ELSEVIER
www.elsevier.com/locate/worlddev

World Development Vol. 40, No. 5, pp. 956–969, 2012

0305-750X/$ - see front matter

doi:10.1016/j.worlddev.2011.11.017

Skill Distribution and Comparative Advantage: A Comparison of China and India

YOKO ASUYAMA *
Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO), Chiba, Japan

Summary. — This paper empirically examines the different comparative advantages of China and India, in relation to their different skill distribution patterns. By utilizing industry export data on China and India from 1983 to 2000, we find that a country with a greater dispersion of skills (i.e., India, especially in the earlier years) has higher exports in industries with shorter production chains, while a country with a more equal dispersion of skills (i.e., China, especially in the later years) has higher exports in industries with longer production chains. The causal relationship is fairly robust and skill sorting mechanism seems to work behind.

Key words — Asia, China, comparative advantage, India, production chains, skill distribution

1. INTRODUCTION

China and India started economic reform around the early 1980s. Since then, both countries have enlarged their presence in the global economy as a result of their rapid growth and large populations. However, the patterns of economic development in these two rapidly emerging economic giants seem fairly different. Economic development in China has been driven by growth in manufacturing industries and exports, whereas the Indian economy has been fueled by growth in service industries and exports such as software, business process outsourcing, and call center services. As of 2005, the shares of manufacturing and service in GDP were 33% and 41% in China, while 15% and 53% in India (World Bank, 2010). The 2009 export value and share in total merchandise exports of manufacturing goods were 1125 billion USD and 94%, respectively in China, while 107 billion USD and 66% in India. In contrast, in 2008, computer and information service exports in India amounted to 36 billion USD, 5.8 times larger than in China (WTO,2010a, 2010b). As will be examined in this paper, the export patterns of China and India also differ within manufacturing sector.

Why have China and India attained these different comparative advantages? This paper aims to provide an answer to this question. In particular, the analysis empirically examines the different patterns of comparative advantages of China and India that result from the differences in skill distribution in each country. It is notable that China historically has been enjoying a much more equal skill distribution with a larger proportion of semi-skilled workers who are equipped with a primary and lower secondary level education. By contrast, the skill distribution in India is characterized by a large number of illiterate populations and relatively large proportion of skilled people with upper secondary and post-secondary education. As of 2005, the shares of employed people who were illiterate or have only received education below the primary level (Level 1), primary and lower secondary education (Level 2), and upper secondary and post-secondary education (Level 3) were 50%, 30%, and 21% in India, while 8%, 73%, and 19% in China. Furthermore, the proportion of the most high-skilled workers who have attained postgraduate (or above) education was 1.5% in India compared with 0.2% in China. [1] Asuyama (2011a) has raised several factors explaining why these different skill distribution patterns emerged between China and India. However, in the present paper, skill distribution is treated as exogenous. The possibility of endogeneity of skill distribution is tested in the robustness analysis.

By utilizing industry export data on China and India from 1983 to 2000, this paper empirically shows that a country with a relatively even skill distribution, that is China, has more exports in industries with longer production chains, whereas a country with a relatively unequal skill distribution, that is India, has more exports in industry with shorter production chains. In this paper, the length of production chains which varies across industry is defined by how many units of domestic nonlabor inputs are required in order to produce one unit of industry output. It measures the size of production linkage with the domestic intermediate input industries. For example, automobile industry has longer production chains since producing a car requires a huge amount of parts, including both software and hardware. By contrast, software industry has shorter production chains, because writing a software program mainly requires labor and computers. As can be easily imagined from this simple example, production chains generally tend to be longer in manufacturing compared with agricultural, mining, and service sectors, although variation also exists within each sector. Although the empirical results of this paper are mostly based on nonservice industries, they indicate that the difference in skill distribution between China and India has influenced the patterns of their comparative advantages. This finding is fairly robust across different specifications, including those which examine manufacturing samples only, correct for selection bias, and control for infrastructure factors.

The rest of the paper is organized as follows. Section 2 presents hypothesized mechanisms and related literature. Section 3 explains the empirical methodology. Section 4 describes the

* I am grateful to three anonymous referees, Moriki Ohara, and Hideaki Goto for helpful comments. I also thank Satoshi Inomata and Hiroshi Kuwamori who provided me with the "Asian International I/O Table" and advised me on the IO Table. I also thank IDE-JETRO for its financial support. All remaining errors are my own. An earlier version of this paper was released as an IDE discussion paper. Final revision accepted: September 27, 2011.

data and explains the construction of the key variables. Section 5 presents the estimation results, including several analyses of robustness. Section 6 concludes.

2. HYPOTHESIZED MECHANISMS AND RELATED LITERATURE

This section briefly presents two hypothesized mechanisms which can explain the causal linkage among skill distribution, length of industry's production chain, and comparative advantage. To begin with, consider a world in which there are two countries A and B and each country has two industries, X and Y. Each country has a uniform distribution of workers' skill with the same mean skill level $\overline{q_L}$, but country B has a more unequal skill distribution, the mean-preserving spread of country A's skill distribution. Industry differs in the number of nonlabor inputs n, that is, the "length of production chains" or the size of production linkage, required to produce one quality-unadjusted output. Industry Y has more n than industry X does. Production function of each industry is supermodular, exhibiting complementarity among input quality.

The first possible mechanism is closely related to the one proposed by Grossman (2004). Let $f_k(q_L, g_k(q_{NL1}, \ldots, q_{NLn}))$ be industry k's quality-adjusted output produced by a team of one assembling/processing worker and n nonlabor inputs. q_L and q_{NLi} indicate skill level of the worker and quality level of nonlabor input i, respectively. As an example, consider the case where $f_k = \sqrt{q_L}\prod_{i=1}^{n}\sqrt{q_{NLi}}$, which is an O-ring type production function proposed by Kremer (1993). Furthermore, let us assume that q_{NLi} has been produced by a worker and that producing perfect-quality nonlabor input requires a certain minimum skill $q_{Low} < \overline{q_L}$. q_{NLi} equals to the probability of success in producing perfect-quality nonlabor input, which corresponds to the proportion of workers in the economy with skill level above q_{Low}. It follows that q_{NLi} becomes smaller in country B which has a greater proportion of very low-skilled workers. Moreover, increase in n leads to smaller $g_k = \prod_{i=1}^{n}\sqrt{q_{NLi}}$ (and thus having larger negative impact on f_k) in both countries, through accumulating the defect occurred while each nonlabor input is produced.

The above setting is very close to the one in Grossman (2004). In his model there are two industries, X′ and Y′. Industry Y′ is characterized by team production (e.g., automobile industry) involving two complementary tasks performed by a team which consists of one manager and one worker. He assumes imperfect labor contract in industry Y′, where an individual's contribution to the output is not observable and team's output is not verifiable. By contrast, an individual's contribution to firm output can be measured perfectly in the other industry X′ (e.g., software industry) which does not involve team production. High-skilled individuals are discouraged from entering industry Y′, where industry productivity, and thus the wages offered, are dragged down by the lower skilled workers. Most talented individuals are sorted into industry X′ where they can earn wages according to their own skill levels. Furthermore, this tendency is much stronger in country B because wages in industry Y′ are further dragged down due to larger proportion of low-skilled workers in the economy. Consequently, Grossman (2004) shows that employment and output of X′ expand in country B, which eventually exports the goods of industry X′ that are produced by the most skilled people. By contrast, country A exports the goods of industry Y′.

If we replace the skill level of a "worker" in a team of Grossman (2004) with the aggregate quality of nonlabor inputs g_k, our X and Y become analogous to Grossman's X′ and Y′. As a result, we can expect similar skill sorting mechanism and reach a similar conclusion by assuming that true contribution of worker and nonlabor inputs is not observable and team's output is not verifiable. In my case, a team in industry Y is constructed by labor and skill-embodied nonlabor inputs. High-skilled workers are sorted into industry X, because in industry Y, their wages are dragged down by the quality of nonlabor inputs g_k.[2] Furthermore, g_k is lower in country B. As in Grossman (2004), country B exports the goods of X, while A exports the goods of Y. Since the negative impact of g_k on industry productivity and wages also becomes worse as n increases, we can expect that country B has a comparative advantage in industries with smaller n (shorter production chain), while country A has one in industries with larger n (longer production chain). However, since the model of Grossman (2004) is based on two-county, two-industry model with zero n in one industry, whether the above logic applies to the multi-county, multi-industry world with variant positive n, is an empirical question.

The second possible mechanism is based on the model proposed by Bombardini, Gallipoli, and Pupato (2009). They theoretically and empirically show that a country with greater dispersion of human capital has more exports in industries with a lower degree of complementarity among workers' skills. They construct multi-country, multi-industry model in which all industry output is supermodular in workers' skill and produced by a mass of workers who inherit country's skill distribution. Unlike Grossman, skill sorting does not occur. They assume that workers' skill is ex-ante unobservable and workers are randomly matched with firms. Industries only differ in the degrees of skill complementarity, resulting from exogenous production technologies. As complementarity becomes higher, the industry's output declines with greater magnitude in a county with greater skill dispersion. In our setting, increase in n (the length of production chains) can play similar role in magnifying the negative impact on industry output, when no skill sorting is assumed. Let us assume that quality of each nonlabor input is determined similarly to the one in the first mechanism. Or alternatively, we can assume that each nonlabor input is produced by a mass of workers whose skill distribution is the same as the entire economy and its quality is supermodular in workers' skill. In this setting, the negative impact of nonlabor inputs on industry output becomes larger (and thus, industry output becomes smaller), as n increases, or as country's skill distribution becomes more disperse. By applying the logic of Bombardini *et al.* (2009), we can conclude that country A exports the goods of industry Y, while country B exports the goods of industry X.

