			-0.004			-0.005			-0.003
			(0.008)			(0.008)			(0.013)
*SK1*Publicfirm*			-0.017**			-0.018***			-0.051***
			(0.007)			(0.007)			(0.018)
*SK1*Smallfirm*			-0.022***			-0.021***			-0.026*
			(0.007)			(0.007)			(0.014)
R-squared	0.530	0.592	0.598	0.513	0.593	0.598	0.500	0.598	0.609
N	25,334	3,467	3,467	23,227	3,384	3,384	6,605	861	861
SK2	0.470*	0.176	0.298	0.525**	0.205	0.334	0.128	2.376	1.742
	(0.252)	(0.879)	(0.858)	(0.259)	(0.881)	(0.860)	(0.409)	(1.519)	(1.491)
SK2*ChainL	-0.091	-0.475**	-0.498**	-0.090	-0.440**	-0.459**	-0.135	-1.797**	-1.603**
	(0.061)	(0.198)	(0.207)	(0.061)	(0.199)	(0.208)	(0.194)	(0.794)	(0.722)
*SK2*ChainQ_Import*	0.279	3.508***	2.729**	0.171	3.491***	2.718**	-0.172	2.033	3.138
	(0.343)	(1.291)	(1.354)	(0.351)	(1.300)	(1.364)	(0.740)	(3.235)	(3.360)
*SK2*ChainQ_Skill*	0.096***	0.199*	0.252**	0.088***	0.185	0.236**	0.153**	0.274	0.319
	(0.030)	(0.118)	(0.116)	(0.031)	(0.119)	(0.116)	(0.064)	(0.235)	(0.232)
*SK2*Temporary*			-0.655**			-0.670**			-0.431
			(0.268)			(0.272)			(0.379)
*SK2*Union*			-0.036			-0.041			0.191
			(0.152)			(0.154)			(0.299)
*SK2*Publicfirm*			-0.385***			-0.395***			-1.154***
			(0.130)			(0.131)			(0.295)
*SK2*Smallfirm*			-0.257**			-0.237*			-0.862***
			(0.128)			(0.129)			(0.285)
R-squared	0.537	0.591	0.596	0.521	0.591	0.596	0.509	0.596	0.612
N	25,334	3,467	3,467	23,227	3,384	3,384	6,605	861	861
SK3(1)	0.662***	0.449	0.569	0.712***	0.448	0.580	0.176	1.890*	2.298**
	(0.229)	(0.510)	(0.517)	(0.237)	(0.511)	(0.519)	(0.413)	(1.088)	(1.075)
SK3(1)*ChainL	-0.149***	-0.289**	-0.272**	-0.142***	-0.273**	-0.255*	-0.152	-1.540**	-1.505**
	(0.054)	(0.131)	(0.134)	(0.054)	(0.132)	(0.135)	(0.185)	(0.643)	(0.587)
*SK3(1)*ChainQ_Import*	0.390	2.022**	1.804*	0.352	1.987**	1.785*	0.195	-0.605	0.577
	(0.290)	(0.876)	(0.930)	(0.292)	(0.882)	(0.934)	(0.677)	(2.217)	(2.380)
*SK3(1)*ChainQ_Skill*	0.082***	0.130*	0.143**	0.072**	0.126*	0.137*	0.155**	0.341**	0.314**
	(0.028)	(0.070)	(0.071)	(0.029)	(0.070)	(0.072)	(0.068)	(0.156)	(0.155)
*SK3(1)*Temporary*			-0.504***			-0.507**			-0.294
			(0.194)			(0.197)			(0.268)
*SK3(1)*Union*			-0.054			-0.060			-0.300
			(0.111)			(0.112)			(0.211)

*SK3(1)*Publicfirm*			-0.182**			-0.184**			-0.554***
			(0.087)			(0.087)			(0.199)
*SK3(1)*Smallfirm*			-0.070			-0.061			-0.551***
			(0.084)			(0.084)			(0.178)
R-squared	0.534	0.578	0.583	0.517	0.579	0.583	0.488	0.552	0.565
N	24,624	3,459	3,459	22,689	3,376	3,376	6,536	860	860
SK3(2)	0.599***	0.428	0.569	0.716***	0.452	0.602	0.191	1.694	2.019*
	(0.224)	(0.545)	(0.549)	(0.230)	(0.546)	(0.551)	(0.414)	(1.096)	(1.082)
SK3(2)*ChainL	-0.130**	-0.294**	-0.288**	-0.120**	-0.273**	-0.265*	-0.171	-1.379**	-1.331**
	(0.053)	(0.137)	(0.142)	(0.053)	(0.138)	(0.142)	(0.181)	(0.641)	(0.572)
*SK3(2)*ChainQ_Import*	0.567**	2.897***	2.546***	0.593**	2.893***	2.558***	0.078	1.224	2.455
	(0.273)	(0.881)	(0.946)	(0.274)	(0.887)	(0.952)	(0.670)	(2.118)	(2.244)
*SK3(2)*ChainQ_Skill*	0.084***	0.124	0.146*	0.063**	0.114	0.135*	0.164**	0.293*	0.268*
	(0.027)	(0.075)	(0.075)	(0.028)	(0.076)	(0.076)	(0.068)	(0.161)	(0.159)
*SK3(2)*Temporary*			-0.546***			-0.551***			-0.363
			(0.201)			(0.203)			(0.265)
*SK3(2)*Union*			-0.073			-0.080			-0.217
			(0.117)			(0.118)			(0.224)
*SK3(2)*Publicfirm*			-0.218**			-0.221**			-0.628***
			(0.092)			(0.093)			(0.212)
*SK3(2)*Smallfirm*			-0.142			-0.129			-0.632***
			(0.093)			(0.092)			(0.201)
R-squared	0.536	0.572	0.578	0.522	0.573	0.579	0.496	0.559	0.573
N	24,624	3,459	3,459	22,689	3,376	3,376	6,536	860	860

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is the logarithm of weekly wages. Other control variables in columns (1), (4), and (7) are the same as those in columns (2), (5), or (8) in Table 4. Other control variables of the remaining columns are the same as those in columns (3), (6), or (9) in Table 4. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 6. Selection-Corrected Skill Wage Premium Regression in 1999

	(1)	(2)	(3)	(4)
Skill index	*SK1*	*SK2*	*SK3(1)*	*SK3(2)*
Skill	0.185***	1.693***	1.378***	1.271***
	(0.014)	(0.153)	(0.130)	(0.135)
Skill*ChainL	-0.0613***	-0.237**	-0.175**	-0.083
	(0.009)	(0.102)	(0.087)	(0.091)
$\hat{\varepsilon}_1$ **Skill*ChainL*	0.030***	0.010	0.003	-0.043
	(0.005)	(0.073)	(0.061)	(0.062)
$\hat{\varepsilon}_1$	-0.332***	0.074	0.068	0.029
	(0.093)	(0.072)	(0.073)	(0.072)
R-squared	0.459	0.467	0.464	0.470
N	12,052	12,043	11,706	11,706
F-test for selection vars.	18.396	0.587	0.503	0.476
p-value	0.000	0.556	0.605	0.621

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is the logarithm of weekly wages. Sample is based on all industries. As mentioned in Section 7.1, $\hat{\varepsilon}_1$ denotes the residual from the skill-sorting regression of equation (4.1) (=columns (3), (6), and (9) in Table 1). "Selection vars." denotes $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_1 * Skill * ChainL$. *Skill*, *ChainL*, *Muslim*, *SG1-3*, *Hhead*, *Married*, *Rural*, State dummies are included as other regressors in all regressions. When using *SK1*, *Exp* and its square are additionally controlled for. Bootstrap standard errors based on 1000 replications are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 7. Alternative Reasons for Inter-industry Skill Wage Differentials (1999)

