INSTITUTE OF DEVELOPING ECONOMIES

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

IDE DISCUSSION PAPER No. 277

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

Yoko ASUYAMA*

February 2011

Abstract
This paper empirically examines the different comparative advantages of two emerging economic giants, China and India, in relation to the different skill distribution patterns in each country. 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, whereas 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. This empirical evidence supports our assumption that the likely mechanism for these results is the negative impact of low-skilled workers on input quality, which accumulates and becomes larger as the length of production chains and the proportion of low-skilled workers in the economy increase.

Keywords: China, Comparative advantage, India, Production chains, Sequential production, Skill distribution

JEL classification: F14, F16

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

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

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

Skill Distribution and Comparative Advantage: A Comparison of China and India[*]

Yoko Asuyama[†]

February 2011

Abstract

This paper empirically examines the different comparative advantages of two emerging economic giants, China and India, in relation to the different skill distribution patterns in each country. 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, whereas 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. This empirical evidence supports our assumption that the likely mechanism for these results is the negative impact of low-skilled workers on input quality, which accumulates and becomes larger as the length of production chains and the proportion of low-skilled workers in the economy increase.

Keywords: China, Comparative advantage, India, Production chains, Sequential production, Skill distribution

JEL codes: F14, F16

[*] I would like to thank Satoshi Inomata and Hiroshi Kuwamori for providing me with the "Asian International I/O Table" and advising me on the input-output tables of China and India. All remaining errors are my own.

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

1. Introduction

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.

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. 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 has more exports in industries with longer production chains, whereas a country with a relatively unequal skill distribution has more exports in industry with shorter production chains. The fact that China has a more equal distribution of skills compared with India indicates that China [India] has a comparative advantage in industries with longer [shorter] production chains. Production chains generally tend to be longer in manufacturing industries compared with agricultural, mining, and service industries. Consequently, the empirical results of this paper 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 key mechanism assumed to explain the above-mentioned causal relationship is that industry productivity is more likely to be dragged down by low-skilled workers in the economy as the length of industry production chains increases. Furthermore, this negative impact on productivity becomes larger in a country with greater dispersion of skills. Therefore, it follows that a country with greater [less] dispersion of skills has a comparative advantage in industries with shorter [longer] production chains.

This paper contributes to the literature in three 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, no existing studies have provided such empirical evidence. Second, in this paper a new index is constructed to show the degree of match between country skill distribution and the length of industry production chains, which contributes to the empirical literature exploring the sources of comparative advantage. When constructing the match index, I assume that the technology and structure of an industry are different across countries. Although existing studies assume the same industrial characteristics across countries, this study takes into account variations in industry characteristics by country. The match index created in this study also has the advantage over previous studies in that it simultaneously identifies comparative advantages in industries with both longer and shorter production chains. Finally, 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.

The rest of the paper is organized as follows. Section 2 provides a theoretical and empirical background to the connections between skill distributions in China and India and the patterns of comparative advantage and economic development in both countries. Section 3 explains the empirical methodology of the analysis in this study. Section 4 describes the data and explains the construction of the key variables. Section 5 presents the estimation results, including several analyses of robustness. Section 6 offers my conclusions.

2. Theoretical and Empirical Background

2.1 Sources of comparative advantage

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. Grossman and Maggi (2000) classify industries into two types. The first type of industry (e.g., large-scale manufacturing) is characterized by complementarity of tasks, referred to as supermodular production technology. In this type of industry, output is increasingly hindered by low-skilled workers as the degree of complementarity becomes higher. The second type of industry (e.g., software or financial services) is characterized by substitutability of tasks, referred to as submodular technology. A typical task in this type of industry requires creativity or problem solving; output is fully determined by the superior performance of the most talented worker in the most extreme case. Their model shows that a country with a relatively homogenous population exports the goods of the first type of industry, whereas a country with a more diverse workforce exports the good of the second type of industry.

Bombardini et al. (2009) extend their two-county, two-sector model to a multi-country, multi-sector model by considering only supermodular production functions, as most production functions are supermodular in the existing trade literature. Industries differ only in terms of degree of complementarity among workers' skills. Under the assumption that workers and firms are randomly matched as a result of labor market friction and unobservability of workers' skills, their model indicates that a country with a greater dispersion of human capital has more exports in sectors with a lower degree of complementarity among workers' skills.

Grossman (2004) constructs a different model showing that when an individual's contribution to firm output can be measured perfectly in one industry (e.g., software) but not in the other industry (e.g., team production such as the automobile industry) because of imperfect labor contracts, a country with the more heterogenous workforce exports the goods of the former industry, which are produced by the most talented individuals. This result occurs because in a country with a greater spread of talents, high-skilled individuals are more discouraged from entering an industry with team production where industry productivity, and thus the wages offered, are dragged down by the larger number of low-skilled workers. Therefore, high-skilled workers are sorted into the other industry where individuals are paid according to their own productivity.

Lastly, Tang (2010) and Ohnsorge and Trefler (2007) also develop models that link the skill distribution of a country to comparative advantage, although both of these studies classify skills according to their attributes, not by the level of skill.[1] Of the five studies mentioned above, only Bombardini et al. (2009) and Tang (2010) empirically examine the linkage between country skill distribution and comparative advantage.[2]

The method of industry characterization in this paper is somewhat similar to that of Bombardini et al. (2009) and Grossman (2004). Despite the different settings, both papers essentially assume that industries are characterized by the extent to which productivity is dragged down by lower-skilled workers, or in short, the degree of

[1] Tang (2010) builds a model in which a country with more protective labor laws encourages workers to acquire firm-specific skills, resulting in the county having relatively more exports in sectors which intensively utilize firm-specific skills. The model developed by Ohnsorge and Trefler (2007) assumes that countries vary in how different types of skills are bundled together in workers. They show that when considering two-dimensional skills (e.g., quantitative and communication skills), a country with a more unequal skill distribution has exports of goods that intensively use either skill, whereas a country with more equal skill distribution has exports of good using both skills.

[2] Although the method of industry characterization and empirical strategy of this paper is close to that of Bombardini et al. (2009), the sample used in their empirical analysis contains neither China nor India.

complementarity among the (unobserved) skills of workers. Under this assumption, a country with a relatively equal skill distribution has a comparative advantage in industries characterized by higher skill complementarity, whereas a country with unequal skill distribution has higher exports in industries characterized by lower skill complementarity. Here, unequal skill distribution is defined as a mean-preserving spread, in which a skill distribution is considered less equal if the minimum skill level becomes lower and the maximum skill level becomes higher with average skill level unchanged.