This paper contributes to the literature in two ways. First, the causal impact of the different skill distribution patterns in China and India on the development patterns in each country is identified empirically. To my knowledge, only a limited number of academic studies have explained why China and India differ in their comparative advantages (Gregory, Nollen, & Tenev, 2009; Lo & Liu, 2009; Ohara & Lin, 2011). Although the existing studies (Lo & Liu, 2009, in particular) regard the skill level of workers as an important factor that influences the pattern of development of China and India, none of them have provided empirical evidence by identifying causal relationship.

Second, this paper proposes a new mechanism which links country skill distribution to comparative advantage via the length of industry production chains, and empirically indentifies its causal relationship. This contributes to the literature exploring the sources of comparative advantage. As Chor

(2010) summarizes the recent surge of empirical studies on sources of comparative advantages, differences between countries in productivity (as predicted by the Ricardian model), factor endowments (as predicted by the Heckscher–Ohlin model), and institutions have been identified as sources of comparative advantage (for more details, see the introduction of Chor (2010) and the papers cited). With regard to the impact of skill distribution on comparative advantage, a few theoretical and empirical studies exist (Bombardini *et al.,* 2009; Grossman, 2004; Grossman & Maggi, 2000; Ohnsorge & Trefler, 2007; Tang, 2010). Although the idea of the present paper is closely related with Grossman (2004) and Bombardini *et al.* (2009), none of the existing studies characterize industry by the length of production chains which affects the quality of nonlabor inputs. In addition, among those studies, only Bombardini *et al.* (2009) and Tang (2010) empirically examine the linkage between country skill distribution and comparative advantage. Furthermore, as explained in Section 4(b), the match index in the present paper has some advantages over previous specifications. By using panel data, which is not common in the literature, and conducting a series of robustness checks, this paper also increases the credibility of the results obtained.

3. EMPIRICAL STRATEGY

I mostly follow the estimation strategy of Bombardini *et al.* (2009). Their estimation equation modifies the gravity equation, which aims to explain the size of bilateral trade flows by various trade barriers as examined in Helpman, Melitz, and Rubinstein (2008). Similar empirical strategies are employed in several recent studies such as Chor (2010), Levchenko (2007), Nunn (2007), and Cuñat and Melitz (2007), which try to detect the sources of comparative advantage by estimating industry trade flows. In this paper, I primarily estimate the following equations:

$$\ln(Export_{xmit}) = \beta MatchIndex_{xit} + X\gamma + FE + \varepsilon_{xmit}$$

where $\ln(Export_{xmit})$ denotes the logarithm of exports from exporter x (i.e., China or India) to importer m in industry i at period t (divided by the product of GDPs of exporter x and importer m); $MatchIndex_{xit}$ is a measure to indicate how well the skill dispersion of exporter x matches with the length of the production chains of the domestic industry i at period t; X denotes other control variables as will be mentioned in Section 4(c). *FE* indicates various fixed effects: exporter-time fixed effects and importer-industry fixed effects (FE1), exporter-industry fixed effects and importer-time fixed effects (FE2), and exporter–importer-industry fixed effects and time fixed effects (FE3). Our main focus is on the coefficient β as will be explained in Section 4(b).

4. DATA

(a) *Exports*

The industry export flow data on China and India are from the "National Bureau of Economic Research-United Nations (NBER-UN) Trade Data, 1962–2000" constructed by Feenstra and Lipsey (for details of the dataset, see Feenstra, Lipsey, Deng, Ma, & Mo, 2005). I extract the values of export of China and India from the dataset, and convert them into real 2000 USD values by deflating with the implicit GDP deflator of each country and fixing both countries' exchange rates as those of year 2000. Then, the original 4-digit Standard International Trade Classification (SITC codes) in the export data is converted to the 19 industry classifications used in my analysis. [3] In order to smooth the year-to-year fractionalization of exports, I use the 3-year average export for the four periods: 1983–85, 1988–90, 1993–95, and 1998–2000. [4] These 3-year average industry exports are divided by the product of GDPs (real 2000 USD price) of exporter x and importer m to control for the size of both economies as sometimes done in standard gravity equations (e.g., Anderson & Wincoop, 2003). The number of importers used in the present empirical analysis amounts to 175.

(b) *Match index of country skill dispersion and industry length of production chains*

First, skill dispersion index is constructed from the "International Data on Educational Attainment: Updates and Implications" constructed by Barro and Lee (for details, see Barro & Lee, 2000). Using their distribution data on educational attainment for the populations over age 15 of China and India, I construct skill dispersion index, CV_{xt}, which is the coefficient of variation of skill computed as follows [5]: $CV_{xt} = \sqrt{\sum_e[(YEDU_{ext} - AVG_{xt})^2 P_{ext}]}/AVG_{xt}$, where AVG_{xt} is average years of education of the population in country x at time t, computed as $AVG_{xt} = \sum_e YEDU_{ext}P_{ext}$. Subscript e denotes the level of educational attainment (no schooling, primary, secondary, and post-secondary); $YEDU_{ext}$ is the allocated years of education for each schooling level (0, 6, 12, and 16 years for China, and 0, 5, 12, and 16 years for India, respectively, considering the typical years of schooling in each country); and P_{ext} denotes the proportion of population with educational attainment level e. [6] I also experimented with two different skill dispersion indices, that is, the Gini coefficient ($Gini_{xt}$) and the proportion of semi-skilled population with primary and secondary education level (MID_{xt}). The three skill dispersion indices and other related skill measures are reported in Table 1. In order to minimize the possibility of reverse causality, following the treatment of Bombardini *et al.* (2009), I use the measure of skill dispersion that is 8–10 years before the year in which the exports occur. [7] All three skill dispersion indices indicate that skill distribution of China is more equal compared with India, and that the degree of skill inequality declined during 1975–90 in both countries. Since the main empirical findings of the present paper generally do not depend on which skill dispersion index we use, this paper only presents the results obtained when using CV_{xt}.

As an index for the length of production chains of industry i of exporter x at time t, the column sum of the Leontief inverse coefficient of each industry ($Leontief_{xit}$) computed from the input–output (IO) tables of China and India is used. It is computed as follows: $Leontief_{xit} = \sum_k leon_{xkit}$, where $leon_{xkit}$ is the Leontief inverse coefficient in cell ki of exporter x's IO table at time t. Subscripts k and i denote row and column of the IO table, respectively. The Leontief inverse coefficient matrix L comprised of $k*i$ $leon_{xkit}$ s is computed as $L = (I - A_d)^{-1}$, where I is the identity matrix; A_d is the input coefficient matrix for domestic inputs, in which the coefficient in cell ki is the amount of domestic input of industry k used by industry i divided by the output of industry i. In sum, $Leontief_{xit}$ measures how many units of domestic nonlabor input industry i requires, both direct and indirect, to produce one unit of output in industry i. [8] I use this $Leontief_{xit}$ as a proxy for the length of production chains of industry i. It should be noted that only domestic inputs are used, since the quality of imported input

Table 1. *Skill dispersion indices of China and India (for the population over age 15)*

Exporter	Year	Average years of education	Percentage of population with no schooling	Percentage of population with post-secondary education	*CV*	*Gini*	*MID*
China	1975	4.380	40.2	0.9	0.925	0.502	58.9
	1980	4.760	34.0	0.9	0.829	0.452	65.0
	1985	4.940	31.5	1.2	0.792	0.431	67.4
	1990	5.850	22.2	1.9	0.653	0.352	75.9
India	1975	2.700	62.6	2.1	1.535	0.706	35.3
	1980	3.270	66.6	2.4	1.561	0.731	31.1
	1985	3.640	61.6	2.8	1.436	0.699	35.5
	1990	4.100	55.8	3.3	1.295	0.656	41.0

Sources: Average years of education: Barro and Lee (2000); other variables: computed by author from Barro and Lee (2000). See Section 4(b).

is assumed to be relatively good and is not affected by the skill of domestic workers. Industry's intensity of imported input usage is separately controlled in regression analyses.[9]

$Leontief_{xit}$ of China is computed from the "Asian International I/O Table" of 1985, 1990, 1995, and 2000, constructed by the Institute of Developing Economies, Japan External Trade Organization (IDE-JETRO). For India, $Leontief_{xit}$ is computed from the "Input–Output Transaction Table" of 1983–84, 1989–90, 1993–94, and 1998–99, published by the Central Statistical Organisation (CSO) of India.[10] Twenty-four industry classifications (see Table 2) are used, as only 24 industry classifications are available for China for all years.[11]

Table 2 reports the four-period average $Leontief_{xit}$. First, as expected, the results show that the length of production chains tends to be longer in manufacturing compared to the agricultural, mining, and service industries in both countries. Second, China tends to have developed longer domestic production chains compared with India.[12] It is possible that Indian firms have more incentives to shorten the length of domestic production chains in order to reduce the negative impact from lower-skilled workers by increasing imported inputs. In this case, the impact of country's skill distribution on comparative advantage is further strengthened through affecting firms' decision on the length of domestic production chains. It is also possible that industry technology is different across countries due to other reasons. The previous studies mentioned earlier constructed industry characteristics using data from only one country (i.e., the United States), and assumed the same industry structure across countries. However, it seems more appropriate to utilize country-specific industry characteristics, as long as the endogeneity problem is avoided.