	All-industry sample			Manufacturing/Service sample			Manufacturing sample		
AltReason =	*Mob*	*Power*	*EW*	*Mob*	*Power*	*EW*	*Mob*	*Power*	*EW*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	0.010	0.029	-0.017	0.008	0.024	-0.021	0.022	0.039	-0.095
	(0.036)	(0.039)	(0.039)	(0.036)	(0.039)	(0.039)	(0.096)	(0.094)	(0.133)
SK1*ChainL	-0.024**	-0.030***	-0.024**	-0.021**	-0.027**	-0.021**	0.008	-0.014	0.007
	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.043)	(0.044)	(0.043)
*SK1*ChainQ_Import*	0.167**	0.160**	0.185***	0.162**	0.154**	0.180***	0.382**	0.345*	0.281
	(0.067)	(0.068)	(0.067)	(0.068)	(0.069)	(0.068)	(0.177)	(0.179)	(0.204)
*SK1*ChainQ_Skill*	0.013***	0.012***	0.013***	0.012***	0.012***	0.013***	-0.005	0.001	0.005
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.012)	(0.013)	(0.015)
*SK1*AltReason*	0.000	-0.001	0.020**	0.000	-0.001	0.021**	0.001	-0.007	0.061
	(0.001)	(0.001)	(0.010)	(0.001)	(0.001)	(0.010)	(0.002)	(0.005)	(0.042)
R-squared	0.598	0.598	0.599	0.599	0.599	0.600	0.609	0.611	0.611
N	3,467	3,467	3,467	3,384	3,384	3,384	861	861	861
SK2	0.322	0.298	-0.155	0.348	0.327	-0.134	1.717	1.905	-2.090
	(0.858)	(0.858)	(0.892)	(0.860)	(0.860)	(0.895)	(1.521)	(1.503)	(2.611)
SK2*ChainL	-0.507**	-0.497**	-0.547***	-0.466**	-0.454**	-0.509**	-1.608**	-2.020***	-1.433**
	(0.208)	(0.210)	(0.206)	(0.210)	(0.211)	(0.207)	(0.726)	(0.766)	(0.712)
*SK2*ChainQ_Import*	2.822**	2.730**	3.404**	2.778**	2.730**	3.406**	3.087	2.058	0.830
	(1.367)	(1.355)	(1.383)	(1.375)	(1.365)	(1.390)	(3.390)	(3.355)	(3.712)
*SK2*ChainQ_Skill*	0.255**	0.252**	0.259**	0.239**	0.235**	0.244**	0.328	0.475*	0.607**
	(0.116)	(0.116)	(0.116)	(0.116)	(0.116)	(0.116)	(0.246)	(0.242)	(0.301)
*SK2*AltReason*	0.018	0.000	0.390**	0.012	0.002	0.400**	0.007	-0.148	1.746*
	(0.029)	(0.006)	(0.183)	(0.029)	(0.006)	(0.184)	(0.063)	(0.094)	(0.922)
R-squared	0.596	0.596	0.597	0.596	0.596	0.597	0.612	0.615	0.614
N	3,467	3,467	3,467	3,384	3,384	3,384	861	861	861
SK3(1)	0.577	0.670	0.585	0.582	0.663	0.593	2.270**	2.186**	0.671
	(0.515)	(0.529)	(0.516)	(0.518)	(0.530)	(0.518)	(1.084)	(1.063)	(1.601)
SK3(1)*ChainL	-0.292**	-0.315**	-0.354**	-0.274**	-0.293**	-0.336**	-1.506**	-1.625***	-1.487**
	(0.135)	(0.146)	(0.143)	(0.137)	(0.149)	(0.145)	(0.590)	(0.589)	(0.583)
*SK3(1)*ChainQ_Import*	1.921**	1.740*	2.361**	1.886**	1.723*	2.330**	0.516	0.487	-0.419
	(0.929)	(0.931)	(0.992)	(0.934)	(0.936)	(0.997)	(2.453)	(2.345)	(2.429)
*SK3(1)*ChainQ_SKill*	0.153**	0.143**	0.104	0.148**	0.138*	0.100	0.322*	0.390**	0.438**
	(0.072)	(0.071)	(0.074)	(0.073)	(0.071)	(0.075)	(0.172)	(0.156)	(0.186)
*SK3(1)*AltReason*	0.028	-0.008	0.310**	0.026	-0.007	0.304*	0.004	-0.097	0.866
	(0.021)	(0.008)	(0.155)	(0.021)	(0.008)	(0.156)	(0.038)	(0.071)	(0.643)
R-squared	0.583	0.583	0.584	0.584	0.584	0.584	0.565	0.567	0.567
N	3,459	3,459	3,459	3,376	3,376	3,376	860	860	860
SK3(2)	0.575	0.580	0.514	0.602	0.607	0.544	2.104*	1.890*	-0.159
	(0.548)	(0.548)	(0.547)	(0.550)	(0.550)	(0.549)	(1.086)	(1.074)	(1.732)
SK3(2)*ChainL	-0.292**	-0.298**	-0.382***	-0.266*	-0.272*	-0.359**	-1.324**	-1.457**	-1.344**
	(0.144)	(0.145)	(0.144)	(0.145)	(0.146)	(0.145)	(0.571)	(0.575)	(0.568)
*SK3(2)*ChainQ_Import*	2.567***	2.531**	3.280***	2.561***	2.547***	3.269***	2.655	2.457	0.781
	(0.952)	(0.947)	(0.994)	(0.959)	(0.952)	(0.999)	(2.304)	(2.222)	(2.447)
*SK3(2)*ChainQ_SKill*	0.148*	0.148*	0.085	0.135*	0.137*	0.077	0.243	0.354**	0.457**
	(0.076)	(0.076)	(0.080)	(0.077)	(0.076)	(0.080)	(0.171)	(0.158)	(0.204)
*SK3(2)*AltReason*	0.006	-0.002	0.521***	0.001	-0.001	0.505***	-0.015	-0.119	1.129
	(0.022)	(0.004)	(0.185)	(0.022)	(0.004)	(0.185)	(0.039)	(0.075)	(0.715)
R-squared	0.578	0.578	0.580	0.579	0.579	0.581	0.573	0.576	0.576
N	3,459	3,459	3,459	3,376	3,376	3,376	860	860	860

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is the logarithm of weekly wages. Other control variables are the same as those in columns (3), (6), or (9) in Table 5. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 8. Individual-level Skill-Sorting Regression (Eq. (4.1)) in 2009

	All-industry sample			Manufacturing/Service sample			Manufacturing sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	-0.025***	-0.033***	-0.024***	-0.028***	-0.035***	-0.024***	-0.001	-0.001	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
ChainL_sfamily			0.376***			0.378***			0.124***
			(0.022)			(0.022)			(0.022)
R-squared	0.074	0.157	0.223	0.081	0.163	0.227	0.001	0.061	0.116
N	29,951	29,927	10,570	28,731	28,708	10,157	4,960	4,956	2,067
F-test for ChainL_sfamily			301.645			294.716			31.054
SK2	-0.409***	-0.443***	-0.307***	-0.427***	-0.457***	-0.314***	-0.049***	-0.023	0.018
	(0.012)	(0.014)	(0.024)	(0.013)	(0.015)	(0.025)	(0.015)	(0.019)	(0.034)
ChainL_sfamily			0.379***			0.382***			0.125***
			(0.022)			(0.022)			(0.022)
R-squared	0.106	0.153	0.220	0.110	0.157	0.222	0.005	0.057	0.113
N	29,322	29,322	10,381	28,148	28,148	9,987	4,917	4,917	2,042
F-test for ChainL_sfamily			300.888			296.118			30.929

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is *ChainL*. The explanatory variables of each column are the same as those in the corresponding column of Table 1. *SK1* in this table is computed based on the detailed educational classification, which is only available for 2009 (see Section 7.3). Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 9. Industry-level Skill-Sorting Regression (Eq. (4.2)) in 2009

Dep. Var.		All-industry sample			Manufacturing/Service sample			Manufacturing sample		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SK1	*ChainL*	-2.957***	-3.052*	-3.582***	-2.947***	-2.674*	-3.116***	-0.639	0.343	0.196
		(0.801)	(1.530)	(1.083)	(0.806)	(1.368)	(0.760)	(1.096)	(1.420)	(0.916)
	ChainQ_Import		0.407	-0.571		-1.904	-3.661**		4.927*	1.451
			(3.643)	(2.514)		(3.605)	(1.405)		(2.773)	(1.869)
	ChainQ_Skill		0.335	-0.351		0.563	-0.052		0.516	0.124
			(0.690)	(0.542)		(0.652)	(0.426)		(0.308)	(0.190)
	R-squared	0.278	0.256	0.649	0.329	0.335	0.770	0.008	0.242	0.558
	N	54	53	53	50	49	49	31	31	31
SK2	*ChainL*	-0.259***	-0.235*	-0.265***	-0.258***	-0.219*	-0.246***	-0.094	-0.045	-0.055
		(0.065)	(0.122)	(0.074)	(0.064)	(0.115)	(0.060)	(0.078)	(0.100)	(0.056)
	ChainQ_Import		-0.030	-0.118		-0.123	-0.264*		0.241	-0.047
			(0.242)	(0.168)		(0.265)	(0.150)		(0.199)	(0.085)
	ChainQ_Skill		0.002	-0.041		0.010	-0.029		0.021	-0.013
			(0.050)	(0.035)		(0.049)	(0.031)		(0.026)	(0.015)
	R-squared	0.369	0.269	0.762	0.390	0.295	0.809	0.038	0.144	0.651
	N	54	53	53	50	49	49	31	31	31