In this paper, I characterize industries by the length of production chains, which indicates the degree to which quality-adjusted industry productivity is negatively influenced by low-skilled workers. This type of industry characterization in this study is similar to Bombardini et al. (2009) and Grossman (2004). However, the mechanism by which the above industry characteristics are generated is different from that of those two papers. In Bombardini et al. (2009), the degree of complementarity among workers' skills varies across industries because of exogenous production technologies which are different across industries. In Grossman (2004), the degree of complementarity among workers' unobserved skills and contributions differs by industry because of the varying degrees of imperfections in labor contracts across industries. In contrast, in the present paper, industries differ in the length of production chains. Longer production chains indicate that more units of non-labor inputs are used. The quality of those inputs is affected by the skill of workers involved at each production stage. Then, the negative effects of low-skilled workers on input quality become larger as the length of production chains increases, because the involvement of low-skilled workers increases. This idea of sequential production and defect accumulation is also very similar to the concept of the O-ring production function developed by Kremer (1993).[3]

[3] Costinot (2009) also characterizes each good by its complexity, defined as the number of production tasks, and develops a model that links average human capital and institutional quality with the country's comparative advantage. However, both Kremer (1993) and Costinot (2009) only theoretically examine the relationship between average skill level and a country's comparative advantage and do not examine the impact of skill distribution.

Similarly to Grossman (2004), high-skilled workers in a country with a more unequal skill distribution may have more incentives to be sorted into industries with shorter production chains, where the productivity of the industry and the wages of high-skilled workers are less affected by low-skilled workers. Hence, a country with a greater [lesser] dispersion of skill has exports in industries with shorter [longer] production chains. This result occurs because, as a result of skill sorting, the average skill level of workers, and thus average productivity, become higher in industries in which each country has a comparative advantage, compared with the industry counterparts in the other country. Alternatively, suppose that workers and firms are randomly matched as in Bombardini et al. (2009) and that skill distributions are therefore identical across industries (i.e., skill sorting does not occur). If we assume an O-ring type production function so that defects of inputs accumulate as the production chains become longer, the length of production chains plays a similar role to the degree of skill complementarity as an industry parameter in Bombardini et al. (2009). Consequently, the negative impact on industrial productivity caused by an increase in the length of production chains is larger in a country with a greater range of skill. Thus, we can reach a similar conclusion as in their paper or the same result mentioned above: a country with a greater dispersion of skill has exports in industries with shorter production chains and a country with a narrower dispersion of skill has exports in industries with longer production chains.

As mentioned in Section 2.3, assuming the identical skill distribution across industries seems unrealistic, although each industry inherits the skill distribution of the country to some degree. In reality, both the forces of skill sorting and the inheritance of the overall country skill distribution seem to have an effect. Thus, the main focus of the current paper is on empirically identifying whether a country with higher skill dispersion has a comparative advantage in industries with shorter production chains (and vice versa) without exploring deeper the mechanism behind this.

2.2 Development patterns of China and India

China seems to have followed the conventional development pattern: As income rises, the share of agriculture declines in the economy and the share of manufacturing increases; and with further development, the share of services increases (Syrquin 1989). As shown in Figure 1, the share of agriculture as a percent of GDP has declined in China with the expansion of manufacturing and later of services. As of 2005, manufacturing accounts for 33% of GDP in China. By contrast, in India, manufacturing is relatively small, accounting for only 15% of GDP. In contrast, the share of service in India is 53% of GDP, which is much larger than the 41% share of GDP in China. China has developed as a global manufacturing hub and provides a huge amount of manufactured goods to the world through exports. According to the trade statistics of World Trade Organization (WTO), the export value of manufactured goods in China was 1,125 billion USD in 2009, which accounted for 94% of the total merchandise exports of China. Manufacturing exports in India were 107 billion USD, less than a tenth of manufacturing exports in China, and accounted for only 66% of total merchandise exports in India (WTO 2010b). In contrast, India has grown as a world offshore service center. In 2008, computer and information service exports in India amounted to 36 billion USD, 5.8 times larger than in China (WTO 2010a).

To my knowledge, only a limited number of academic studies have explained why China and India differ in their comparative advantages.[4] In particular, no empirical study exists which has identified the sources of both countries' comparative advantages and their causal relationship. Gregory et al. (2009) focus on software and hardware industries and examine the reasons why India is internationally successful in the former, while China is successful in the latter. Their analytical framework is comprehensive and they compare many possible reasons, such as differences in production inputs (including

[4] Gregory et al. (2009) and Lo and Liu (2009) also support my view (Gregory et al., 2009, p.10; Lo and Liu, 2009, p.237).

worker skill), management processes, and business environment. Lo and Liu (2009) construct a model to explain why China has disproportionately attracted foreign offshore manufacturing activities, while India has mainly attracted offshore service activities. They argue that the differences in industry-specific technology capabilities have caused the observed comparative advantages for each country: China has greater technological capabilities in manufacturing due to the relative abundance of human capital, while the relative abundance of IT software professionals and English-speaking population has enhanced the technological capabilities of India for the software industry. Although it is not a comparative study of China and India, Kochhar et al. (2006) examine the unique development pattern of India, including the underdevelopment of labor-intensive manufacturing industries. Their study is also comprehensive and they argue that the legacy of several policy combinations enacted since Independence in 1947, including the emphasis on higher education compared to primary education, have affected the development pattern of India. Although all three studies (Lo and Liu, 2009, in particular) regard the level and availability of worker skill in a country as important factors that influence the pattern of development, none of these studies have empirically examined causality on comparative advantage from skill level and availability. Furthermore, these studies link skill and comparative advantage in a way that is rather simple and direct. This approach is different from mine, which assumes that comparative advantage is generated through the interactions between country skill distribution and the varying length of industry production chains.

2.3 Skill distribution in China and India

Figure 2 shows the change in educational attainment of the employed population in China and India from the early 1980s to 2005. In both countries, the educational level of the workforce improved over time. However, it is notable that China enjoys 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 proportion of semi-skilled workers is smaller in India. 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. Even as of 2005, the share of employed people who were illiterate or have only received education below the primary level was 50% in India, while it was only 8% in China. The share of workers receiving primary and lower secondary education was only 30% in India, compared to 73% in China. In contrast, the share of upper secondary and post-secondary education was 21% in India, which is 2% higher than in China (19%). Furthermore, the proportion of high-skilled workers who have attained postgraduate (or above) education was 1.5% in India compared with 0.2% in China.