Finally, $MatchIndex_{xit}$ (of industry i of exporter x at time t) is constructed as the product of the standardized skill dispersion index (of exporter x at 8–10 years before time t) and the standardized index for the length of production chains

Table 2. *Four-period average $Leontief_{xit}$ and $MatchIndex(CV)_{xit}$*

Sector	Industry	4-period average *Leontief* China	(Rank)	India	(Rank)	4-period average *MatchIndex(CV)* China	(Rank)	India	(Rank)
Agri-cultrue	1 Paddy	1.687	(21)	1.572	(15)	-0.530	(22)	0.728	(10)
	2 Other agricultural products	1.592	(23)	1.408	(20)	-0.628	(23)	1.088	(5)
	3 Livestock and poultry	1.907	(18)	1.873	(12)	0.041	(18)	0.022	(15)
	4 Forestry	1.553	(24)	1.154	(23)	-0.781	(24)	1.666	(2)
	5 Fishery	1.685	(22)	1.223	(21)	-0.437	(21)	1.528	(4)
Mining	6 Crude petroleum and natural gas	1.791	(19)	1.192	(22)	-0.275	(20)	1.601	(3)
	7 Other mining	2.185	(13)	1.449	(19)	0.604	(13)	1.002	(6)
Manu-facturing	8 Food, beverage and tobacco	2.233	(12)	1.881	(11)	0.698	(12)	0.106	(13)
	9 Textile, leather, and the products thereof	2.458	(4)	2.186	(1)	1.164	(6)	-0.555	(24)
	10 Timber and wooden products	2.313	(10)	1.768	(14)	0.940	(9)	0.333	(11)
	11 Pulp, paper and printing	2.336	(9)	2.051	(6)	0.864	(10)	-0.295	(20)
	12 Chemical products	2.365	(7)	2.054	(5)	1.013	(8)	-0.308	(21)
	13 Petroleum and petro products	2.134	(14)	1.563	(16)	0.479	(15)	0.774	(9)
	14 Rubber products	2.235	(11)	2.101	(3)	0.754	(11)	-0.333	(22)
	15 Non-metallic mineral products	2.441	(6)	1.841	(13)	1.185	(5)	0.148	(12)
	16 Metal products	2.555	(2)	2.137	(2)	1.457	(1)	-0.512	(23)
	17 Machinery	2.456	(5)	1.926	(9)	1.189	(4)	-0.006	(17)
	18 Transport equipment	2.509	(3)	2.055	(4)	1.342	(3)	-0.246	(19)
	19 Other manufacturing products	2.361	(8)	1.929	(8)	1.015	(7)	0.081	(14)
Service	20 Electricity, gas, and water supply	2.122	(15)	2.002	(7)	0.507	(14)	-0.165	(18)
	21 Construction	2.570	(1)	1.904	(10)	1.431	(2)	0.006	(16)
	22 Trade and transport	1.999	(16)	1.515	(17)	0.206	(16)	0.855	(8)
	23 Services	1.953	(17)	1.490	(18)	0.102	(17)	0.942	(7)
	24 Public administration	1.732	(20)	1.000	(24)	-0.173	(19)	1.997	(1)
	Average	2.132		1.720		0.507		0.436	

Note: Cells are highlighted when the rank is less than or equal to 12.

(of industry i of exporter x at time t), multiplied by negative one (-1).[13] The key here is that by standardizing both the skill dispersion index and $Leontief_{xit}$, $MatchIndex_{xit}$ is constructed so that it becomes larger either when higher skill dispersion and shorter production chains are multiplied (matched) or when lower skill dispersion and longer production chains are multiplied (matched).[14] In this way, we can simultaneously test whether a country with higher skill dispersion has more exports in industries with shorter production chains and whether a county with lower skill dispersion has more exports in industries with longer production chains. A positive coefficient for the $MatchIndex_{xit}$ indicates that exports become larger when skill distribution and the length of production chains match better, thus supporting my hypothesis. This $MatchIndex_{xit}$ has an advantage over the specification used by Bombardini *et al.* (2009), which does not simultaneously test the two directions of comparative advantage.

Table 2, which displays the four-period average of $MatchIndex(CV)_{xit}$, shows that China has a relatively larger $MatchIndex(CV)_{xit}$ in most of the manufacturing industries, which are characterized by longer production chains, while India has larger $MatchIndex(CV)_{xit}$ in industries with relatively shorter production chains including agricultural, mining, and services, and some manufacturing industries such as "Timber and wooden products," "Petroleum and petro products," and "Nonmetallic mineral products."

(c) *Other control variables*

The first group of additional control variables accounts for various conventional trade barriers between exporter–importer pairs. They include logarithm of distance (ln(*Distance*)_*xm*); the presence of colonial ties (*Colonial_xm*); geographically contiguity (*Contiguous_xm*); shared legal systems, languages, and religions (*Legalsystem_xm*, *Language_xm*, and *Religion_xm*); and the number of exporter/importer who are members of GATT or WTO (*GATT_WTO_xm*). Those variables are mostly constructed from Helpman *et al.* (2008), CEPII's data, and Barro and McCleary (2005).

The second group of control variables includes endowment characteristics of the exporting country and its industries. These control variables include the imported input ratio of industry (*ImportRatio_i*); capital intensity of the exporting country (*Kintensity_x*); the interaction term of *Kintensity_x* with capital intensity of industry (*Kintensity_x∗i*); educational level of population in the exporting country (*StdAvgEdu_x*); skill intensity of industry (*Skillintensity_i*); the interaction term of *Skillintensity_i* with the skill intensity of the exporting country (*Skillintensity_x∗i*). *Kintensity_x∗i* and *Skillintensity_x∗i* are added to control for effects predicted by the Heckscher–Ohlin model that a country exports relatively more in industries using relatively more abundant factors in the country. Data on the skill distribution of workers at a detailed industry-level are only available for India. Thus, *Skillintensity_i* and *Skillintensity_x∗i* are added as controls in regressions, when restricting the sample to India. More detailed explanations on the above control variables are provided in Appendix Table 7.

5. ESTIMATION RESULTS

(a) *Baseline results*

The baseline regression results with three types of fixed effects mentioned in Section 3 are reported in Table 3. The estimates in columns (1), (3), and (5) use all exports in the 19 nonservice industries of China and India. However, exports from primary industries might be substantially affected by inputs such as land, weather, and natural resources, which are not included as inputs in the IO tables nor affected by worker skill levels. Considering those unobserved factors on the primary industries, specifications (2), (4), and (6) restrict the sample to only the 12 manufacturing industries. The regression result reported in column (7) is based on the sample of the bilateral nonservice industry export data of 103 countries. Although the sample is cross-section (1998–2000 average), and $Leontief_{xit}$ is computed from the 1997 IO table of the United States and applied for all countries as in the literature, the result serves as reference point and contributes to supporting those obtained from China and India sample, in which the variation in skill dispersion is relatively small.[15] Results based on the India sample with *Skillintensity_i* and *Skillintensity_x∗i* additionally controlled are reported in columns (1)–(4) of Appendix Table 8.

Consistent with my hypothesis, the estimated coefficients for $MatchIndex(CV)_{xit}$ are positive and statistically significant in all specifications. For example, in column (1), a unit increase in $MatchIndex(CV)_{xit}$ is associated with a 36.3% (=[exp(0.310) − 1] ∗ 100) increase in industry exports. Similarly, a unit increase in $MatchIndex(CV)_{xit}$ in columns (2)–(6) is associated with 10.8%, 50.8%, 39.7%, 33.3%, and 26.9% increase in industry exports, respectively. Even under the most modest estimate of 10.8%, for instance, if the $MatchIndex(CV)_{xit}$ of the machinery industry in India (−0.229) had been the same as that of China (1.200) in the third period, mainly by achieving more even skill distribution, Indian exports in machinery would have been larger by 15.8% point (=[exp{0.103 ∗ [1.200 − (−0.229)]} − 1] ∗ 100) after controlling for various other determinants. The statistically significant coefficient of $MatchIndex(CV)_{xit}$ in column (7) makes the results obtained from China and India sample robust.

The estimated coefficients for standard trade barriers between exporter-importer paris exhibit the predicted signs except for GATT/WTO membership based on the China and India sample. Distance is negatively associated with exports. Geographic continuity, colonial ties, common legal systems, languages, and religions are all positively associated with exports, although not all estimates are statistically significant. Somewhat unexpectedly, capital intensity of the exporting country is negatively associated with exports in most columns. The average educational level of the population in the exporting country is positively associated with exports. The estimated positive relationship between exports and the interaction terms of capital/skill intensity of exporting country and industry confirms the prediction of the Heckscher–Ohlin model. The coefficient of industry ratio of imported input is positive with FE2 and FE3, but negative with FE1.