Notes: Figures are WLS estimates with weight = employment size of each industry. "Dep. Var." denotes the dependent variable. The explanatory variables of each column are the same as those in the corresponding column of Table 2. *SK1* in this table is computed based on the detailed educational classification as in Table 8. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table 10. Skill Wage Premium Regression (Eq. (4.4)) in 2009

			(1)	(2)	(3)	(4)	(5)	(6)
SK1	All-industry sample	*SK1*	0.106***	0.059***	0.073***	-0.018	0.067	0.073
			(0.004)	(0.007)	(0.011)	(0.038)	(0.057)	(0.058)
		SK1*ChainL	-0.016***	0.004	-0.013*	0.003	-0.014	-0.016
			(0.002)	(0.004)	(0.007)	(0.007)	(0.011)	(0.012)
		*SK1*ChainQ_Import*				0.054***	0.026	0.028
						(0.019)	(0.037)	(0.038)
		*SK1*ChainQ_Skill*				0.008**	0.000	0.000
						(0.004)	(0.006)	(0.006)
		R-squared	0.490	0.549	0.616	0.543	0.636	0.638
		N	29,322	29,322	7,169	23,019	5,273	5,273
	Manufacturing/Service sample	*SK1*	0.107***	0.063***	0.072***	-0.038	0.065	0.068
			(0.004)	(0.007)	(0.011)	(0.038)	(0.057)	(0.058)
		SK1*ChainL	-0.015***	0.002	-0.013*	0.003	-0.014	-0.017
			(0.002)	(0.004)	(0.007)	(0.007)	(0.011)	(0.012)
		*SK1*ChainQ_Import*				0.041**	0.033	0.035
						(0.019)	(0.038)	(0.039)
		*SK1*ChainQ_Skill*				0.011***	0.001	0.001
						(0.004)	(0.006)	(0.006)
		R-squared	0.483	0.543	0.623	0.539	0.644	0.647
		N	28,148	28,148	6,981	21,845	5,085	5,085
	Manufacturing sample	*SK1*	0.102***	0.076**	0.105**	-0.066	-0.072	-0.065
			(0.014)	(0.031)	(0.049)	(0.053)	(0.082)	(0.081)
		SK1*ChainL	-0.016**	-0.008	-0.024	-0.002	-0.019	-0.022
			(0.007)	(0.015)	(0.023)	(0.017)	(0.028)	(0.029)
		*SK1*ChainQ_Import*				0.032	0.055	0.042
						(0.035)	(0.058)	(0.058)
		*SK1*ChainQ_Skill*				0.016***	0.020**	0.020**
						(0.005)	(0.009)	(0.009)
		R-squared	0.507	0.551	0.680	0.555	0.684	0.688
		N	4,917	4,917	887	4,917	887	887
SK2	All-industry sample	*SK2*	1.134***	0.901***	1.484***	0.531	1.772*	1.958**
			(0.089)	(0.097)	(0.185)	(0.495)	(0.964)	(0.967)
		SK2*ChainL	0.068	0.064	-0.334***	-0.117	-0.625***	-0.663***
			(0.053)	(0.057)	(0.118)	(0.092)	(0.188)	(0.190)
		*SK2*ChainQ_Import*				1.004***	1.689**	1.706**
						(0.254)	(0.686)	(0.679)
		*SK2*ChainQ_Skill*				0.064	-0.002	-0.030
						(0.049)	(0.098)	(0.098)
		R-squared	0.495	0.557	0.620	0.554	0.641	0.644
		N	29,322	29,322	7,169	23,019	5,273	5,273

SK2	Manufacturing/Service sample	*SK2*	1.173***	0.948***	1.464***	0.360	1.871*	2.031**
			(0.089)	(0.096)	(0.186)	(0.483)	(0.968)	(0.969)
		SK2*ChainL	0.051	0.048	-0.331***	-0.123	-0.634***	-0.671***
			(0.053)	(0.057)	(0.118)	(0.091)	(0.188)	(0.190)
		*SK2*ChainQ_Import*				0.951***	1.783***	1.805***
						(0.252)	(0.688)	(0.684)
		*SK2*ChainQ_Skill*				0.090*	-0.017	-0.041
						(0.048)	(0.098)	(0.098)
		R-squared	0.488	0.551	0.626	0.550	0.649	0.653
		N	28,148	28,148	6,981	21,845	5,085	5,085
	Manufacturing sample	*SK2*	1.147***	1.891***	3.267***	-0.014	-0.539	-0.479
			(0.391)	(0.398)	(0.945)	(0.624)	(1.303)	(1.347)
		SK2*ChainL	-0.023	-0.430**	-1.073**	-0.490**	-0.674	-0.561
			(0.185)	(0.189)	(0.438)	(0.226)	(0.484)	(0.504)
		*SK2*ChainQ_Import*				0.007	2.592**	2.495**
						(0.453)	(1.005)	(0.974)
		*SK2*ChainQ_Skill*				0.255***	0.302*	0.240
						(0.068)	(0.161)	(0.162)
		R-squared	0.522	0.564	0.686	0.568	0.695	0.700
		N	4,917	4,917	887	4,917	887	887
	Control variables are the same as those in:		column (1)	column (2) in Table 4	column (3)	column (1)	column (2) in Table 5	column (3)

Notes: Figures are WLS estimates with weight = NSS survey weight. The dependent variable is the logarithm of weekly wages. *SK1* in this table is computed based on the detailed educational classification as in Table 8. Robust standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Appendix A. Description of Variables and Summary Statistics

(a) Industry-level statistics

Variable	Description	1999		2009	
		Mean	Std. Dev.	Mean	Std. Dev.
SK1	Skill index 1	8.442	2.017	9.126	2.071
SK2	Skill index 2	-0.053	0.173	-0.060	0.157
SK3(1)	Skill index 3(1)	-0.062	0.201	-0.026	0.098
SK3(2)	Skill index 3(2)	-0.053	0.197	-0.012	0.097
ChainL	Length of domestic production chains	1.838	0.360	1.891	0.400
ChainQ_Import	Dependence on imported inputs	0.109	0.090	0.181	0.150
ChainQ_Skill	Skill level (years of education) embodied in inputs from other industries	6.938	0.762	7.860	0.681
Import	% of final goods imports in industry output	11.965	16.753	11.363	16.042
Export	% of final goods exports in industry output	11.492	17.330	9.205	10.610
Smallfirm	Employment % of small firms with fewer than 10 employees	56.091	29.154	55.943	27.040

Notes: Number of observations (industries) is 56 for *ChainQ_Skill* and 57 for other variables.