Why have these different skill distribution patterns emerged between China and India? Asuyama (2010) has raised several factors by comprehensively examining the skill formation systems of both countries. That study highlights how different skill distribution patterns are generated as a result of various differences between the two countries in terms of (1) government policies on education and training (historical development, incentive structures, financing), (2) individual incentives for skill acquisition, and (3) firms' demand for skills. Based on this finding, I proceed by linking skill distribution and comparative advantage. It should be noted that in the present paper, skill distribution is treated as exogenous. The possibility of endogeneity of skill distribution is tested in the robustness analysis.

Figure 3 shows the breakdown of the educational attainment by industry of the employed population of China and India in 2005. The overall skill distribution patterns of China and India are mostly inherited to the industry skill distribution in each country. At the same time, Figure 3 also indicates differences in skill intensities, or the presence of skill sorting across industries in both countries. Most service industries employ relatively more educated people than the agricultural, mining, and manufacturing industries. Figure 3 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. Furthermore, as will be reported in Section 4.2.2, 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 model described in Section 2.1, 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.

3. Empirical Strategy

I mostly follow the estimation strategy of Bombardini et al. (2009). Their estimation equation modified the gravity equation, which aims to explain the size of bilateral trade flows (e.g., total exports from country x to county m) by various trade barriers as examined in Helpman et al. (2008). Specifically, Bombardini et al. (2009) split total country exports into exports by industry, and add an interaction term of the skill dispersion of the exporting country and the technological characteristics of the industry as shown by the degree of skill complementarity. 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.[5] In this paper, I estimate the following equation:

[5] Instead of using export volume of country x to country m in industry i as the dependent variable, Levchenko (2007) uses normalized country x's share in U.S. import in industry i; Nunn (2007) and Cuñat and Melitz (2007) use export volume of country x to the world in industry i.

$$(1) \quad \ln(Export_{xmit}) = \beta MatchIndex_{xit} + X_1\gamma + \alpha_{xt} + \alpha_{mi} + \varepsilon_{xmit}, \text{ or}$$

$$(2) \quad \ln(Export_{xmit}) = \beta' MatchIndex_{xit} + X_2\gamma' + \alpha_t + \alpha_{xmi} + \varepsilon'_{xmit},$$

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*);[6] $MatchIndex_{xit}$ is a measure to indicate how well the skill dispersion of exporter *x* matches with the characteristics of the domestic industry *i* (i.e., length of the production chains) at period *t*; α_{xt} and α_{mi} in equation (1) denote exporter-time fixed effects and importer-industry fixed effects, and α_t and α_{xmi} in equation (2) are time fixed effects and exporter-importer-industry fixed effects; X_1 in equation (1) denotes other control variables including conventional trade barriers between exporter and importer such as distance, and endowment characteristics of the exporter and its industry (e.g., capital [or skill] intensity of the exporter and its industry and their interaction term, and the ratio of imported input); X_2 in equation (2) is similar to X_1, but excludes time-invariant trade barriers between exporter and importer because they are absorbed by α_{xmi}. Equation (1) is useful to check whether conventional trade barriers explain the export patterns of China and India. Equation (2) is superior to (1) in the sense that it controls for all the unobserved effects between exporter and importer by exporter-importer fixed effects. However, equation (2) may not control all time-variant characteristics of exporters, which are controlled for by exporter-time fixed effects in equation (1). Thus, we estimate both equations.

Our main focus is on the coefficient β (or β'), which shows whether $MatchIndex_{xit}$ explains the patterns of industry export flows of China and India. As we have discussed above, we expect that a county with higher skill dispersion has relatively more exports in industries with shorter production chains, and conversely that a country with a more equal skill distribution has relatively more exports in industries with longer

[6] Log transformation of export values omit samples with zero export. In order to correct for selection bias due to such omission, the value of exports including zero trade values is used as the dependent variable in the Poisson regressions in Section 5.2.2.

production chains. In other words, higher skill dispersion of a country matches better with the development of industries with shorter production chains, and vice versa. As will be explained in detail in the next section, $MatchIndex_{xit}$ measures this degree of match. The coefficient β (or β') indicates whether exports become relatively larger in industries with a better match, and thus whether the differences in skill dispersion of China and India explains their patterns of industry exports.

After running the above-mentioned baseline fixed effects regressions, I test the existences of reverse causality from export patterns to $MatchIndex_{xit}$, correct selection bias caused by the occurrence of zero exports, and add an additional match index between the level of infrastructure in a country and length of production chains by industry. Details on these robustness checks are explained in Section 5.2.

4. Data

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

4.1 Exports

The industry export flow data on China and India are from the "National Bureau of Economic Research-United Nations (NEBR-UN) Trade Data, 1962-2000" constructed by Feenstra and Lipsey. Their dataset contains worldwide annual bilateral trade flows from 1962 to 2000, which are classified using the 4-digit Standard International Trade Classification (SITC) codes, revision 2 (for details of the dataset, see Feenstra et al., 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 4-digit SITC codes in the export data are converted to the 19 industry classifications used in my analysis.[7] In order to smooth the year-to-year fractionalization of exports, I use the three-year average export for the four periods: 1983-1985 (the first period, or t = 1), 1988-1990 (t = 2), 1993-1995 (t = 3), and 1998-2000 (t = 4). These three-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 and van Wincoop, 2003). The number of importers used in the present empirical analysis amounts to 175.

4.2 Match index of country skill dispersion and industry length of production chains

4.2.1 Skill dispersion indices of China and India

First, skill dispersion indices are constructed from the "International Data on Educational Attainment: Updates and Implications" constructed by Barro and Lee. Their data contains the distribution of educational attainment at various levels for the populations over age 15 and over age 25 in 138 countries at five-year intervals for the years 1960-2000 (for details, see Barro and Lee, 2000). Using the distribution data on educational attainment for the populations over age 15 of China and India, I constructed the three skill dispersion indices: CV_{xt}, $Gini_{xt}$, and MID_{xt}.