(b) *Robustness analyses*

(i) *Test for endogeneity*

The patterns of industry exports may influence the skill distribution and the length of industry production chains of the exporting country. In addition to such a reverse causality problem, $MatchIndex(CV)_{xit}$ might be correlated with unobserved factors in the error term. Such endogeneity problems of the match index would lead to the biased estimates for the coefficients of $MatchIndex(CV)_{xit}$ in Table 3.

In order to take into account the possibility of endogeneity, I construct an instrument for $MatchIndex(CV)_{xit}$ and test whether $MatchIndex(CV)_{xit}$ is endogenous or not.[16] The instrument ($MatchIndex(IV)_{xit}$) is constructed similarly to

Table 3. *Determinants of comparative advantage (baseline results)*

Sample	China and India						(Ref.) World, nonservice (1998–2000 average)
	Nonservice (1)	Manufacturing (2)	Nonservice (3)	Manufacturing (4)	Nonservice (5)	Manufacturing (6)	(7)
MatchIndex(CV)_x∗i	0.310***	0.103**	0.411***	0.334***	0.287***	0.238***	0.332***
	(0.025)	(0.042)	(0.056)	(0.062)	(0.072)	(0.075)	(0.020)
ImportRatio_i	−0.004	−0.010***	0.026***	0.009	0.028***	0.011	
	(0.003)	(0.003)	(0.006)	(0.006)	(0.007)	(0.007)	
Kintensity_x			−0.016***	−0.012***	−0.001	0.006	
			(0.003)	(0.004)	(0.006)	(0.007)	
Kintensity_x∗i	0.000	0.000*	0.000***	0.000	0.000***	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
StdAvgEdu_x			0.402***	0.340**	−0.006	−0.139	
			(0.121)	(0.146)	(0.211)	(0.245)	
ln(*Distance*)_*xm*	−0.868***	−0.787**	−0.917***	−0.808**			−0.932***
	(0.303)	(0.345)	(0.263)	(0.316)			(0.022)
Contiguous_xm	0.796*	0.624	0.824**	0.867**			0.429***
	(0.456)	(0.493)	(0.345)	(0.393)			(0.080)
Legalsystem_xm	0.113	0.112	0.071	0.089			0.179***
	(0.117)	(0.128)	(0.096)	(0.107)			(0.025)
Colonial_xm	0.184	0.497**	0.305	0.507**			0.616***
	(0.212)	(0.229)	(0.207)	(0.224)			(0.074)
Language_xm	0.056	0.113	0.071	0.083			0.191***
	(0.157)	(0.173)	(0.131)	(0.147)			(0.039)
Religion_xm	0.000***	0.000***	0.000***	0.000***	0.002***	0.003***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
GATT_WTO_xm	−0.141	−0.112	−1.956***	−2.000***	−0.104	−0.081	1.723***
	(0.150)	(0.158)	(0.171)	(0.182)	(0.171)	(0.182)	(0.412)
Fixed effects	FE1	FE1	FE2	FE2	FE3	FE3	FE4
Number of observations	11,017	8555	11,017	8555	11,017	8555	72,822
R-squared	0.757	0.752	0.625	0.649	0.857	0.854	0.427
F-statistics	45.06	52.29	139.09	152.73	24.70	30.48	222.19

Notes: FE1 indicates exporter-time fixed effects (FE) and importer-industry FE. FE2 indicates exporter-industry FE and importer-time FE. FE3 indicates exporter–importer-industry FE and time FE. FE4 indicates exporter FE, importer FE, and industry FE. The dependent variable is the logarithm of 3-year average exports (divided by [GDP of exporter x ∗ GDP of importer m]) from exporter x to importer m in industry i. Robust standard errors, clustered by exporter–importer pair are reported in parentheses. In (7), dummies which indicate whether exporter x and importer m belong to a common regional trade arrangement (*RTA*) and share a currency (*CU*; taken from Head, Mayer, and Ries (2010)), and number of islands and landlocked countries in exporter–importer pair (*islands* and *landlocked*; taken from Helpman *et al.* (2008)) are also controlled. Those control variables are not included in (1)–(6) due to either perfect collinearity or a very small variation.
* 10% significance level.
** 5% significance level.
*** 1% significance level.

$MatchIndex(CV)_{xit}$ by multiplying the two standardized variables. The first is equal to one minus the 3-year average ratio of primary and secondary school enrollment to the population, which predates the skill dispersion measures by 10 years. The other is $Leontief_{xit}$ computed for Thailand using the IDE-JETRO's "Asian International I/O Table" of 1985, 1990, 1995, and 2000. Thailand is chosen because it is a relatively large developing country in Asia and exports varieties of goods including both manufacturing and agricultural goods. Conducting the Hansen (or Sargan) test which examines the validity of the instrument is not feasible, because the test requires more than one instrument. However, since both variables influence the patterns of industry exports only indirectly through CV_{xt} and $Leontief_{xit}$ by construction, $MatchIndex(IV)_{xit}$ satisfies the exclusion restriction necessary in order for the instrument to be valid.

Using the above instrument, endogeneity of $MatchIndex(CV)_{xit}$ is tested following the method of Wooldridge (Wooldridge, 2006, pp. 532–533). In the first stage, $MatchIndex(CV)_{xit}$ is regressed on $MatchIndex(IV)_{xit}$ and all the exogenous variables in the baseline estimation (columns (1), (3), and (5) of Table 3). In the second stage, the residual obtained from the first stage regression is added to the baseline estimation equation. If the coefficient of the first stage residual is statistically different from zero, we can conclude that our match index is endogenous. The first-stage estimation results show that our instrument is very strong and statistically significant at the 1% level (the F statistics of $MatchIndex(IV)_{xit}$ is 4481, 1185, and 243, depending on specification). However, the second-stage estimation results indicate that we cannot conclude that our match index is endogenous, since the estimated coefficient for the first stage residual is not significantly different from zero, even at a 10% significance level. Thus, I assume that our match index is exogenous.

(ii) *Selection corrections*

The baseline estimations omit observations with export values of zero, due to their log transformation. However, observations with zero exports constitute about half of the export sample (e.g., 54.0% in the nonservices export sample and 43.4% in manufacturing export sample), although both China and India have positive values for total exports in each industry. Zero *bilateral* industry trade indicates that firms decide not to export to a certain country due to the presence of high

Table 4. *Selection corrected estimates (PPML estimation)*

Sample	China and India						(Ref.) World, nonservice (1998–2000 average)
	Nonservice (1)	Manufacturing (2)	Nonservice (3)	Manufacturing (4)	Nonservice (5)	Manufacturing (6)	(7)
MatchIndex(CV)_x∗i	0.163**	0.174*	0.365**	0.038	0.315*	−0.018	0.359**
	(0.073)	(0.095)	(0.173)	(0.128)	(0.173)	(0.128)	(0.162)
ImportRatio_i	0.013**	0.004	0.051***	0.019**	0.052***	0.021**	
	(0.006)	(0.006)	(0.010)	(0.009)	(0.010)	(0.010)	
Kintensity_x			0.011	0.028***	0.008	0.025**	
			(0.008)	(0.007)	(0.013)	(0.010)	
Kintensity_x∗i	0.000*	0.000	0.000	0.000	0.000	0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
StdAvgEdu_x			0.167	−0.543*	0.594	−0.106	
			(0.327)	(0.281)	(0.517)	(0.430)	
ln(*Distance*)_*xm*	−1.184***	−1.004***	−1.283***	−1.083***			−0.936***
	(0.351)	(0.329)	(0.375)	(0.356)			(0.082)
Contiguous_xm	0.941***	1.060***	1.000***	1.104***			0.300**
	(0.253)	(0.251)	(0.260)	(0.250)			(0.149)
Legalsystem_xm	0.087	0.107	0.035	0.049			0.139
	(0.121)	(0.128)	(0.118)	(0.125)			(0.117)
Colonial_xm	0.847***	0.907***	0.747***	0.804***			1.329***
	(0.252)	(0.273)	(0.277)	(0.298)			(0.313)
Language_xm	0.089	−0.072	0.126	−0.004			0.591***
	(0.178)	(0.197)	(0.178)	(0.198)			(0.169)
Religion_xm	0.000**	0.000**	0.000	0.000*	0.001***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GATT_WTO_xm	0.009	−0.010	0.128	−0.044	−0.041	−0.083	0.138
	(0.242)	(0.232)	(0.161)	(0.160)	(0.235)	(0.218)	(0.549)
Fixed effects	FE1	FE1	FE2	FE2	FE3	FE3	FE4
Number of observations	17,182	12,752	20,985	13,224	15,052	11,483	305,292

Notes: PPML estimators are computed using *ppml* command of STATA developed by Silva and Tenreyro. The dependent variable is the 3-year average exports (divided by [GDP of exporter x ∗ GDP of importer m]) from exporter x to importer m in industry i. Robust standard errors clustered by exporter–importer pair are reported in parentheses. Notations for fixed effects are the same as in Table 3. In (7), *RTA*, *CU*, *islands*, *landlocked* are also controlled.
* 10% significance level.
** 5% significance level.
*** 1% significance level.