(b) Individual-level statistics

Variable	Description	1999		2009	
		Mean	Std. Dev.	Mean	Std. Dev.
***Sample 1**: Sample used in the regressions with specifications from column(1) in Table 4*					
SK1	Skill index 1 (in 2009, figures in the second row use a more detailed educational classification. See Section 7.3)	8.934	4.712	9.919	4.461
		NA	NA	10.121	4.689
SK2	Skill index 2	0.000	0.364	0.000	0.344
SK3(1)	Skill index 3(1)	0.000	0.424	0.000	0.320
SK3(2)	Skill index 3(2)	0.000	0.422	0.000	0.326
Wage	Weekly wage (rupees)	1,000	952	2,313	2,652
ChainL	Length of domestic production chains of affiliated industry	1.557	0.427	1.580	0.432
Age	Age	36.477	11.034	36.095	11.188
Exp	Estimated years of work experience	22.543	11.659	21.176	11.903
Muslim	Dummy : 1 if religion is Islam, 0 otherwise	0.098	0.298	0.103	0.304
SG1	Dummy : 1 if social group is scheduled tribe, 0 otherwise	0.051	0.220	0.046	0.209
SG2	Dummy : 1 if social group is scheduled caste, 0 otherwise	0.147	0.354	0.165	0.371

SG3	Dummy : 1 if social group is other backward class, 0 otherwise	0.295	0.456	0.358	0.480
SG4	Dummy : 1 if social group is others, 0 otherwise	0.507	0.500	0.431	0.495
Hhead	Dummy : 1 if head of the household, 0 otherwise	0.713	0.453	0.668	0.471
Married	Dummy : 1 if currently married, 0 otherwise	0.799	0.401	0.760	0.427
Occ1	Dummy : 1 if professionals, 0 otherwise	0.122	0.327	0.153	0.360
Occ2	Dummy : 1 if technicians, 0 otherwise	0.063	0.243	0.027	0.162
Occ3	Dummy : 1 if government administrators or executive officials, 0 otherwise	0.011	0.103	0.007	0.083
Occ4	Dummy : 1 if managers, 0 otherwise	0.024	0.152	0.040	0.196
Occ5	Dummy : 1 if clerical and related workers, 0 otherwise	0.123	0.328	0.128	0.334
Occ6	Dummy : 1 if sales workers, 0 otherwise	0.069	0.254	0.089	0.285
Occ7	Dummy : 1 if service workers, 0 otherwise	0.139	0.346	0.133	0.339
Occ8	Dummy : 1 if farmers, fishermen, hunters, loggers, or related workers, 0 otherwise	0.049	0.216	0.035	0.184
Occ9	Dummy : 1 if production and related workers, transport equipment operators and laborers (supervisors and foremen), 0 otherwise	0.026	0.160	0.387	0.487
Occ10	Dummy : 1 if production and related workers, transport equipment operators and laborers (other than supervisors and foremen), 0 otherwise	0.351	0.477		
Occ11	Dummy : 1 if occupation is not classified, 0 otherwise	0.023	0.150	0.001	0.035
Rural	Dummy : 1 if rural sample, 0 otherwise	0.376	0.484	0.359	0.480
***Sample 2**: Sample used in the regressions with specifications from column(3) in Table 4*					
Temporary	Dummy : 1 if has temporary employment, 0 if has permanent employment	0.104	0.306	0.132	0.339
Union	Dummy : 1 if union/association member, 0 otherwise	0.834	0.372	0.826	0.379
Publicfirm	Dummy : 1 if working for public or semi-public enterprise, 0 otherwise	0.646	0.478	0.586	0.493
Smallfirm	Dummy : 1 if the number of workers in the enterprise is fewer than 10, 0 otherwise	0.272	0.445	0.298	0.457
SS	Dummy : 1 if covered under Provident Fund (in 1999) or eligible for social security benefits (in 2009), 0 otherwise	0.815	0.388	0.753	0.431
***Sample 3**: Sample used in the regressions with specifications from column (4) in Table 1*					
ChainL _sfamily	*Average ChainL* of other family members of the same gender	1.584	0.328	1.647	0.354

Notes: Sample size is as follows. Sample 1: 32,612 in 1999 (31,805 for *SK3*) and 29,322 in 2009 (29,230 for *SK3*). Sample2: 4,897 in 1999 and 7,169 in 2009. Sample 3: 12,645 in 1999 and 10,570 in 2009. For more details on the occupation dummies (*Occ1*-*Occ11*), see Appendix Table B.4(a).

Appendix B (Supplementary Materials)

Table B.1. Summary Statistics of Key Variables by Industry in 1999

(sorted by *ChainL*)

No.	Sector	Industry name	*ChainL*	*SK1*	*SK1*	*SK2*	*SK3(1)*	*SK3(2)*	*ChainQ _Import*	*ChainQ _Skill*
57	S	Public administration/defense	1.000	10.256	10.112	0.172	0.181	0.174	0.000	
2	P	Forestry/logging	1.169	7.283	2.724	-0.020	-0.150	-0.073	0.013	6.742
54	S	Education/Research	1.206	12.655	12.697	0.305	0.408	0.382	0.008	6.896
5	P	Other mining	1.216	8.759	3.797	0.006	-0.006	0.018	0.040	7.436
51	S	Post/Telecommunication	1.232	10.838	10.840	0.144	0.151	0.126	0.032	7.325
3	P	Fishing	1.240	5.299	3.245	-0.212	-0.323	-0.275	0.012	6.813
6	M	Dairy product	1.246	8.470	7.010	-0.072	-0.131	-0.121	0.016	4.778
47	S	Wholesale/Retail	1.292	8.334	7.178	-0.161	-0.175	-0.217	0.046	7.392
52	S	Finance (banking etc.)	1.320	12.538	12.417	0.296	0.320	0.326	0.010	6.727
1	P	Agriculture	1.415	4.217	3.676	-0.220	-0.320	-0.261	0.027	8.004
53	S	Insurance	1.427	14.007	13.663	0.421	0.461	0.448	0.046	7.882
23	M	Basic chemicals	1.476	10.305	8.701	0.060	0.044	0.061	0.532	7.239
45	S	Gas/Water	1.482	8.599	8.304	0.055	0.072	0.069	0.070	6.083
4	P	Coal/lignite mining	1.503	5.902	5.586	-0.031	-0.077	-0.045	0.018	7.466
56	S	Other services	1.505	8.385	6.323	-0.074	-0.093	-0.106	0.099	7.117
19	M	Wood and its product (excl. furniture)	1.585	5.436	4.068	-0.292	-0.333	-0.292	0.126	6.117
18	M	Wooden furniture	1.658	5.873	5.565	-0.317	-0.283	-0.238	0.129	6.114
29	M	Structural clay/ceramic products	1.670	6.480	3.357	-0.135	-0.168	-0.155	0.263	6.406
31	M	Other non-metalic mineral products	1.673	6.867	4.392	-0.195	-0.242	-0.231	0.246	6.857
48	S	Hotel/Restaurant	1.816	6.465	5.484	-0.256	-0.317	-0.317	0.148	5.303
22	M	Refined petroleum/Coke product	1.823	10.124	9.088	0.113	0.119	0.140	0.155	5.284
34	M	Hand tools/General hardware	1.839	7.974	5.275	-0.259	-0.247	-0.240	0.208	7.576
50	S	Other transport/storage	1.883	7.461	5.530	-0.130	-0.177	-0.133	0.054	7.332
46	S	Construction	1.911	6.841	4.241	-0.143	-0.125	-0.106	0.038	6.821
40	M	Radio/TV/Communicatio n equipment	1.934	10.157	9.595	-0.073	-0.007	-0.041	0.224	7.928
9	M	Beverage/Tobacco	1.939	7.559	4.129	-0.120	-0.197	-0.167	0.056	6.213
13	M	Jute etc. textile	1.948	5.513	5.281	-0.110	-0.204	-0.180	0.098	6.306
36	M	General/Special purpose/office/other non-electrical machinery	1.949	9.841	8.864	0.034	0.047	0.066	0.216	7.735
30	M	Cement	1.954	8.418	7.727	0.013	-0.006	0.030	0.167	6.784
49	S	Railway transport	1.968	8.589	8.450	0.102	0.091	0.126	0.034	8.234
25	M	Pharmaceuticals	1.971	10.931	10.618	0.044	0.021	0.032	0.180	7.304
11	M	Silk textile	1.974	7.481	4.706	-0.079	-0.123	-0.127	0.057	6.310
38	M	Electrical appliances	1.983	8.477	8.275	-0.234	-0.153	-0.184	0.188	7.549
32	M	Basic iron/steel	2.001	9.869	9.100	0.098	0.112	0.132	0.173	7.240
26	M	Other chemicals	2.003	9.945	9.224	0.062	0.047	0.055	0.167	7.253