CV_{xt} is the coefficient of variation that results from dividing the variance of years of education (VAR_{xt}) by average years of education (AVG_{xt}) of the population in country *x* at time *t*, where $VAR_{xt} = \sum_e [(YEDU_{ext} - AVG_{xt})^2 P_{e\,xt}]$ and $AVG_{xt} = \sum_e YEDU_{ext} P_{ext}$. Subscript *e* denotes the level of educational attainment (no schooling, primary, secondary, and post-secondary); $YEDU_{ext}$ is the allocated years of

[7] Nineteen industries are equal to the 24 industries used in input-output table (see Section 4.2.2) minus the five service industries which are not included in the trade data. The 19 industry classifications may seem too broad, but they have advantages when we analyze the relatively small amount of exports from developing countries and allow the same industry classification across input-output tables of different countries to be maintained when computing the length of production chains.

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*.

$Gini_{xt}$ indicates the Gini coefficient computed from the distribution of years of education of the population as follows:

$$Gini_{xt} = [\sum_{e=2}^{4} \sum_{j=1}^{e-1} P_{ext} P_{jxt} (YEDU_{ext} - YEDU_{jxt})] / AVG_{xt} .$$

MID_{xt} is just the proportion of the population with primary and secondary education level. This index also measures the degrees of skill dispersion of China and India, as skill dispersion in India is characterized by larger proportions of populations with both no schooling and post-secondary level compared to China (Table 1).

Table 1 reports the three skill dispersion indices used in the analysis and other related skill measures in 1975, 1980, 1985, and 1990. In order to minimize the possibility of reverse causality, I use the measure of skill dispersion that is 8-10 years before the year in which the exports occur. 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 between 1975 and 1990 in both countries. The standardized indices used when constructing the match index are also reported in Table 1 (the average of each standardized indices equals zero).

4.2.2 Index for the length of industry production chains

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 tables of China and India is used. $Leontief_{xit}$ measures how many units of input industry *i* requires, both direct and indirect, to produce one unit of output in industry *i*.[8] The size of $Leontief_{xit}$ depends on the

[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

technology of the industry and varies across industries. I use this $Leontief_{xit}$ as a proxy for the length of production chains of industry *i*.

$Leontief_{xit}$ 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). Twenty-four industry classifications are used, as only 24 industry classifications are available for all years. 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 (CSO, 1990, 1997, 2000, 2005). The original 115 industries are consolidated into the 24 industries mentioned above by using the concordance table of Saluja and Yadav (2009) as a reference. In contrast to the skill dispersion indices, the timing of exports and the year for calculating $Leontief_{xit}$ are almost the same.

In developing countries in particular, firms sometimes have to depend on imported inputs. With imported inputs, firms may be able to produce good quality products and export them, regardless of the skill level of workers of the country. Thus, I calculate $Leontief_{xit}$ based on domestic inputs only, the quality of which is assumed to be negatively affected by low-skilled domestic workers (additional details are provided in the Appendix II).

Table 2 reports the calculated $Leontief_{xit}$ for 24 industries in China and India. Cells are highlighted when the length of the production chain is larger than the average of all 24 industries in both countries for all time periods. Table 3 ranks the industries according to the size of 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. This

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 generate chains of demand for inputs.

evidence indicates the existence of differences in input structure or in the technology of the industries between China and India. 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, Table 3 clearly shows that we need to take into account the differences in industry characteristics across countries.

4.2.3 Match index

$MatchIndex_{xit}$ (of industry *i* of exporter *x* at time *t*) is calculated in the following ways. $MatchIndex_{xit}(\text{MID})$ is constructed by multiplying the standardized skill dispersion index MID_{xt} (of exporter *x* at 8-10 years before time *t*) and the standardized $Leontief_{xit}$ (of industry *i* of exporter *x* at time *t*). $MatchIndex_{xit}(\text{CV})$ and $MatchIndex_{xit}(\text{Gini})$ are similarly computed by using CV_{xt} and $Gini_{xt}$, respectively, but negative one (-1) is multiplied further in the end; for example, $MatchIndex_{xit}(\text{CV}) =$ standardized CV_{xt} * standardized $Leontief_{xit}$ * (-1). 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).[9] 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

[9] 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.

length of production chains match better (i.e., when $MatchIndex_{xit}$ becomes larger), thus supporting my hypothesis.

The specification used by Bombardini et al. (2010) does not simultaneously test the two hypotheses above. They just examine whether countries with higher skill dispersion have more exports in industries characterized by higher degrees of substitutability across worker skills, by testing the coefficient for the interaction term of skill dispersion index and industry skill substitutability index. Since both indices are positive values, their specification is not suitable to identify the relationship that a country with lower skill dispersion has more exports in industries with lower skill substitutability. The relationship between their interaction term and the industry exports of a county is expected to be U-shaped rather than linear.

The relationships between standardized $Leontief_{xit}$ and $MatchIndex_{xit}(\mathrm{CV})$ of China and India are illustrated in Figure 4. It clearly illustrates that $MatchIndex_{xit}(\mathrm{CV})$ of China, which has a more equal skill distribution, becomes larger in industries with longer production chains, whereas $MatchIndex_{xit}(\mathrm{CV})$ of India, which has a more unequal skill distribution, becomes smaller in industries with shorter production chains. Table 4, which displays the four-period average of $MatchIndex_{xit}(\mathrm{CV})$, shows that China has a relatively larger $MatchIndex_{xit}(\mathrm{CV})$ in most of the manufacturing industries, while India has larger $MatchIndex_{xit}(\mathrm{CV})$ in agricultural, mining, and services, and some manufacturing industries such as “Timber and wooden products,” “Petroleum and petro products,” and “Non-metallic mineral products.”

4.3 Other control variables

The first group of additional control variables accounts for various conventional trade barriers between exporter–importer pairs. These control variables include logarithm of distance (ln(Distance)_*xm*); presence of colonial ties (Colonial ties_*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*).[10] Those variables are mostly constructed from the dataset built by Helpman et al. (2008), CEPII, 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 capital intensity of the exporting country, defined as capital stock divided by GDP (Kintensity_*x*); the interaction term of capital intensity of the exporting country with capital intensity of industry (Kintensity_*x*i*); standardized average years of education of the population over age 15 in the exporting country (StdAvgEdu_*x*); skill intensity of industry, defined as the ratio of working population over age 15 with post-secondary level education (Skillintensity_*i*); the interaction term of the skill intensity of industry with the skill intensity of the exporting country, defined as the ratio of population with post-secondary level education (Skillintensity_*x*i*); and the ratio of imported input to total input by industry (ImportRatio_*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 (e.g., capital, skilled workers) 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. Kintensity_*x* and StdAvgEdu_*x* are not included in estimating equation (1), since they are absorbed by exporter-time fixed effects.