Table 5. *Infrastructure match index added (PPML estimation)*

Sample	Nonservice	Manufacturing	Nonservice	Manufacturing	Nonservice	Manufacturing
Infrastructure variables	*Powerloss*	*Powerloss*	*Road*	*Road*	*Powerloss*	*Powerloss*
	(1)	(2)	(3)	(4)	(5)	(6)
MatchIndex(CV)_x∗i	0.539***	0.478***	0.627***	0.453***	0.673***	0.609***
	(0.135)	(0.154)	(0.128)	(0.145)	(0.228)	(0.178)
MatchIndex(Powerloss)_x∗i	−0.423***	−0.373***			−0.591***	−0.733***
	(0.102)	(0.126)			(0.142)	(0.184)
MatchIndex(Road)_x∗i			0.523***	0.346***		
			(0.130)	(0.122)		
Fixed effects	FE1	FE1	FE1	FE1	FE2	FE2
Number of observations	17,182	12,752	17,182	12,752	20,985	13,224
Sample	Nonservice	Manufacturing	Nonservice	Manufacturing	Nonservice	Manufacturing
Infrastructure variables	*Road*	*Road*	*Powerloss*	*Powerloss*	*Road*	*Road*
	(7)	(8)	(9)	(10)	(11)	(12)
MatchIndex(CV)_x∗i	0.540**	0.413***	0.609***	0.519***	0.479**	0.295*
	(0.218)	(0.154)	(0.236)	(0.190)	(0.221)	(0.164)
MatchIndex(Powerloss)_x∗i			0.563***	−0.682***		
			(0.150)	(0.178)		
MatchIndex(Road)_x∗i	0.640***	0.585***			0.608***	0.489***
	(0.151)	(0.165)			(0.157)	(0.167)
Fixed effects	FE2	FE2	FE3	FE3	FE3	FE3
Number of observations	20,985	13,224	15,052	11,483	15,052	11,483

Notes: Estimation is based on the China and India sample. Control variables other than infrastructure match index are the same as in Table 4. See notes for the columns (1)–(6) in Table 4.

trade barriers at time t.[17] If so, excluding zero export observations may generate biased estimates by introducing the correlation between observed and unobserved trade barriers, as Helpman *et al.* (2008) suggest. To correct for such selection bias, I utilize the Poisson pseudo-maximum-likelihood (PPML) method, which can include zero export value as the dependent variable. By comparing several estimation methods, Silva and Tenreyro (2006) propose this PPML method to deal with zeros in trade data, as well as a heteroskedasticity bias, when estimating gravity equations.[18]

Tables 4 and columns (5)–(8) of Appendix Table 8 report the PPML estimation results. In all but one specification (column (6) of Table 4), the estimated coefficient of the match index is positive, consistent with my hypothesis. Seven of them are statistically significant at a 5% or 10% level. The signs of the estimated coefficients for the conventional bilateral trade barriers are almost the same as those in Table 3. The somewhat unexpected negative signs in Table 3 now either become positive (*ImportRatio_i*, *Kintensity_x*, and *Skillintensity_i*) or insignificant (*GATT_WTO_xm*).

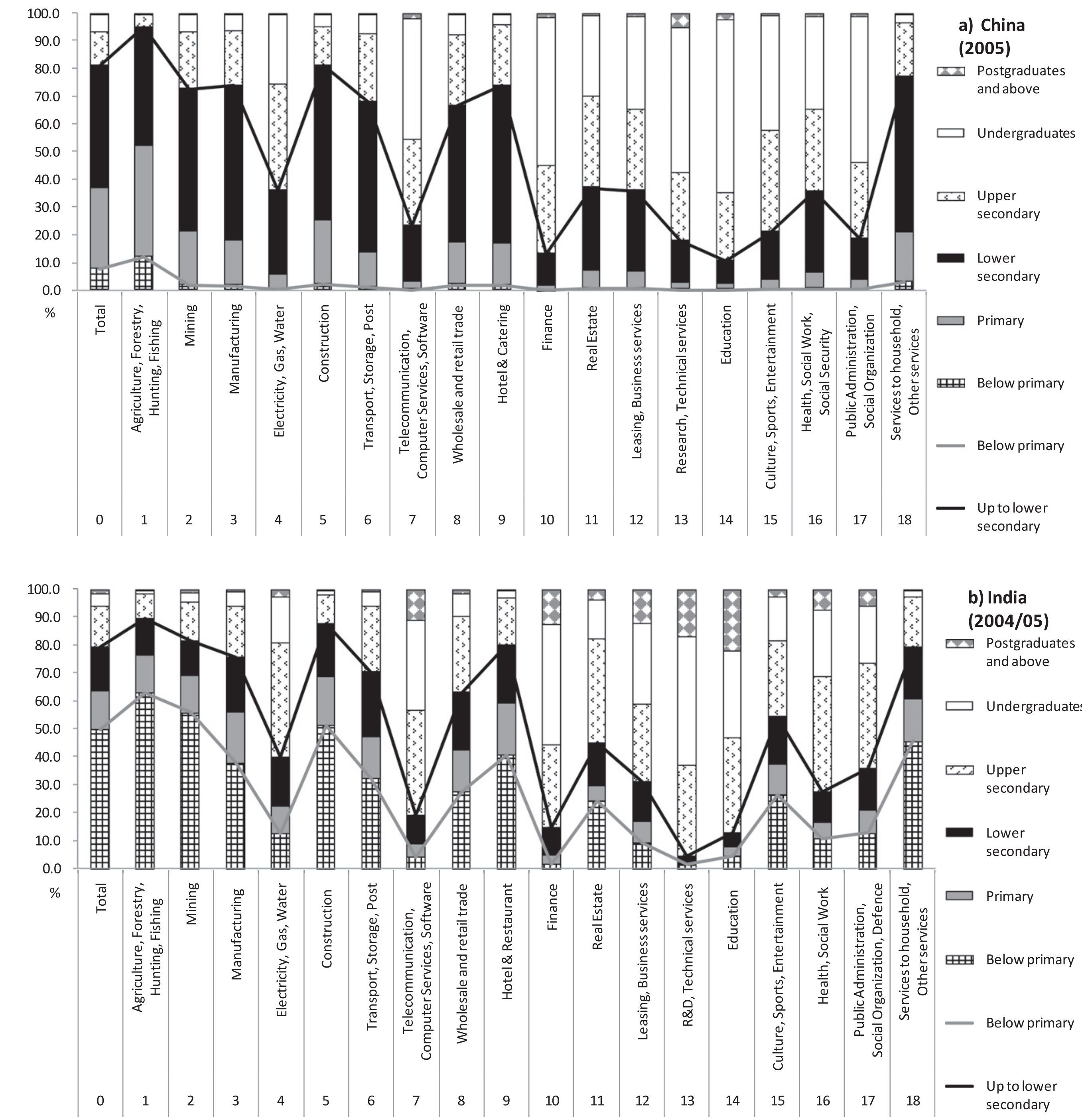

Figure 1. *Educational attainment of employed population by industry. Note and Source: The same as in endnote 1.*

(iii) *Additional match index controlled*

One of the key hypotheses of this paper is that the productivity of an industry with longer production chains is more likely to be dragged down by low-skilled workers involved across the chains. However, negative impacts resulted from low capacity and poor quality of production infrastructure, may also accumulate across the production chains. In order to control for these negative impacts of poor infrastructure, I construct two match indices: $MatchIndex(Powerloss)_{xit}$ and $MatchIndex(Road)_{xit}$. Each index is constructed similarly to the skill match index (e.g., $MatchIndex(CV)_{xit}$). Instead of using the skill dispersion index CV_{xt}, I use data on electric power transmission and distribution loss in the exporting country ($Powerloss_{xt}$, measured as a percentage of output), and $(-1)^*$ road density in China and India ($Road_{xt}$, kilometers of total road network divided by 100 km^2 of land area), respectively.

Table 5 and the columns (9)–(16) in Appendix Table 8 report the PPML estimation results having introduced either $MatchIndex(Powerloss)_{xit}$ or $MatchIndex(Road)_{xit}$ as an additional control variable. Each index is added separately, since adding both indices together creates a multicollinearity problem. Even after controlling for infrastructure, positive coefficients of $MatchIndex(CV)_{xit}$ are obtained in all specifications. They are also statistically significant, except for columns (11) and (15) in Appendix Table 8. In Table 5, the sizes of the coefficients become even larger compared to those of Table 4. The effect of $MatchIndex(Powerloss)_{xit}$ is unexpectedly negative in all specifications. This may be because our variable, $Powerloss_{xt}$ just indicates the management technology or quality of the power supply system, and is not related with the quality of power actually used in the production stage. In order to control for the quality of power used in actual production, we might need data such as the frequency of blackouts that are not restored by generator. The effect of $MatchIndex(Road)_{xit}$ is positive in all specifications as anticipated.

In summary, the series of robustness checks presented above largely confirm the positive impact of the skill match index on the size of exports.