21	M	Publishing/printing	2.036	8.737	9.023	-0.071	-0.066	-0.089	0.075	7.199
24	M	Fertilizer	2.046	11.423	11.224	0.280	0.338	0.385	0.153	7.131
44	S	Electricity	2.052	9.635	9.401	0.117	0.192	0.185	0.123	7.224
33	M	Basic non-ferrous metals	2.053	8.193	7.774	-0.111	-0.116	-0.138	0.193	7.481
39	M	Electric motors etc.	2.089	9.435	9.348	0.011	0.076	0.055	0.174	7.843
27	M	Rubber products	2.101	9.709	9.110	0.014	0.008	0.047	0.152	6.713
37	M	Wire/cable etc.	2.104	10.275	10.197	0.067	0.146	0.128	0.186	7.620
7	M	Sugar	2.112	8.696	6.716	0.057	0.020	0.042	0.040	4.951
35	M	Other metal product	2.120	7.714	7.136	-0.133	-0.141	-0.140	0.169	7.869
43	M	Other manufacturing	2.134	7.391	6.655	-0.243	-0.203	-0.264	0.161	7.553
41	M	Transport equipment (Motor vehicle etc.)	2.139	10.570	9.964	0.168	0.209	0.206	0.126	7.613
14	M	Other textile products	2.148	6.286	4.972	-0.276	-0.333	-0.303	0.071	6.569
42	M	Other transport equipment	2.160	5.155	5.206	-0.348	-0.388	-0.386	0.125	7.806
55	S	Medical and health services	2.161	10.883	11.355	0.162	0.225	0.218	0.023	7.101
28	M	Plastic products	2.162	9.345	9.616	-0.124	-0.157	-0.138	0.108	7.107
10	M	Cotton/woolen textile	2.183	6.423	5.625	-0.192	-0.274	-0.268	0.069	6.628
17	M	Leather footwear	2.186	7.304	4.118	-0.189	-0.165	-0.169	0.034	6.687
15	M	Wearing apparel	2.239	7.167	6.321	-0.254	-0.256	-0.248	0.083	6.504
20	M	Paper and its products	2.283	8.099	7.904	-0.117	-0.178	-0.153	0.076	7.083
8	M	Other food products	2.284	6.568	5.908	-0.196	-0.242	-0.249	0.054	5.659
12	M	Man-made fiber textiles	2.332	8.253	8.333	-0.175	-0.238	-0.223	0.103	6.904
16	M	Leather and its product (excl. footwear)	2.414	7.797	6.761	-0.249	-0.211	-0.209	0.035	6.747

Notes: P, M, and S in the column "sector" denote primary, manufacturing, and service sectors, respectively.

Table B.2. Industry-Level Skill-Sorting Regression When Using *ChainQ_Skilltotal*: 1999 and 2009 panel

Dep. Var.		All-industry sample		Manufacturing/ Service sample		Manufacturing sample	
		(1)	(2)	(3)	(4)	(5)	(6)
SK1	*ChainL*	-1.045*	-0.892	-0.974*	-1.006	-1.718**	-2.114***
		(0.540)	(0.568)	(0.573)	(0.613)	(0.643)	(0.605)
	ChainQ_Import	-0.616	-0.265	-0.811	-0.517	-2.419	-2.074
		(1.208)	(1.350)	(1.229)	(1.204)	(2.471)	(2.269)
	ChainQ_Skilltotal	0.805***	0.787***	0.664***	0.541***	0.672	0.745
		(0.249)	(0.236)	(0.166)	(0.193)	(0.404)	(0.457)
	R-squared	0.768	0.776	0.782	0.797	0.474	0.523
	N	114	114	104	104	76	76
SK2	*ChainL*	-0.090*	-0.100*	-0.075	-0.091	-0.229***	-0.240***
		(0.050)	(0.054)	(0.058)	(0.067)	(0.063)	(0.063)
	ChainQ_Import	-0.012	-0.026	-0.018	-0.035	-0.255*	-0.211
		(0.107)	(0.096)	(0.110)	(0.098)	(0.130)	(0.133)
	ChainQ_Skilltotal	0.041**	0.047***	0.038**	0.041**	0.069*	0.063
		(0.016)	(0.014)	(0.016)	(0.018)	(0.040)	(0.042)
	R-squared	0.288	0.336	0.249	0.274	0.324	0.347
	N	114	114	104	104	76	76

Notes: "Dep. Var." denotes the dependent variable. Figures are WLS estimates with weight = employment share of each industry averaged over the two periods. In columns (1), (3), and (5), year 2009 and industry dummies are controlled for. In columns (2), (4), and (6), *Import*, *Export*, and *Smallfirm* are additionally controlled for. Robust standard errors clustered by industry are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1

Table B.3. Concordance Table on Industry Classification

(a) 57-Industry Classification (1999 and 2009)

No.	Sector	Industry Name	1999 1998–1999 (1993–1994) IO code	2009 2007–2008 IO code	1999/2009 WIOD code	1999/2009 NIC-1998 codes in 1999–2000 NSS < NIC-2004 code in 2009-2010 NSS >*
1	P	Agriculture	1-17, 19, 20	1-20, 22-24	C1	01, 852 < 01 (excl. 01136), 852 >
2	P	Forestry/logging	21	25		02 < 02, 01136 >
3	P	Fishing	22	26		05, 1512
4	P	Coal/lignite mining	23	27	C2	10
5	P	Other mining	24-32	28-37		11-14 < 11-14, 402 >
6	M	Dairy products	18	21	C3	152
7	M	Sugar	33, 34	38, 39		1542
8	M	Other food products	35-38	40-43		151 (excl. 1512), 153, 154 (excl. 1542)
9	M	Beverage/Tobacco	39, 40	44, 45		155, 16
10	M	Cotton/woolen textile	41-43	46-48	C4	17111, 17113, 17115, 17117, 17121, 17123 < 17111, 17113, 17115, 17117, 17121, 17123, 17131-17133, 17139, 17141, 17142, 17149, 17126, 17129, 17134-17136, 17143 >
11	M	Silk textile	44	49		17112, 17116, 17122 < 17112, 17116, 17122, 17144 >
12	M	Man-made fiber textiles	45	50		17114, 17118, 17124 < 17114, 17118, 17124, 17137, 17145 >
13	M	Jute textile etc.	46	51		17119, 17125
14	M	Other textile products	47, 49	52, 54		1722, 1723, 1729, 173 < 1722, 1723, 1729, 173, 17252-17255, 1724, 17251, 17259 >
15	M	Wearing apparel	48	53		1721, 1810
16	M	Leather and its products (excl. footwear)	55	60	C5	1820, 191
17	M	Leather footwear	54	59		19201, 52601 < 19201, 19209, 52601 >
18	M	Wooden furniture	50	55	C6	36101

19	M	Wood and its products (excl. furniture)	51	56		20
20	M	Paper and its products	52	57	C7	21
21	M	Publishing/printing	53	58		22
22	M	Refined petroleum/Coke products	58, 59	63, 64	C8	23
23	M	Basic chemicals	60, 61	65, 66	C9	2411 (excl. 24113, 24114, 24115)
24	M	Fertilizer	62	67		2412
25	M	Pharmaceuticals	65	70		2423
26	M	Other chemicals	63, 64, 66-68	68, 69, 71-73		242 (excl. 2423), 2413, 24113-24115, 243
27	M	Rubber products	56	61	C10	251, 19202
28	M	Plastic products	57	62		252, 36103
29	M	Structural clay/ceramic products	69	74	C11	2692, 2693
30	M	Cement	70	75		26941, 26942
31	M	Other non-metallic mineral products	71	76		261, 2691, 26943, 26944, 26945, 2695, 2696, 2699 < 261, 2691, 26943, 26944, 26945, 26949, 2695, 2696, 2699 >
32	M	Basic iron/steel	72	77	C12	271
33	M	Basic non-ferrous metals	73,75	78, 80		272, 273
34	M	Hand tools/General hardware	76	81		2893 (excl. 28931)
35	M	Other metal products	74, 77	79, 82		281, 2891, 2892, 28931, 2899, 36102
36	M	General/Special purpose/office/other non-electrical machinery	78-83	83-87	C13	291, 292, 29301, 29302, 29306, 29307, 29309, 30 < 291, 292, 29301, 29302, 29306, 29307, 29309 >
37	M	Wire/cable etc.	85,86	89, 90	C14	313, 314
38	M	Electrical appliances	87	91		315, 29303-29305, 29308, 52602
39	M	Electric motors etc.	84,89	88, 93		311, 312, 319
40	M	Radio/TV/Communication equipment	88, 90	92, 94		32, 52603
41	M	Transport equipment (Motor vehicle etc.)	91-94	95-98	C15	34, 351, 352, 3591
42	M	Other transport equipment	95,96	99, 100		3592, 52605, 3599
43	M	Other manufacturing	97,98	101-105	C16	33, 37, 353, 369, 36104, 36109, 52609, 52604 < 33, 37, 353, 369, 36104, 36109, 52609, 52604, 30 >

44	S	Electricity	100	107	C17	401
45	S	Gas/Water (No Gas in 2009)	101,102	108		402, 403, 410 < 403, 410 >
46	S	Construction	99	106	C18	45
47	S	Wholesale/Retail	107	116	C19, 20, 21	50-52 (excl. 526)
48	S	Hotel/Restaurant	108	117	C22	55
49	S	Railway transport	103	109	C23	601
50	S	Other transport/storage	104,105	110-114	C24, 25, 26	602, 603, 61-63
51	S	Post/Telecommunication	106	115	C27	64
52	S	Finance (banking etc.)	109	118	C28	65, 671
53	S	Insurance	110	119		66, 672
54	S	Education/Research	112	121	C32	73, 80
55	S	Medical and health services	113	122	C33	851
56	S	Other services	111,114	120, 123-129	C29, 30, 34, 35	70-72, 74, 853, 90-93, 95, 99
57	S	Public administration/defense	115	130	C31	75

Notes: P, M, and S in the column "sector" denote primary, manufacturing, and service sectors, respectively. NIC-2004 codes in the 2009–2010 NSS are reported in brackets < > only when the corresponding codes differ from the NIC-1998 codes.
Sources: Created by author based on the industry descriptions of India's IO table, NIC-1998 and NIC-2004 classifications (Ministry of Statistics and Programme Implementation's website[32]), and WIOD.