5. Estimation Results

5.1 Baseline results

[10] Dummies which indicate that exporter/importer is landlocked or island country are omitted from the regression due to perfect collinearity.

The fixed effects regression results using $MatchIndex_{xit}(\text{CV})$ as the match index are reported in Table 5. Columns (1) and (3) show the results of estimation equation (1), and columns (2) and (4) show the results of equation (2). The estimates in columns (1) and (2) use all exports in the 19 non-service industries of China and India. However, exports of agricultural goods might also be affected by natural inputs such as land and weather, which are neither included as inputs in the input-output tables nor affected by worker skill levels. Similarly, exports from mining industry might depend on natural resource endowments of the country. Considering those unobserved factors on the primary industries, specifications (3) and (4) restrict the sample to only the 12 manufacturing industries. In order to control for industry differences in skill intensity and its interaction term with exporter skill intensity, specifications (5) and (6) restrict the sample to the observations for non-service and manufacturing industries in India, respectively.

Consistent with my hypothesis, the estimated coefficients for $MatchIndex_{xit}(\text{CV})$ are positive and statistically significant in all specifications. For example, in specification (1), a unit increase in $MatchIndex_{xit}(\text{CV})$ is associated with a 36.3% (=[exp(0.310)-1]*100) increase in industry exports in specification (1). Similarly, a unit increase in $MatchIndex_{xit}(\text{CV})$ in specifications (2)-(6) is associated with 33.3%, 10.8%, 26.9%, 25.0%, and 37.6% increase in industry exports, respectively. Even under the most modest estimate of 10.8%, for instance, if the $MatchIndex_{xit}(\text{CV})$ of the machinery industry in India (-0.229) had been the same as that of China (1.200) in the third period, 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 industry export determinants.

The estimated coefficients for standard trade barriers between exporter and importer exhibit the predicted signs except for GATT/WTO membership in the exporter-importer pair. Distance is negatively associated with exports, and geographic continuity, colonial ties, common legal systems, languages, and religions are all positively associated with exports, although the relationships are not statistically significant for common legal systems and languages. Capital intensity of the exporting country is positively associated with exports to a small degree, when restricting the sample to manufacturing. 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 average educational level of the population in the exporting country does not significantly influence the amount of exports. Somewhat unexpectedly, the degree of skill intensity in an industry is negatively correlated with exports in the Indian sample, as shown in columns (5) and (6), although it is insignificant for the manufacturing sector sample. Industry ratio of imported input is positively associated with exports when controlling time-invariant exporter-importer factors, as shown in columns (2) and (4), but is negatively associated with exports without such control.

I also obtain similar regression results when using $MatchIndex_{xit}(\text{Gini})$ and $MatchIndex_{xit}(\text{MID})$ as the match index. Table 6 reports only the estimated coefficients for the match index. Except for the coefficients for $MatchIndex_{xit}(\text{Gini})$ and $MatchIndex_{xit}(\text{MID})$ in specification (5) using the observations for Indian non-service industries, all the estimated coefficients for the match index are positive and statistically significant. The sizes of the coefficients across the three match indices are similar in each specification.

5.2 Robustness analyses

5.2.1 Test for endogeneity (reverse causality)

The patterns of industry exports may influence the skill distribution and the length of industry production chains of the exporting country. For instance, a high level of exports in an industry that intensively uses skilled workers may increase demand for such skilled workers. This demand raises the skill wage premium and induces more people to attain higher education, and thus changes the skill distribution of the economy. Similarly, a high level of export experience in an industry may increase the length of production chains by increasing the demand for high quality domestic suppliers and fostering those suppliers. On the other hand, a high level of export experience may decrease the length of production chains by increasing the demand for better quality inputs and thus result in the use of more imported inputs. The existence of such reverse causality from past exports to the match index would lead to the biased estimates for the coefficients of the match index in Tables 5 and 6.

In order to take into account the possibility of reverse causality, I construct an instrument for the match index and test whether the match index is endogenous or not.[11] The instrument ($MatchIndex_{xit}\_iv$) is constructed similarly to the match index by multiplying the two standardized variables.[12] The first is equal to one minus the three-year average ratio of primary and secondary enrollment to the population, which predates the skill dispersion measures by ten years.[13] The other is $Leontief_{xit}$ computed for Thailand using the "Asian International I/O Tables" by IDE-JETRO. Thailand is chosen because it is a relatively large developing country in Asia (although not as large as China or India) and exports varieties of goods including both manufacturing and agricultural goods. The population ratio of past school enrollment at the primary and secondary levels would affect the patterns of industry exports only through changing the proportion of middle-skilled workers. Similarly, the length of production chains in Thailand would only affect the industry exports of China and India

[11] 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. This treatment follows the approach taken by Bombardini et al. (2009). Accordingly, reverse causality only becomes a problem when exports in 1983-1985 affected the skill distribution patterns in 1990, which are the most recent skill distribution used in the estimation.
[12] As in the construction of match index, negative one is multiplied further when constructing an instrument for the *MatchIndex*(CV) and *MatchIndex*(Gini).
[13] When constructing an instrument for the *MatchIndex*(MID), [three-year average ratio of primary and secondary enrollment to the population] is used instead of [1 - three-year average ratio of primary and secondary enrollment to the population].

through the associations resulting from common industry-specific technological characteristics. Thus, $MatchIndex_{xit}\_iv$ satisfies the exclusion restriction necessary in order for the instrument to be valid.

In the first stage of testing for endogeneity in the match index, the index is regressed on the instrument ($MatchIndex_{xit}\_iv$) and all the exogenous variables in the baseline estimation. 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 (Wooldridge, 2006: pp.532-533).

Table 7 reports the estimation results for the first-stage and second-stage regressions with different fixed effect specifications. The first-stage estimation results in columns (1) an (3) show that our instrument is very strong and statistically significant at the one percent level. The F statistics of the instrument ($MatchIndex_{xit}\_iv$) is 2464 in column (1) and 1083 in column (3). The second-stage estimation results in columns (2) and (4) 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. Since the instrumental variable (IV) estimator is less efficient than ordinary least squares (OLS) when the explanatory variables are exogenous, I do not use the IV estimator as I now assume that our match index is exogenous and that there is no reverse causality from export to the match index in my analysis.