(c) *Skill sorting hypothesis v.s. random matching hypothesis*

The results obtained so far have shown that a country with a greater [lesser] dispersion of skill has higher exports in industries with shorter [longer] production chains. Is the skill sorting mechanism, that is the first hypothesis presented in Introduction, working behind? Or does the random matching between workers and firms prevail as proposed by the second hypothesis? We can find some supporting evidence for the skill sorting hypothesis based on descriptive statistics and the India sample. First, Figure 1 indicates the presence of skill sorting across industries in both countries, under the assumption of high correlation between unobserved skill and educational attainment. Figure 1 shows the breakdown of the educational attainment by industry of the employed population of China and India in 2005, although industry classification is different from the one used in the previous regression analysis. Most service industries employ relatively more educated people than the agricultural, mining, and manufacturing industries. Figure 1 also shows that the skill level of workers in China in manufacturing is higher and more homogeneous, while the skill level of workers in India in several service industries (telecommunication and computer services, wholesale and retail trade, leasing and business services, R&D and technical services, and health and social work) is higher in terms of the proportion of workers with more than lower secondary education. This proportion is also higher in India in the agriculture, forestry, hunting, and fishing industries. As we have seen, production chains tend to be longer in manufacturing and shorter in agricultural, mining, and service industries. Hence, the skill distribution patterns by industry of China and India seem to be consistent with the skill sorting hypoth-

Table 6. *Regression results for skill sorting (India sample)*

Sample	India, all industries			India, nonservice		
	(1)	(2)	(3)	(4)	(5)	(6)
leontief	−8.303***	−3.646**	−2.908	−4.544	−3.189*	−1.726
	(3.124)	(1.486)	(2.122)	(4.171)	(1.880)	(2.746)
bpleon		−0.388***	−0.403***		−0.380***	−0.394***
		(0.029)	(0.037)		(0.036)	(0.041)
Time dummy	Yes	Yes	Yes	Yes	Yes	Yes
Sector dummy	Yes	No	Yes	Yes	No	Yes
N	96	96	96	76	76	76
R-squared	0.267	0.683	0.684	0.192	0.656	0.659
F-statistics	5.40	38.73	27.24	3.32	26.69	22.19
	India, manufacturing and service			India, manufacturing		
	(7)	(8)	(9)	(10)	(11)	
leontief	−8.200**	−3.574*	−4.108*	−0.337	−5.322	
	(3.602)	(1.910)	(2.292)	(5.722)	(3.557)	
bpleon		−0.457***	−0.458***		−0.473***	
		(0.046)	(0.046)		(0.056)	
Time dummy	Yes	Yes	Yes	Yes	Yes	
Sector dummy	Yes	No	Yes	No	No	
N	68	68	68	48	48	
R-squared	0.218	0.698	0.699	0.143	0.685	
F-statistics	3.46	28.62	23.57	1.80	18.30	

Notes: The dependent variable is the percentage of workers who have attained post-secondary education in industry *i*.
* 10% significance level.
** 5% significance level.
*** 1% significance level.

esis, where high-skilled people are encouraged more to be sorted into industries with shorter production chains in a country like India which has a greater dispersion of skill.

Second, Appendix Table 8 shows that after correcting for selection bias, most of the coefficients of industry's skill intensity (*Skillintensity_i*) become positive supporting the skill sorting hypothesis, although they are mostly insignificant. Third, we can also observe skill sorting trend by conducting simple OLS regressions based on the India sample. In Table 6, the percentage of workers who have attained post-secondary education in industry i ($postsec_{it}$) is regressed on $Leontief_{it}$ and other control variables such as the percentage of low-skilled workers involved across the chains ($bpleon_{it}$), time dummies (four periods), and broad sector dummies (primary, manufacturing, and service). $bpleon_{it}$ is defined as the percentage of low-skilled workers (i.e., illiterate workers and literate workers without formal schooling or with below primary-level education), who are involved across all the production chains of industry i. It is computed as $bpleon_{it} = \sum_k bpwin_{kt} * (leon_{kit}/Leontief_{it})$, where $bpwin_{kt}$ is the percentage of low-skilled people working within industry k, which is computed from the NSSO's data (NSSO, 1983, 1987–88, 1993–94, 1999–00). $bpleon_{it}$ is controlled because in addition to the length of production chains, the degree of involvement by low-skilled workers across the chains determines how much high-skilled workers' wages are dragged down.

Columns of Table 6 differ in sample and control variables as expressed in the table. Although some coefficients are not statistically significant, $Leontief_{it}$ and $postsec_{it}$ are negatively associated in all specifications, as expected by the skill sorting hypothesis: The proportion of the high-skilled workers in industry i declines as the industry production chains become longer. The similar results are obtained in Asuyama (2011b) which examines skill sorting across 42 Indian industries in 1999–2000.

6. CONCLUDING REMARKS

This paper empirically examines the different comparative advantages of the two emerging economic giants, China and India that result from the different skill distribution patterns in each country. By utilizing industry export data on China and India from 1983 to 2000, this paper finds that a country with a greater dispersion of skills (i.e., India, especially in the earlier years) has higher exports in industries with shorter production chains. Conversely, a country with a more equal dispersion of skills (i.e., China, especially in the later years) is found to have higher exports in industries with longer production chains. The causal relationship is fairly robust across different specifications, including those which examine manufacturing samples only, test for endogeneity, correct for selection bias, and control for infrastructure factors. Although further research is necessary, skill sorting mechanism seems to explain this causal relationship.

Although skill distributions are becoming more equal over time in both countries, China has enjoyed more equal skill distribution compared with India. Skill distribution in India is characterized by a much narrower semi-skilled labor force in the middle and much larger proportions of illiterate and skilled workers at opposite ends of the spectrum. The length of production chains tends to be longer in most manufacturing industries, while shorter in the agricultural, mining, and service industries. Although our main estimation results are only strictly applicable to nonservice industries, our results are consistent with the fact that China, a country with narrower dispersion of skills, has a comparative advantage in large-scale manufacturing industries with longer production chains, while India, a country with a greater dispersion of skills, has a comparative advantage in offshore service industries with shorter production chains.[19] This finding indicates that if India would like to foster large-scale manufacturing industries with long domestic supply chains and increase exports in these industries, it needs to increase the number of semi-skilled workers with primary or secondary education and make skill distribution more equal. As Asuyama (2011a) has examined, potential solutions may include various reforms in education and training policies, such as redesigning the financing system for education and the incentive structure for teachers and local government officials, as well as simply improving the quantity and quality of primary and secondary education.

In addition to further exploring the linkage between skill sorting and comparative advantage, fully incorporating service industries into the empirical analysis is left for future research. Another possible extension of this research is to take into account the different skill content across each production chain, in addition to the length of production chains. Finally, building a multi-country, multi-industry skill sorting model which endogenizes the length of production chains is also left for future research.

NOTES

1. Skill distribution data are taken from the SC and NBS (State Council and National Bureau of Statistics of China) (2007) and NSSO (2004–05, 61st round). In India, a person is considered working based on the usual principal activity status.

2. Similar positive assortative matching between skill of labor and quality of nonlabor inputs, which is caused by different quality of nonlabor inputs across industries, is also suggested by Sampson (2011).

3. Nineteen industries are equal to the 24 industries used in input–output table (see Section 4(b)) minus the five service industries which are not included in the trade data.

4. I also experimented with using all year export data instead of 3-year average as the dependent variable, and obtained very similar regression results.

5. Taking into account for the prevalence of child labor in India would make the skill distribution of potential Indian labor force more unequal. Thus, including child labor would further strengthen the impact of country's skill distribution on comparative advantage.

6. If we assume random matching between workers and firms and ex-ante unobservable skills as in the second hypothesized mechanism, it is better to use distribution of unobserved skills as Bombardini *et al.* (2009) did, instead of using observable skills measured by education. However, with the limited data availability, it is reasonable to assume a high correlation between education and unobserved skills, at least for a large economic sphere such as country and industry. In fact, the unobserved and observed skill distribution of a country is highly correlated with 0.857 coefficient of correlation in Bombardini *et al.* (2009) (see St Dev and St Dev Resid in their Table 1).

7. Using skill dispersion measure which predates exports by 3–5 years generated similar regression results with larger β with higher statistical significance in general.

8. For example, suppose that in order to produce one unit of output, an automobile industry directly uses 0.4 units of input from the automobile industry itself, 0.2 units from the steel industry, and 0.1 units from the computer industry (the remaining 0.3 units are value added). Consequently, the 0.4 units of input from the automobile industry further require $0.4 * 0.4$ units of input from the automobile industry, $0.4 * 0.2$ units from steel industry, and $0.4 * 0.1$ units from computer industry. Again, to produce the $0.4 * 0.4$ units of input from the automobile industry, the automobile industry requires $0.4 * 0.4 * 0.4$ inputs from the automobile industry itself ... and so on. In this way, one output generated by an industry also indirectly generates chains of demand for nonlabor inputs.

9. According to the recent studies on Chinese exports (Dean, Fung, & Wang, 2011; Koopman, Wang, & Wei, 2008), my *Leontief* is likely be overestimated (and industry's intensity of imported input usage is likely to be underestimated) especially for China. This problem occurs because my *Leontief* assumes that industry's intensity of imported input usage is the same, regardless of whether industry output is for processing exports which utilize substantial amount of imported inputs, or for normal exports or domestic consumption. In my empirical study, this overestimation problem is largely controlled by exporter-industry fixed effects or exporter-time fixed effects. In addition, this problem is likely to occur especially in later periods when processing exports account for around half of the total Chinese exports. Thus, I experimented with restricting the sample to the first period when the share of processing exports was still low (around 10%) in China. The PPML estimators of β, which control for selection bias with several fixed effects specifications are still positive and statistically significant for manufacturing sample, although they are statistically insignificant and slightly negative for nonservice sample. These results are available from the author upon request.