(b) 54-Industry Classification (2009)

No.	Sector	Industry Name	1993–1994 IO code	2007–2008 IO code	WIOD code	NIC-2004 code in 2009–2010 NSS
1	P	Agriculture, Fishing	1-17, 19, 20, 22	1-20, 22-24, 26	C1	01 (excl. 01136), 852, 05, 1512
2	P	Forestry	21	25		02, 01136
3	P	Coal/lignite mining	23	27	C2	10
4	P	Other mining	24-32	28-37		11-14, 402
5	M	Sugar, Dairy products	18, 33, 34	21, 38, 39	C3	152, 1542
6	M	Other food products	35-38	40-43		151 (excl. 1512), 153, 154 (excl. 1542)
7	M	Beverage/Tobacco	39, 40	44, 45		155, 16

[32] http://mospi.nic.in/Mospi_New/site/inner.aspx?status=2&menu_id=129

8	M	Cotton textile	41, 42	46, 47	C4	17111, 17115, 17121, 17131-17133, 17139, 17141, 17142, 17149, 17126, 17129
9	M	Woolen, Silk, Jute etc. textile	43, 44, 46	48, 49, 51		17113, 17117, 17123, 17134-17136, 17143, 17112, 17116, 17122, 17144, 17119, 17125
10	M	Man-made fiber textiles	45	50		17114, 17118, 17124, 17137, 17145
11	M	Other textile products	47, 49	52, 54		1722, 1723, 1729, 173, 17252-17255, 1724, 17251, 17259
12	M	Wearing apparel	48	53		1721, 1810
13	M	Leather and its products (incl. footwear)	54, 55	59, 60	C5	1820, 191, 19201, 19209, 52601
14	M	Wooden furniture	50	55	C6	36101
15	M	Wood (excl. furniture)	51	56		20
16	M	Paper and its products	52	57	C7	21
17	M	Publishing/printing	53	58		22
18	M	Refined Petroleum/Coke product, Basic chemicals, Fertilizer	58-62, 101	63-67	C8, 9	23, 2411 (excl. 24113-24115), 2412
19	M	Pharmaceutical	65	70		2423
20	M	Other chemicals	63, 64, 66-68	68, 69, 71-73		242 (excl. 2423), 2413, 24113-24115, 243
21	M	Rubber and Plastic products	56, 57	61, 62	C10	251, 19202, 252, 36103
22	M	Structural clay/ceramic products, Cement	69, 70	74, 75	C11	2692, 2693, 26941, 26942
23	M	Other non-metallic mineral products	71	76		261, 2691, 26943, 26944, 26945, 26949, 2695, 2696, 2699
24	M	Basic iron/steel	72	77	C12	271
25	M	Basic non-ferrous metals, Hand tools/General hardware	73, 75, 76	78, 80, 81		272, 273, 2893 (excl. 28931)
26	M	Other metal products	74, 77	79, 82		281, 2891, 2892, 28931, 2899, 36102
27	M	General/Special purpose/other non-electrical machinery	78-81, 83	83-87	C13	291, 292, 29301, 29302, 29306, 29307, 29309
28	M	Electrical Appliance, Batteries	86, 87	90, 91	C14	314, 315, 29303-29305, 29308, 52602
29	M	Electric motors etc.	84, 89	88, 93		311, 312, 319
30	M	Radio/TV/Communication equipment, Electrical cables and wires	85, 88, 90	89, 92, 94		313, 32, 52603

31	M	Motor vehicle & Motor cycles and scooters	93, 94	97, 98	C15	34, 3591
32	M	Ships, boats, and Rail equipment	91, 92	95, 96		351, 352
33	M	Bicycles, cycle-rickshaw, Other transport equip.	95, 96	99, 100		3592, 52605, 3599
34	M	Gems and Jewelry	82, 97, 98	103	C16	3691
35	M	Other manufacturing		101, 102, 104, 105		33, 37, 353, 369 (excl. 3691), 36104, 36109, 52609, 52604, 30
36	S	Electricity	100	107	C17	401
37	S	Gas, Water (No Gas in 2009)	102	108		403, 410
38	S	Construction	99	106	C18	45
39	S	Wholesale/Retail	107	116	C19, 20, 21	50-52 (excl. 526)
40	S	Hotel/Restaurant	108	117	C22	55
41	S	Railway transport	103	109	C23	601
42	S	Land transport	104, 105	110		602, 603
43	S	Water and air transport		111, 112	C24, 25	61, 62
44	S	Auxiliary transport, storage and warehousing		113, 114	C26	63
45	S	Post/Telecommunication	106	115	C27	64
46	S	Finance (banking etc.)	109	118	C28	65, 671
47	S	Insurance	110	119		66, 672
48	S	Education/Research	112	121	C32	73, 80
49	S	Medical and health services	113	122	C33	851
50	S	Computer services	111, 114	124	C29, 30, 35	72
51	S	Business services		123		74 (excl. 7411)
52	S	Community, social, personal services		128	C34	853, 91, 93, 95
53	S	Other services		120, 125-127, 129	C29, 30, 35	70, 71, 7411, 90, 92, 99
54	S	Public administration/defense	115	130	C31	75

Notes: P, M, and S in the column “sector” denote primary, manufacturing, and service sectors, respectively.
Sources: Created by author based on the industry descriptions in India's IO table and NIC-2004 classifications.

Table B.4. Concordance Table on Occupation Classification

(a) 11 Occupation Categories (*Occ1*–*Occ11*) in Appendix A (b)

	Description	NCO-1968 code in 1999–2000 NSS	NCO-2004 code In 2009–2010 NSS
Occ1	Professionals	00, 02, 05, 07, 10-19 (excl. 192, 199)	2 (excl. 223), 312, 313, 324, 33, 347
Occ2	Technicians, etc.	01, 03, 04, 06, 08, 09, 199, 30, 572	223, 3 (excl. 312, 313, 315, 324, 33, 341, 343, 345, 347)
Occ3	Government administrators and executive officials	20, 21	11
Occ4	Managers	22-29, 360, 60	12-13
Occ5	Clerical and related workers	3 (excl. 30, 357, 358, 360, 370, 371)	4, 343
Occ6	Sales workers	4	341, 522, 523, 911
Occ7	Service workers	5 (excl. 541, 572), 192, 357, 358, 370, 371	5 (excl. 522, 523), 315, 345, 912-915
Occ8	Farmers, fishermen, hunters, loggers, and related workers	6 (excl. 60)	6, 92
Occ9	Production and related workers, transport equipment operators and laborers (supervisors and foremen)	71-98 (excl. all 3-digit codes ending with zero (e.g. 710, 720, …, 980))	7, 8, 916, 93
Occ10	Production and related workers, transport equipment operators and laborers (other than supervisors and foremen)	7-9 (excl. those recorded as *Occ9*), 541	
Occ11	Not classified	X	X

Sources: Created by author based on India's NCO-1968 and NCO-2004 codes (Directorate General of Training, Ministry of Skill Development and Entrepreneurship, Government of India website[33]).