5.2.2 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 non-services export sample and 43.4% in manufacturing export sample), although both China and India have positive values for exports in each industry when aggregating industry exports across all importing countries. Zero bilateral industry trade indicates that exporting firms decide not to export their products to a certain country due to the presence of high trade barriers at time *t*. 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 pursue two strategies. The first strategy is to conduct a Poisson regression, which can include zero export value as the dependent variable. By comparing several estimation methods, Silva and Tenreyro (2006) proposed this Poisson method to deal with zeros in trade data (as well as a heteroskedasticity bias) when estimating gravity equations.

The second strategy is to employ a correction procedure for the panel data developed by Wooldridge (1995). Since we can observe not only the selection indicator to show whether export value is positive or zero but also the exact value of exports, I apply Procedures 3.1 and 4.1.1 of Wooldridge (1995), which use the Tobit-form selection equation. First, for each time period, residuals ($\hat{v}_{xmit}$, or Tobit_residuals) are obtained using the following cross-section Tobit model:

$$Export_{xmit} = \max(0, z_{xmit}\delta_t + v_{xmit}),$$

where z_{xmit} indicates all explanatory variables, including match index, in the baseline estimation. In the second stage, when we correct for selection bias in baseline equation (2), the following equation is estimated by time fixed effects over the observations for which export value is positive:

$$\ln(Export_{xmit}) = \beta''MatchIndex_{xit} + x_2\gamma'' + \rho\hat{v}_{xmit} + \alpha_t + \alpha_{xmi} + \varepsilon'_{xmit}.,$$

where the predicted residuals are obtained by $\hat{v}_{xmit} = Export_{xmit} - z_{xmit}\hat{\delta}_t$ from the first stage Tobit model.[14] x_2 equals to X_2 in baseline equation (2) minus one variable which satisfies the exclusion restrictions. If the estimated coefficient ρ is statistically significantly different from zero, we can conclude that selection bias exists. The

[14] I also experimented with adding the interactions of the predicted Tobit residuals and time dummies to allow ρ to vary across time and obtained similar results as reported in Table 9.

estimation for β'' is the unbiased estimator after controlling for selection.[15] In a similar manner, selection bias is controlled for in baseline equation (1).[16]

The excluded variable in the second stage needs to satisfy the exclusion restrictions, that is, it should affect selection but should not have a direct effect on the dependent variable ($\ln(Export_{xmit})$) in the second stage. In other words, it is necessary that the excluded variable affect $\ln(Export_{xmit})$ only through the selection mechanism. For such an excluded variable, I use either Religion_*xm* or Language_*xm*.[17] Both variables are identified by Helpman et al. (2008) to satisfy the exclusion restrictions in gravity equations. They are also significantly correlated with the dependent variable in most periods when estimating the first stage Tobit equations.

Tables 8 reports the estimation results from the Poisson regression. In all but one specification (column (4) with manufacturing industry samples with exporter-importer-industry fixed effects), the estimated coefficient of the match index is positive, consistent with my hypothesis. Four of them are statistically significant at a 5% or 10% level. Compared to the baseline estimation in Table 5, the size of the coefficient becomes larger in some specifications, as shown in columns (2), (3), and (5), and smaller in the remaining specifications, as shown in columns (1), (4), and (6). Similar results are found when using other skill dispersion indices (see Table A.1 in Appendix I).

The signs of the estimated coefficients for the conventional bilateral trade barriers are almost the same as those in Table 5. The somewhat unexpected negative signs now either become positive (ImportRatio_*i*, in most cases) or insignificant

[15] Significance tests are conducted based on the standard errors obtained by bootstrap with 300 replications. This bootstrap method is different from Wooldridge (1995) but similar to Zimmer (2010) who also applied the panel selection model proposed by Wooldridge (1995).
[16] When using baseline equation (1), "time" is not actual time but is defined as exporter-time pair. Thus, there are eight (2 exporters * 4 period) "time" indicators and eight cross-section Tobit regressions are conducted.
[17] Due to the construction of those variables, Religion_*xm* is time-variant but Language_*xm* is time-invariant between exporter-importer pair. Thus, Language_*xm* cannot be excluded from the second stage estimation when correcting for baseline equation (2) because Language_*xm* is absorbed by the exporter-importer-industry fixed effects.

(GATT_WTO_*xm* and Skillintensity_*i*). The interaction effect of skill intensity (Skillintensity_*x*i*) become weaker and insignificant.

Table 9 reports the second stage estimation results from the panel selection model proposed by Wooldridge (1995). The significant associations between the predicted Tobit residuals and the exports in all specifications indicate the presence of selection bias. The estimated coefficients of the match index after correcting for selection bias are positive and statistically significant in all specifications. Furthermore, the sizes of those coefficients are mostly similar to those obtained in Table 5. Similar results are obtained when using $MatchIndex_{xit}$ (Gini) and $MatchIndex_{xit}$ (MID) (see Table A.2 in Appendix I).

However, we need to be careful when interpreting the results in Table 9, because at least one of our instruments is not likely to satisfy the exclusion restrictions. As Helpman et al. (2008) argue, the exclusion restrictions require that after correcting for selection bias with one valid instrument, there should be no significant correlation between exports and the other instrument in the second stage estimation (Helpman el al., 2008, p.466). For example, if we assume that our Language_*xm* is a valid instrument, the significant correlation between Religion_*xm* and exports after controlling for selection, shown in columns (1) and (4), indicates that Religion_*xm* has a direct impact on exports and thus does not satisfy the exclusion restrictions. Similarly, conditional on the assumption that Religion_*xm* is a valid instrument, the significant coefficients of Language_*xm* in the second stage, shown in columns (2) and (5), suggest that Language_*xm* does not work as a valid instrument. Considering the validity of our instruments, our first strategy, the Poisson method, seems more appropriate to correct for selection bias.

5.2.3 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. However, in addition to worker skill, capacity and quality of production infrastructure, may also have different sizes of impacts on industry productivity according to the length of industry production chains. For instance, a poor quality power generation and distribution system that results in power failures may deteriorate the quality of inputs at each production stage. The underdevelopment of road networks and rugged roads may also damage the quality of input when it is transported from one stage of the production chain to the next. Similar to the impacts of low-skilled workers, these negative impacts may accumulate and thus cause more damage at the end of longer production chains.