10. Since the original IO tables of India do not distinguish domestic and imported inputs, the transaction tables of domestic inputs are computed by using original import flow matrix, use table, and make table. The import flow matrices of 1983–84 and 1998–99 are estimated from those of 1989–90 and 1993–94, respectively, since they are not available. It is assumed that the share of imported input for each column i in the total import is unchanged from 1983 to 1989, and from 1993 to 1998.

11. The original 115 industries in India's IO table are consolidated into the 24 industries by using the concordance table of Saluja and Yadav (2009) as a reference.

12. As mentioned in endnote 9, our Leontief might be overestimated for China. However, even comparing Leontief in the first period when the share of processing exports in China was still low, Leontief is longer in China than in India.

13. If *MID* is used as the skill dispersion index, multiplication of negative one is not necessary.

14. For example, suppose that the standardized *CV* of skill distribution of countries A and B at time t are -2 and 2, respectively, and the standardized *Leontief* of industries X and Y are -1.5 and 2.0 in country A and -2.0 and 1.5 in country B. So country A's skill distribution is more equal, and industry Y has longer production chains in both countries. Then, $MatchIndex(CV)$ for the (country, industry) pair is -3.0 $(=(-2.0 * -1.5) * (-1))$ for (A, X); 4.0 $(=(-2.0 * 2.0) * (-1))$ for (A, Y); 4.0 $(=(2.0 * -2.0) * (-1))$ for (B, X); and -3.0 $(=(2.0 * 1.5) * (-1))$ for (B, Y). $MatchIndex(CV)$ becomes larger in both (A, Y) and (B, X) combinations, i.e., lower [higher] skill dispersion matches with industries with longer [shorter] production chains.

15. The 1997 IO table of the US is released by the Bureau of the Economic Analysis. Both domestic and imported inputs are included when computing *Leontief*, since import structure is likely to differ substantially across countries. Skill dispersion index is computed from Barro and Lee (2000) for each country allocating 0, 6, 12, and 16 years of education for each of the four schooling levels.

16. As mentioned previously, the use of skill distribution measures that are 8–10 years before the data on exports minimizes the possibility of reverse causality. In addition, endogeneity problem is largely controlled by various fixed effects.

17. Trade values also become zero simply because the NBER-UN Trade Data does not contain trade flows which do not exceed $100,000 per year at the 4-digit SITC Rev.2 level (Feenstra *et al., 2005*).

18. I also employed a correction procedure for the panel data developed by Wooldridge (1995), in particular, his procedures 3.1 and 4.1.1, which use the Tobit-form selection equation. As the excluded variables, I use either *Religion_xm* or *Language_xm*, both of which are identified by Helpman *et al.* (2008) to satisfy the exclusion restrictions in gravity equations. The regression results of the first-stage indicate the presence of selection bias. The estimated coefficients of the match index after correcting for selection bias in the second-stage regression are positive and statistically significant in all specifications. Furthermore, the sizes of those coefficients are mostly similar to those obtained in Table 3. However, since the results also indicate that at least one of our instruments is not likely to satisfy the exclusion restrictions, the PPML estimation seems more appropriate to correct for selection bias. The regression results mentioned in this endnote are available from the author upon request.

19. In particular, we can easily extend our logic to software industry. As mentioned in Introduction, India exports more from software industry, that has shorter production chains: the *Leontief* of the computer and related activities which include software industry ranks 124th out of 130 industries in India in 2003, although this *Leontief* is computed from all inputs including domestic and imported inputs (India IO table, 2003–04).

REFERENCES

Anderson, J. E., & Wincoop, E. V. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review, 93*(1), 170–192.

Asuyama, Y. (2011a). Skill formation through education and training: A comparison of China and India. In M. Ohara, M. Vijayabaskar, & H. Lin (Eds.), *Industrial dynamics in China and India: Firms, clusters, and different growth paths* (pp. 107–134). Basingstoke, Hampshire: Palgrave Macmillan.

Asuyama, Y. (2011b). Skill sorting, inter-industry skill wage premium, and production chains: Evidence from India 1999–2000. IDE Discussion paper, 278.

Barro, R. J., & Lee, J. W. (2000). International data on educational attainment: Updates and implications. CID Working Paper, 42.

Barro, R. J., & McCleary, R. M. (2005). Which countries have state religions? *Quarterly Journal of Economics, 120*(4), 1331–1370.

Bombardini, M., Gallipoli, G., & Pupato, G. (2009). Skill dispersion and trade flows. NBER Working Paper, 15097.

CEPII (Centre D'Etudes Prospectives et D'Informations Internationales). *Distances measures*. Retrieved from: <http://www.cepii.fr/anglaisgraph/bdd/distances.htm/>.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics, 82*(2), 152–167.

CSO (Central Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India). *Input–output transactions table 1983–84, 1989–90, 1993–94, 1998–99, 2003–04*.

CSO. *National accounts statistics*. Retrieved from: <http://mospi.nic.in/mospi_cso_rept_pubn.htm/>.

Cuñat, A., & Melitz, M. J. (2007). Volatility, labor market flexibility, and the pattern of comparative advantage. NBER Working Paper, 13062.

DCS-NBS (Department of Comprehensive Statistics of National Bureau of Statistics) (2005). *China compendium of statistics 1999–2004*. Beijing: China Statistics Press.

Dean, J. M., Fung, K. C., & Wang, Z. (2011). Measuring vertical specialization: The case of China. *Review of International Economics, 19*(4), 609–625.

EPW Research Foundation (2002). *Annual survey of industries 1973–74 to 1997–98: A data base on the industrial sector in India*. Mumbai: EPW Research Foundation.

Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., & Mo, H. (2005). World trade flows: 1962–2000. NBER Working Paper, 11040.

Ghosh, B., & De, P. (2005). *India infrastructure database 2005 (Vol. II)*. New Delhi: Bookwell.

Gregory, N., Nollen, S., & Tenev, S. (2009). *New industries from new places: The emergence of the software and hardware industries in China and India*. Stanford, California/Washington, DC: Stanford University Press/World Bank.

Grossman, G. M. (2004). The distribution of talent and the pattern and consequences of international trade. *Journal of Political Economy, 112*(1), 209–239.

Grossman, G. M., & Maggi, G. (2000). Diversity and trade. *American Economic Review, 90*(5), 1255–1275.

Head, K., Mayer, T., & Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics, 81*(1), 1–14.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics, 123*(2), 441–487.

Holz, C. A. (2006). New capital estimates for China. *China Economic Review, 17*(2), 142–185.

IDE-JETRO (Institute of Developing Economies, Japan External Trade Organization). *Asian international I/O table, 1985, 1990, 1995, 2000*.

IMF (International Monetary Fund). 2010. *World economic outlook database, October 2010*. Retrieved from: <http://www.imf.org/external/pubs/ft/weo/2010/02/weodata/index.aspx/>.

Koopman, R., Wang, Z., & Wei, S. J. (2008). How much of Chinese exports is really made in China? Assessing domestic value-added when processing trade is pervasive. NBER Working Paper, 14109.

Kremer, M. (1993). The O-ring theory of economic development. *Quarterly Journal of Economics, 108*(3), 551–575.

Levchenko, A. A. (2007). Institutional quality and international trade. *Review of Economic Studies, 74*(3), 791–819.

Lo, C. P., & Liu, B. J. (2009). Why India is mainly engaged in offshore service activities, while China is disproportionately engaged in manufacturing? *China Economic Review, 20*(2), 236–245.

Mayer, T., & Zignago, S. (2006). *Notes on CEPII's distances measures*. Retrieved from: <http://www.cepii.fr/anglaisgraph/bdd/distances.htm/>.

MHRD (Ministry of Human Resource Development). (2008). *Selected educational statistics: 2005–06, all India, time-series*. Retrieved from: <http://www.educationforallinindia.com/SESall-india-time-series-2005-06.pdf/>.

MHRD and NIC (Ministry of Human Resource Development, Department of Education, and National Informatics Centre). (n.d.). *Compilation on 50 years of Indian Education: 1947–1997*. Retrieved from: <http://education.nic.in/cd50years/home.htm/>.

NBS (National Bureau of Statistics of China) (2009). *China statistical yearbook 2009*. Beijing: China Statistics Press.

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, 1983 (38th round), 1987–88 (43rd round), 1993–94 (50th round), 1999–00 (55th round), 2004–05 (61st round)*.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *Quarterly Journal of Economics, 122*(2), 569–600.

Ohara, M., & Lin, H. (2011). Competition and management in the manufacturing sector in China and India: A statistical overview. In M. Ohara, M. Vijayabaskar, & H. Lin (Eds.), *Industrial dynamics in China and India: Firms, clusters, and different growth paths* (pp. 19–39). Basingstoke, Hampshire: Palgrave Macmillan.

Ohnsorge, F., & Trefler, D. (2007). Sorting it out: International trade with heterogeneous workers. *Journal of Political Economy, 115*(5), 868–892.

Saluja, M. R., & Yadav, B. (2009). Construction of input–output table for Indian economy-2005. In H. Kuwamori, Y. Uchida, & S. Inomata (Eds.), *Compilation and use of the 2005 international input–output tables*. Chiba: IDE-JETRO.

Sampson, T. (2011). *Assignment reversals: Trade, skill allocation and wage inequality*. Retrieved from: <http://www.people.fas.harvard.edu/~tsampson/papers.html/>.