[33] http://dget.nic.in/content/innerpage/nco---20php

(b) Occupation Concordance Used when Constructing *SK3*

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
000	69	53
001	73	45
002	75	51
003	74	43
009	69	53
010	224, 225	151, 162
011		
012		
014		
015		
017		
018		
019		
020	43	2
021	53	11
022	55	12
023	57	14
024	48	10
025	45	15
026	47	20, 21
027	56	13
028	218	161
029	59, 235	23, 173
030	217	152
031	214	162
032		153
033		155
034		151
035		162
036		
037	218	161
039	214	162
040	226	163
041		170
042		
043	829	221, 661, 701
044		173
045	226, 829	173, 221, 661, 701
049		
050	78	44, 52
051	83	54
052	79	25
053	77	42
054	76	54
057		
059		
060	223	150
061		
063		
069		

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
070	84	65
071		
072		
073		
074	85	62
075	86	72
076	96	64
077	97	74
078	208	85
079	88, 89	61, 71, 73
080	208	85
081	445	921
082	447	922
083		
084	95	75
085	89	924
086	206	83
087	87	63
088	99, 104, 105	76
089	447	922, 925
090	208	85
091		
092		
093		
094		
095		
099		
100	68	35
101		36
102		34
103	64, 235, 229	3-5, 173
104	386	375
105	68	35, 36
107		
109		
110	166	91
111		
113	166	91
114		
116		
117		
119		
120	23	1
121		
122		
123		
124		
127		
129		

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
130	169	94
131		92, 96
132		96
133	167	93
134	164, 165	32, 33
135	169	96
136	27, 163	56, 174
137	174, 177	100, 954
139	169	96
140	178	31
141		30
142	234	173
143		173
147		
149		
150	154	102-105, 110-116, 120-126, 130-135, 140
151	157	144
152		
153	156	142
154	155	143
155	158	145
156	159	145
157		
158		
159		
160	183	181
161	195	184
162	13, 194	192, 194
163		
169		
170	188	190
171	185	183, 425
172	195	184
173	189	191
177	194	194
179		
180	186	185
181	193	182
182	187	175
183		194
184	194	194
186		
187		
188		
189		
190	176	86
191		90
192	469	933
193	199	180

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
194	235	173
195		
199		
200	4	222
201		
202		
206		
209		
210		
211		
212		
214		
217		
213		
219		
220	22	220, 233, 245
221		
222	22	220, 245
223		
224		
225		
227		
229		
230	7	202
231		
234		
239		
240	22	220, 245
241		
242		
243		
244		
245		
246		
248		
249		
250		
251		
252		
253		
255		
256		
259		
260		
261		
262		
263		
264		
267		
268		

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
269	22	220, 245
290		
291		
292		
293		
294		
295		
296		
299		
300	303	312
301		
302	389	311, 394, 395
303		
304		
305		
306		
307		
309		
310		
311		
312		
315		
316		
318		
319		
320	313	370-372, 376
321	315	391
322		
323	385	345
324	313, 315, 385	345, 370-372, 376, 391
328		
329		
330	337	305
331	276	310
336	276, 337, 383	310, 301, 305
337		
338		
339		
340	308	343, 350
341		
342		
343		
345		
347		
349		
350	379	394, 395
351	365	381
352	319	364
353	329	330
354	338	360

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
355	389	394, 395
356	318	390
357	471	394, 395
358	469	933, 945
359	347, 378, 384, 389	313, 344, 355, 362, 394, 395
360	22, 471	220, 245
361	22, 303	220, 224, 245, 312
362	471	220, 245
363		
364		
366		
368		
369		
370	823	226, 704
371		
372		
374		
375		
377		
379		
380	355	331
381	357	333, 383
384	346, 354-356	331, 332, 361
385		
386		
389		
390	348	385
391	349	384
392		
396	349	384, 385
398		
399		
400	274	281, 282, 285
401	275	284
402	274, 275	281, 282, 284, 285
403		
404		
405		
407		
409		
410	243	281, 282
411	29, 33	205, 225
412	274, 275	281
413		
414		
415		
417		
419		
420		281, 282, 284, 285
421		

<table>
<tr><th>NCO-1968 code in 1999–2000 NSS</th><th>occ1990dd code used in Autor and Dorn (2013)</th><th>1970 U.S. census code used in Yamaguchi (2012)</th></tr>
<tr><td>422</td><td rowspan="3">274, 275</td><td rowspan="3">281, 282, 284, 285</td></tr>
<tr><td>427</td></tr>
<tr><td>429</td></tr>
<tr><td>430</td><td>274, 275, 283</td><td>282-285, 262</td></tr>
<tr><td>431</td><td>277</td><td>264, 266</td></tr>
<tr><td>432</td><td rowspan="6">274, 275</td><td rowspan="6">282-285</td></tr>
<tr><td>433</td></tr>
<tr><td>434</td></tr>
<tr><td>435</td></tr>
<tr><td>437</td></tr>
<tr><td>439</td></tr>
<tr><td>440</td><td>253</td><td>265</td></tr>
<tr><td>441</td><td>254</td><td>270</td></tr>
<tr><td>442</td><td>255</td><td>271</td></tr>
<tr><td>443</td><td>256</td><td>260</td></tr>
<tr><td>444</td><td>274</td><td>261</td></tr>
<tr><td>445</td><td>375</td><td>326, 363</td></tr>
<tr><td>449</td><td>253-256, 274</td><td>260, 261, 265, 270, 271, 363</td></tr>
<tr><td>450</td><td rowspan="8">255</td><td rowspan="8">285</td></tr>
<tr><td>451</td></tr>
<tr><td>452</td></tr>
<tr><td>453</td></tr>
<tr><td>454</td></tr>
<tr><td>455</td></tr>
<tr><td>457</td></tr>
<tr><td>459</td></tr>
<tr><td>490</td><td rowspan="5">274, 275</td><td rowspan="5">281-285</td></tr>
<tr><td>491</td></tr>
<tr><td>493</td></tr>
<tr><td>494</td></tr>
<tr><td>499</td></tr>
<tr><td>500</td><td rowspan="4">22</td><td rowspan="4">230, 245</td></tr>
<tr><td>501</td></tr>
<tr><td>502</td></tr>
<tr><td>509</td></tr>
<tr><td>510</td><td rowspan="6">405</td><td rowspan="6">940, 950, 982</td></tr>
<tr><td>511</td></tr>
<tr><td>513</td></tr>
<tr><td>514</td></tr>
<tr><td>517</td></tr>
<tr><td>519</td></tr>
<tr><td>520</td><td>436</td><td>912, 981</td></tr>
<tr><td>521</td><td>405, 435</td><td>915</td></tr>
<tr><td>522</td><td>434</td><td>910</td></tr>
<tr><td>523</td><td rowspan="6">434-436</td><td rowspan="6">910, 915, 912, 981</td></tr>
<tr><td>524</td></tr>
<tr><td>525</td></tr>
<tr><td>526</td></tr>
<tr><td>528</td></tr>
<tr><td>529</td></tr>
</table>

<table>
<tr><th>NCO-1968 code in 1999–2000 NSS</th><th>occ1990dd code used in Autor and Dorn (2013)</th><th>1970 U.S. census code used in Yamaguchi (2012)</th></tr>
<tr><td>530</td><td rowspan="8">405</td><td>901, 984</td></tr>
<tr><td>531</td><td>984</td></tr>
<tr><td>532</td><td rowspan="6">901, 931, 940, 950</td></tr>
<tr><td>533</td></tr>
<tr><td>534</td></tr>
<tr><td>537</td></tr>
<tr><td>538</td></tr>
<tr><td>539</td></tr>
<tr><td>540</td><td>453</td><td>903</td></tr>
<tr><td>541</td><td rowspan="6">405</td><td>902</td></tr>
<tr><td>542</td><td>780, 785</td></tr>
<tr><td>543</td><td rowspan="4">780, 785, 902</td></tr>
<tr><td>544</td></tr>
<tr><td>547</td></tr>
<tr><td>549</td></tr>
<tr><td>550</td><td rowspan="7">408, 764</td><td>630, 983</td></tr>
<tr><td>551</td><td rowspan="6">611, 630, 983</td></tr>
<tr><td>552</td></tr>
<tr><td>554</td></tr>
<tr><td>556</td></tr>
<tr><td>556</td></tr>
<tr><td>559</td></tr>
<tr><td>560</td><td rowspan="4">457, 458</td><td rowspan="4">935, 944</td></tr>
<tr><td>561</td></tr>
<tr><td>562</td></tr>
<tr><td>569</td></tr>
<tr><td>570</td><td>417</td><td>961</td></tr>
<tr><td>571</td><td>418</td><td>964</td></tr>
<tr><td>572</td><td>36</td><td>215</td></tr>
<tr><td>573</td><td rowspan="2">426</td><td rowspan="2">962</td></tr>
<tr><td>574</td></tr>
<tr><td>575</td><td rowspan="4">427</td><td rowspan="4">960-965</td></tr>
<tr><td>576</td></tr>
<tr><td>577</td></tr>
<tr><td>579</td></tr>
<tr><td>590</td><td>461</td><td>932, 933</td></tr>
<tr><td>591</td><td>469</td><td>211, 933</td></tr>
<tr><td>592</td><td rowspan="5">405, 462, 469</td><td rowspan="5">933, 941, 953</td></tr>
<tr><td>593</td></tr>
<tr><td>595</td></tr>
<tr><td>597</td></tr>
<tr><td>599</td></tr>
<tr><td>600</td><td rowspan="7">475</td><td rowspan="7">802, 821</td></tr>
<tr><td>601</td></tr>
<tr><td>602</td></tr>
<tr><td>603</td></tr>
<tr><td>604</td></tr>
<tr><td>605</td></tr>
<tr><td>609</td></tr>
</table>