In order to control for these negative impacts of poor infrastructure, which may also accumulate and become larger as the length of production chains increases, I construct two match indices: $MatchIndex_{xit}(\text{Powerloss})$ and $MatchIndex_{xit}(\text{Road})$. Each index is constructed similarly to $MatchIndex_{xit}(\text{CV})$ and $MatchIndex_{xit}(\text{MID})$, respectively. When constructing $MatchIndex_{xit}(\text{Powerloss})$, 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). In the case of $MatchIndex_{xit}(\text{Road})$, I use data on road density in China and India ($Road_{xt}$, kilometers of total road network divided by 100 sq. km of land area), instead of MID_{xt}.[18]

Table 10 reports the Poisson regression results having introduced either $MatchIndex_{xit}(\text{Powerloss})$ or $MatchIndex_{xit}(\text{Road})$ as an additional control variable. Each index is added separately, since adding both indices together creates a multicollinearity problem. The Poisson method is chosen because our religion or language instruments may not be valid corrections for selection bias as explained before. Results are only presented where all time-invariant exporter-importer factors are controlled by fixed effects. It is clear from Table 10 that even after controlling

[18] Both $Powerloss_{xt}$ and $Road_{xt}$ may not be the most adequate indicators to address the quantity and quality of power and transportation infrastructures, but they are the best variables I can obtain for both China and India for all four periods.

infrastructure using the infrastructure match index, positive coefficients of $MatchIndex_{xit}(\mathrm{CV})$ are obtained in all specifications, although in the India sample they are not statistically significant, as shown in columns (5)-(8). When both the China and India samples are used, the sizes of the coefficients become even larger compared to those in the columns (2) and (4) of Table 8. Turning our attention to the infrastructure match indices, we can notice that the effect of $MatchIndex_{xit}(\mathrm{Powerloss})$ is unexpectedly negative in all specifications. This may be because our variable, $Powerloss_{xt}$ just indicates the level of technology and quality of management of the power supply system, and is not related with the quality of power actually used in the production stage. In order to control the quality of power used in actual production, we might need data such as the frequency of power failures that are not restored by generator. The effect of $MatchIndex_{xit}(\mathrm{Road})$ is positive in all specifications as anticipated. Similar results are also obtained when using $MatchIndex_{xit}(\mathrm{Gini})$ and $MatchIndex_{xit}(\mathrm{MID})$ (see Table A.3 in Appendix I).

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

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 reverse causality, correct for selection bias, and control for infrastructure factors.

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. Furthermore, production chains tend to be shorter in India than in China, when comparing the same industries. Although the export data used in the empirical analysis do not cover service industries, the estimation results obtained 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. This finding indicates that if India would like to foster large-scale manufacturing industries 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 (2010) 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.

Incorporating service industries into this 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. Even when the length of production chains of a certain industry is long and the skill distribution of the entire economy is unequal, the negative impact of defect accumulation might be very small if industry inputs are produced only by high-skilled workers at all stages. Although it is difficult to obtain skill distribution data at a detailed industry level for many countries, including China, considering the skill content as well as the length of production chains would provide for more rigorous analysis. If we also increase the number of exporting countries, we can expand the comparison beyond only China and India, and thus increase skill distribution patterns that are examined. This possible future extension would also make the result of this paper more general.

References

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170-192.

Asuyama, Yoko. 2010. "Skill Formation through Education and Training: A Comparison of China and India." Mimeo. Institute of Developing Economies.

Barro, Robert J., and Jong-Wha Lee. 2000. "International Data on Educational Attainment: Updates and Implications." *CID Working Paper*, 42.

Barro, Robert J., and Rachel M. McCleary. 2005. "Which Countries Have State Religions?" *Quarterly Journal of Economics*, 120(4): 1331–1370.

Bombardini, Matilde, Giovanni Gallipoli, and German Pupato. 2009. "Skill Dispersion and Trade Flows." *NBER Working Paper*, 15097.

Chor, Davin. 2010. "Unpacking Sources of Comparative Advantage: A Quantitative Approach." *Journal of International Economics*, 82 (2): 152–167.

Cuñat, Alejandro, and Marc J. Melitz. 2007. "Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage." *NBER Working Paper*, 13062.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "World Trade Flows: 1962-2000." *NBER Working Paper*, 11040.

Gregory, Neil, Stanley Nollen, and Stoyan Tenev. 2009. *New Industries from New Places: The Emergence of the Software and Hardware Industries in China and India*. Washington, D.C and Stanford, California: Stanford University Press and World Bank.

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

Grossman, Gene M., and Giovanni Maggi. 2000. "Diversity and Trade." *American Economic Review*, 90(5): 1255-1275.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. "Estimating Trade Flows:

Trading Partners and Trading Volumes." *Quarterly Journal of Economics*, 123(2): 441-487.

Kochhar, Kalpana, Utsav Kumar, Raghuram Rajan, Arvind Subramanian, and Ioannis Tokatlidis. 2006. "India's Pattern of Development: What Happened, What Follows?" *Journal of Monetary Economics*, 53(5): 981-1019.

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

Levchenko, Andrei A. 2007. "Institutional Quality and International Trade." *Review of Economic Studies*, 74(3): 791–819.

Lo, Chu-Ping and Bih Jane Liu. 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, Thierry and Soledad Zignago. 2006. "Notes on CEPII's Distances Measures." http://www.cepii.fr/anglaisgraph/bdd/distances.htm (accessed October 21, 2010).

Nunn, Nathan. 2007. "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade." *Quarterly Journal of Economics*, 122(2): 569-600.

Ohnsorge, Franziska, and Daniel Trefler. 2007. "Sorting It Out: International Trade with Heterogeneous Workers." *Journal of Political Economy*, 115(5): 868-892.

Saluja, M.R. and Bhupesh Yadav. 2009. "Construction of Input-Output Table for Indian Economy-2005." In Kuwamori, Hiroshi, Yoko Uchida, and Satoshi Inomata, eds. *Compilation and Use of the 2005 International Input-Output Tables*. Chiba: IDE-JETRO.

Silva, J. M. C. Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Syrquin, Moshe. 1989. "Patterns of Structural Change." In Chenery, Hollis and T.N. Srinivasan, eds. *Handbook of Development Economics*, Volume I. Amsterdam:

Elsevier Science B.V.

Tang, Heiwai. 2010. "Labor Market Institutions, Firm-specific Skills, and Trade Patterns." *Centro Studi Luca D'Agliano Development Studies Working Paper*, 301.