SC and NBS (State Council and National Bureau of Statistics of China) (2007). *2005 national 1% population sample survey*. Beijing: China Statistics Press.

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

Tang, H. (2010). Labor market institutions, firm-specific skills, and trade patterns. Centro Studi Luca D'Agliano Development Studies Working Paper, 301.

UN (United Nations). (n.d.). *National accounts main aggregates database*. Retrieved from: <http://unstats.un.org/unsd/snaama/Introduction.asp/>.

UN (1993). *Statistical yearbook 1990/91*. New York: United Nations.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics, 68*(1), 115–132.

Wooldridge, J. M. (2006). *Introductory econometrics: A modern approach* (3rd ed.). Mason, Ohio: Thomson/South-Western.

World Bank. (2010). *World development indicators & global development finance*. Retrieved from: <http://data.worldbank.org/data-catalog/world-development-indicators/>.

WTO (World Trade Organization). (2010a). *International trade statistics 2010*. Retrieved from: <http://www.wto.org/english/res_e/statis_e/its2010_e/its10_toc_e.htm/>.

WTO. (2010b). *WTO trade profiles*. Retrieved from: <http://stat.wto.org/Home/WSDBHome.aspx?Language=E/>.

APPENDIX

See Tables 7 and 8.

Table 7. *Description of variables and sources*

Variable	Sources/construction method
Export_xmi	See Section 4(a). The GDP data of China and India are taken from DCS-NBS (Department of Comprehensive Statistics of National Bureau of Statistics) (2005), NBS (National Bureau of Statistics of China) (2009), and National Account Statistics published by CSO. GDPs of other countries and exchange rates are taken from World Bank (2010), supplemented with the UN (n.d.), UN (1993), and IMF (International Monetary Fund) (2010)
CV_{xt}	See Section 4(b)
$Leontief_{xit}$	See Section 4(b)
MatchIndex(CV)_x∗i	See Section 4(b)
MatchIndex(IV)_x∗i	See Section 5(b)(i). School enrollment data are taken from DCS-NBS (2005), MHRD & NIC (n.d.), and MHRD (Ministry of Human Resource Development) (2008). Population data are from World Bank (2010)
ImportRatio_i	Percentage of the imported inputs to total input value in each industry Source: The same as $Leontief_{xit}$
Kintensity_x	Capital intensity of exporting country (hereafter, exporter) = capital stock/GDP ∗ 100(%). Both capital stock and GDP are real 2000 prices and 3-year average (1983–85, 1988–90, 1993–95, 1998–2000) Source: Capital stock data are from: Holz (2006), p. 170, Table 6, BC3, for China, and National Account Statistics published by CSO, for India. GDP is from World Bank (2010)
Kintensity_x∗i	Interaction of *Kintensity_x* and *Kintensity_i*, which is capital intensity of industry *i* (=capital stock/gross value added (GVA) ∗ 100(%)). Both capital stock and GVA are real 2000 prices *Kintensity_i* is estimated as follows: China: Capital stock is estimated as [Depreciation/0.05] assuming 5% depreciation rate. Depreciation and GVA data are from IO table. Implicit deflators are computed from Holz (2006), p. 178, Table 8 (columns (2)/(4)) and China's GDP data India: For manufacturing industries, assuming the same capital intensity between registered and unregistered sectors, fixed asset and GVA (or net value added plus depreciation) data from the Annual Survey of Industries (ASI) (EPW Research Foundation, 2002; ASI website of Ministry of Statistics and Programme Implementation), which cover only registered manufacturing are used. Implicit deflators are computed from National Account Statistics. For the nonmanufacturing industries, net fixed capital stock (NFCS) and GDP data from National Account Statistics are used. *Kintensity_i* of agriculture is applied for three industries (Nos. 1, 2, and 3), and that of mining and quarrying is applied for two industries (Nos. 6 and 7), due to broad industry classification of NFCS and GDP data
StdAvgEdu_x	Standardized average years of education of the population over age 15 in exporting country Source: Barro and Lee (2000)
*Skillintensity*_i (only for India)	Skill intensity of industry = percentage of working population who have completed secondary or post-secondary education. A person is considered working based on the usual principal activity status Source: NSSO, 1983, 1987–88, 1993–94, 1999–00
Skillintensity_x∗i (only for India)	Interaction of *Skillintensity_i* and skill intensity of exporter (=percentage of population over age 15 who have completed secondary education or have attained some post-secondary education). This ratio is calculated from Barro and Lee (2000)
ln(Distance)_xm	Logarithm of the distance between the capital cities of exporter and importing country (hereafter, importer) Source: CEPII's distance data (see Mayer & Zignago, 2006)
Contiguous_xm	Dummy = 1 if exporter and importer are contiguous Source: CEPII's distance data (see Mayer & Zignago, 2006)
Legalsystem_xm	Dummy = 1 if exporter and importer share the same legal origin Source: Helpman *et al.* (2008)
Colonial_xm	Dummy = 1 if importer ever colonized exporter or vice versa Source: Helpman *et al.* (2008)
Language_xm	Dummy = 1 if a language is spoken by at least 9% of the population in both exporter and importer Source: CEPII's distance data (see Mayer & Zignago, 2006)
Religion_xm	Time-variant index which indicates the degree of shared religion between exporter and importer. It is constructed as follows by applying the method of Helpman *et al.* (2008) *Religion_xm* = $\sum_k$(% *religion_k* in exporter ∗ % *religion_k* in importer), where % *religion_k* indicates percentage of population who are adherent to religion *k*. There are nine religions (Catholic, Protestant, other Christian, Orthodox, Muslim, Hindu, Buddhist, Other Eastern religions, and Jewish). Since only 1970 and 2000 data are available, the religion indices for four periods (1984, 1989, 1994, and 1999) are estimated by assuming the constant growth rate of the index from 1970 to 2000 Source: Barro and McCleary (2005)
GATT_WTO_xm	Number of exporter/importer who are members of GATT or WTO Source: Helpman *et al.* (2008) and WTO website
MatchIndex(Powerloss)_x∗i	For the construction of index, see Section 5(b)(iii) Source: World Bank (2010)
MatchIndex(Road)_x∗i	For construction of the index, see Section 5(b)(iii) Source: World Bank (2010), and Ghosh and De (2005), p. 754, Table 6.11

Table 8. *Regression results obtained when using the India sample only*

Sample	A. Baseline fixed effects regression				B. PPML estimation			
	Nonservice (1)	Manufacturing (2)	Nonservice (3)	Manufacturing (4)	Nonservice (5)	Manufacturing (6)	Nonservice (7)	Manufacturing (8)
MatchIndex(CV)_x∗i	0.217**	0.333***	0.385***	0.461***	0.320**	0.244	0.318**	0.251
	(0.091)	(0.118)	(0.081)	(0.114)	(0.145)	(0.173)	(0.139)	(0.171)
Skillintensity_i	−0.028***	−0.010	−0.044***	−0.036***	0.004	0.016	0.012	0.026***
	(0.011)	(0.012)	(0.011)	(0.013)	(0.011)	(0.013)	(0.011)	(0.010)
Skillintensity_x∗i	0.004***	0.002	0.005***	0.004***	0.000	−0.001	−0.001	−0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Fixed effects	FE1′	FE1′	FE2′	FE2′	FE1′	FE1′	FE2′	FE2′
Number of observations	5110	3886	5110	3886	7068	5255	10,146	6348

	C. PPML estimation with infrastructure controlled							
Sample	Nonservice	Manufacturing	Nonservice	Manufacturing	Nonservice	Manufacturing	Nonservice	Manufacturing
Infrastructure variables	*Powerloss*	*Powerloss*	*Road*	*Road*	*Powerloss*	*Powerloss*	*Road*	*Road*
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
MatchIndex(CV)_x∗i	0.245*	0.321*	0.176	0.310*	0.248*	0.366**	0.164	0.353**
	(0.133)	(0.185)	(0.140)	(0.176)	(0.128)	(0.183)	(0.135)	(0.174)
MatchIndex(Powerloss)_x∗i	−0.533***	−0.288*			−0.573***	−0.421**		
	(0.170)	(0.175)			(0.156)	(0.200)		
MatchIndex(Road)_x∗i			0.562***	0.287*			0.594***	0.440**
			(0.108)	(0.174)			(0.100)	(0.200)
Skillintensity_i	−0.005	0.010	−0.004	0.009	0.002	0.017	0.003	0.015
	(0.011)	(0.014)	(0.010)	(0.014)	(0.011)	(0.011)	(0.010)	(0.011)
Skillintensity_x∗i	0.000	−0.001	−0.001	−0.001	−0.001	−0.002	−0.002	−0.002
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Fixed effects	FE1′	FE1′	FE1′	FE1′	FE2′	FE2′	FE2′	FE2′
Number of observations	7068	5255	7068	5255	10,146	6348	10,146	6348

Notes: FE1′ indicates time fixed effects (FE) and importer-industry FE. FE2′ indicates industry FE and importer-time FE. *ImportRatio_i*, *Kintensity_x∗i*. *Religion_xm* (only with FE1′), and *GATT_WTO_xm* (only with FE1′) are also controlled. See notes of Table 3 for panel A, those of Table 4 for panel B, and those of Table 5 for panel C, respectively.