<table>
<tr><th>NCO-1968 code in 1999–2000 NSS</th><th>occ1990dd code used in Autor and Dorn (2013)</th><th>1970 U.S. census code used in Yamaguchi (2012)</th></tr>
<tr><td>610</td><td rowspan="18">473</td><td rowspan="18">801</td></tr>
<tr><td>611</td></tr>
<tr><td>612</td></tr>
<tr><td>613</td></tr>
<tr><td>614</td></tr>
<tr><td>616</td></tr>
<tr><td>618</td></tr>
<tr><td>619</td></tr>
<tr><td>620</td></tr>
<tr><td>621</td></tr>
<tr><td>622</td></tr>
<tr><td>623</td></tr>
<tr><td>624</td></tr>
<tr><td>625</td></tr>
<tr><td>626</td></tr>
<tr><td>627</td></tr>
<tr><td>628</td></tr>
<tr><td>629</td></tr>
<tr><td>630</td><td rowspan="13">479</td><td rowspan="13">822-824</td></tr>
<tr><td>631</td></tr>
<tr><td>632</td></tr>
<tr><td>633</td></tr>
<tr><td>635</td></tr>
<tr><td>636</td></tr>
<tr><td>637</td></tr>
<tr><td>639</td></tr>
<tr><td>640</td></tr>
<tr><td>641</td></tr>
<tr><td>643</td></tr>
<tr><td>645</td></tr>
<tr><td>649</td></tr>
<tr><td>650</td><td>779</td><td>690, 692</td></tr>
<tr><td>651</td><td>479</td><td>822-824</td></tr>
<tr><td>652</td><td>451</td><td>755</td></tr>
<tr><td>655</td><td rowspan="4">479</td><td rowspan="4">822-824</td></tr>
<tr><td>657</td></tr>
<tr><td>658</td></tr>
<tr><td>659</td></tr>
<tr><td>660</td><td rowspan="5">496</td><td rowspan="5">450, 761</td></tr>
<tr><td>661</td></tr>
<tr><td>662</td></tr>
<tr><td>663</td></tr>
<tr><td>669</td></tr>
<tr><td>670</td><td rowspan="7">498</td><td rowspan="7">752</td></tr>
<tr><td>671</td></tr>
<tr><td>672</td></tr>
<tr><td>673</td></tr>
<tr><td>676</td></tr>
<tr><td>678</td></tr>
<tr><td>679</td></tr>
</table>

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
680	498	752
681		
682		
683		
684		
686		
687		
688		
689		
710	628	441
711	616	640
712		
713	598	614
714	615	603
715	616, 617	640
716	614	614
717		
718	617	640
719		
720	628	441
721	766	622
722	707	533
723	766	622
724	719	503
725		
726	724	446, 626
727	755	503
728	723	635
729	653	575
730	628	441
731	658, 766	443
732	727	662
733	779	690, 692, 694, 695
734		
735		
737		
739		
740	628	441
741	756	641
742	724	446
743	757	690, 692, 694, 695
744	779	
745		
747	766, 779	622, 666, 690, 692, 694, 695
749		
750	628	441
751	749	670, 674
752	738	672, 681

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
753	749	674
754		
755	739	673
756		
757		671
758	743	674
759	799	610
760	628	441
761	669	444
762		
764		
769		
770	628	441
771	769	501
772		690, 692, 694, 695
773		
774	686	631, 633
775	754	604, 643
776	769	690, 692, 694, 695
777	687, 763	402
778	763	575
779	754, 769	501, 604, 612, 643
780	628	441
781	763, 779	690, 692, 694, 695
782		
783		
784		
785		
786		
787		
788		
789		
790	628	441
791	666	551, 613
792	669	444
793		636
794	645, 185	183, 425, 514
795	744	663
796	668	563
797	669	551, 663
798		
799		
800	628	441
801	669	542
802	669, 745	542, 664
803	669	575
804		
805		

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
807	669	575
808		
809		
810	628	441
811	567	415
812	657	413
813	729, 733	690, 692, 694, 695
814	658	443
815		
816		
817		
819		
820	628	441
821	675	546
822		
823		
824		
825		
827		
828		
829		
830	628	441
831	549, 713	403, 442
832		
833	634	561
834		
835	703	652, 653
836	709	621, 651
837	779	690, 692, 694, 695
838		
839		
840	503	441
841	535	492, 495
842	549, 785	492, 495, 602
843	505, 785	473, 602
844	508, 785	471, 602
845	507, 509, 516, 518, 526, 534, 544, 549	480-482, 470. 492, 495, 502
847	549, 785	492, 495, 602
848		
849		
850	503	441
851	575	430
852	533	492, 495
853	523, 785	492, 495, 602
854	523	485
855	575	430
856	527	552
857		554

NCO-1968 code in 1999–2000 NSS	*occ1990dd* code used in Autor and Dorn (2013)	1970 U.S. census code used in Yamaguchi (2012)
859	523, 533	492, 495
860	228, 467	171, 505
861	228	171
862	228, 467	171, 505
864		
869		
870	628	441
871	585	522
872	783	680
873	653	535
874	597	550
878	653	522, 535, 550, 680
879		
880	628	441
881	535	453
882	535, 649	435, 453
883	649	435
884	535, 649	435, 453
887		
889		
890	628	441
891	589, 677	445, 506
892	675	575
893	766	622
894	649	435
895	789	644
896	756, 779, 799	610, 624, 641
898		
899		
900	628	441
901	734, 755, 779	690, 692, 694, 695
902	719, 779	
903	719, 755, 779	
906	779	
907		
908		
909		
910	628	441
911	765	690, 692, 694, 695
912		
914		
915		
919		
920	628	441
921	736	422
922		
923	734	530
924		434
925	649	435
926		515
927	679	405
928	774	645
929	734	690, 692, 694, 695
930	628	441
931	579	510
932	789	543
933		644
935		
936		
937		
938		
939		
940	628	441
941	535	516
942	658	443
943	753, 779	575
944	684, 779	575, 690, 692, 694, 695
945		
946		
947		
949		
950	558	441
951	563, 594	410, 560
952	588	421
953	595	534
954	599	440
955	584	520
956	593	601
957	589	445
958	595	534
959	583, 599	512, 751
960	628	441
961	696	545
962		666
963	519, 887	642, 764
964	694, 696	545
966		
968		
969		
970	628	441
971	889	753
972	527	554
973	848	424
974	594, 844, 853	412, 436
975	368	392, 610
976	888	625
977	804, 834	763
978		
979		
980	803	441
981	829	661, 701
982		
983	823	704
984	824	456
985	825	712, 713
986	804, 808, 809, 834	703, 710, 714, 715
987	804	763
988	834	780, 785
989		712, 713
990	889	780, 785
991		
992		
993		
994		
995		
996		
997		
998		
999		

Notes: Cells highlighted in gray are those not listed in India's NCO-1968 classification but appearing in the 1999–2000 NSS data. I assign the *occ1990dd* and 1970 U.S. Census codes of the neighboring cell to those codes.

Sources: Created by author based on the occupation descriptions of India's NCO-1968 code, *occ1990dd* code in Autor and Dorn (2013) and Dorn (2009), and 1970 U.S. census codes.