Wooldridge, Jeffrey M. 1995. "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions." *Journal of Econometrics*, 68 (1): 115-132.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics: A Modern Approach*, 3rd edition. Mason, Ohio: Thomson/South-Western.

Zimmer, David M. 2010. "The Role of Health Insurance in Labor Supply Decisions of Divorced Females." *Quarterly Review of Economics and Finance*, 50 (2): 121–131.

Statistics

Barro, Robert J., and Jong-Wha Lee. 2000. *International Data on Educational Attainment: Updates and Implications*. http://www.cid.harvard.edu/ciddata/ciddata.html (accessed October 6, 2008).

Barro, Robert J., and Rachel M. McCleary. *Religion Adherence Data*. http://www.hks.harvard.edu/fs/rmcclea/data.html (accessed October 21, 2010).

CSO (Central Statistical Organisation, Ministry of Statistics and Programme Implementation, Government of India). 1990. *Input-Output Transactions Table 1983-84*.

CSO. 1997. *Input-Output Transactions Table 1989-90*.

CSO. 2000. *Input-Output Transactions Table 1993-94*.

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

CSO. 2007. *National Accounts Statistics –Back series 1950-51 to 1999-2000*. New

Delhi: CSO.

CSO. 2008. *National Accounts Statistics 2008*. New Delhi: CSO.

CEPII (Centre D'Etudes Prospectives et D'Informations Internationales). *Distances Measures*. http://www.cepii.fr/anglaisgraph/bdd/distances.htm (accessed October 21, 2010).

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

EPW Research Foundation. 2002. *Annual Survey of Industries 1973-74 to 1997-98: A Data Base on the Industrial Sector in India*. Mumbai, India: EPW Research Foundation.

Feenstra, Robert, and Robert Lipsey. *NBER-United Nations Trade Data, 1962-2000*. http://cid.econ.ucdavis.edu/data/undata/undata.html (accessed October 3, 2008)

Ghosh, Buddhadeb and Prabir De. 2005. *India Infrastructure Database 2005*, Vol.II. New Delhi: Bookwell.

Holz, Carsten 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*. http://www.imf.org/external/pubs/ft/weo/2010/02/weodata/index.aspx (accessed October 22, 2010).

MHRD (Ministry of Human Resource Development). 2008. *Selected Educational Statistics: 2005-06, All India, Time-Series*. http://www.educationforallinindia.com/SESall-india-time-series-2005-06.pdf (accessed December 10. 2010).

MHRD and NIC (Ministry of Human Resource Development, Department of Education, and National Informatics Centre). *Compilation on 50 years of Indian Education: 1947-1997*. http://education.nic.in/cd50years/home.htm (accessed December 10. 2010).

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).*

Quan guo ren kou chou yang diao cha ban gong shi. 1997. *1995 National 1% Population Sample Survey*. Beijing: China Statistics Press.

SC and NBS (State Council and National Bureau of Statistics of China). 2002. *Tabulation on the 2000 Population Census of the People's Republic of China*. Beijing: China Statistics Press.

SC and NBS. 2007. *2005 National 1% Population Sample Survey*. Beijing: China Statistics Press.

SC and SSB (State Council and State Statistical Bureau of China). 1985. *1982 Population Census of China: Results of Computer Tabulation*. Beijing: China Statistics Press.

SC and SSB. 1993. *Tabulation on the 1990 Population Census of the People's Republic of China*. Beijing: China Statistics Press.

SSB (State Statistical Bureau of China). 1988. *Tabulations of China 1% Population Sample Survey*. Beijing: China Statistics Press.

UN (United Nations). *National Accounts Main Aggregates Database*. http://unstats.un.org/unsd/snaama/Introduction.asp (accessed October 22, 2010).

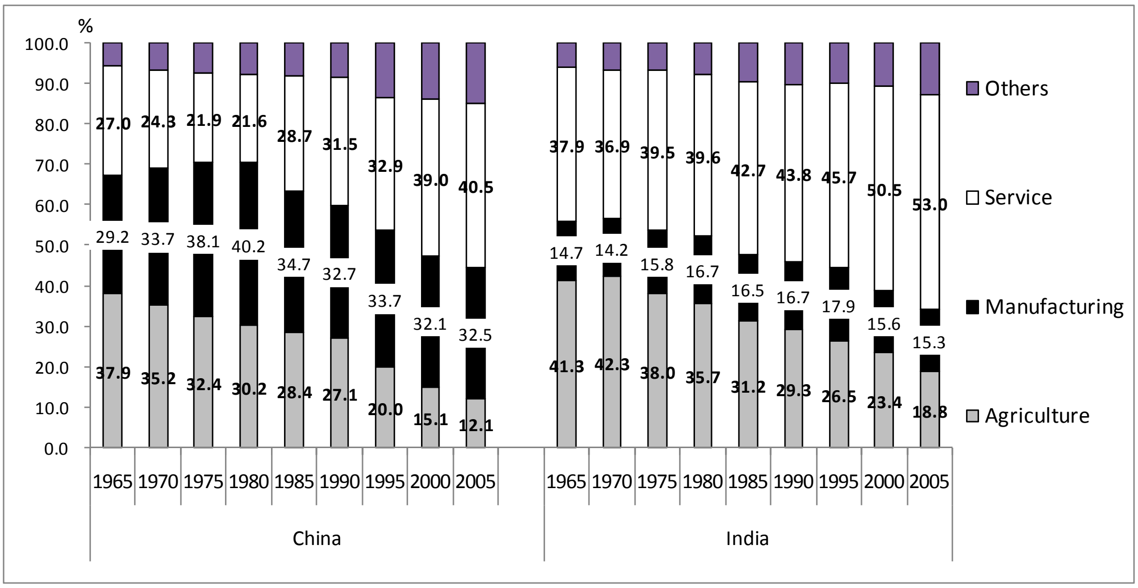
UN. 1993. *Statistical Yearbook 1990/91*. New York: United Nations.

World Bank. 2010. *World Development Indicators & Global Development Finance*. http://data.worldbank.org/data-catalog/world-development-indicators (accessed October 4, 2010).

WTO (World Trade Organization). 2010a. *International Trade Statistics 2010*. http://www.wto.org/english/res_e/statis_e/its2010_e/its10_toc_e.htm (accessed December 28, 2010).

WTO. 2010b. *WTO Trade Profiles*. http://stat.wto.org/Home/WSDBHome.aspx?Language=E (accessed December 28, 2010).

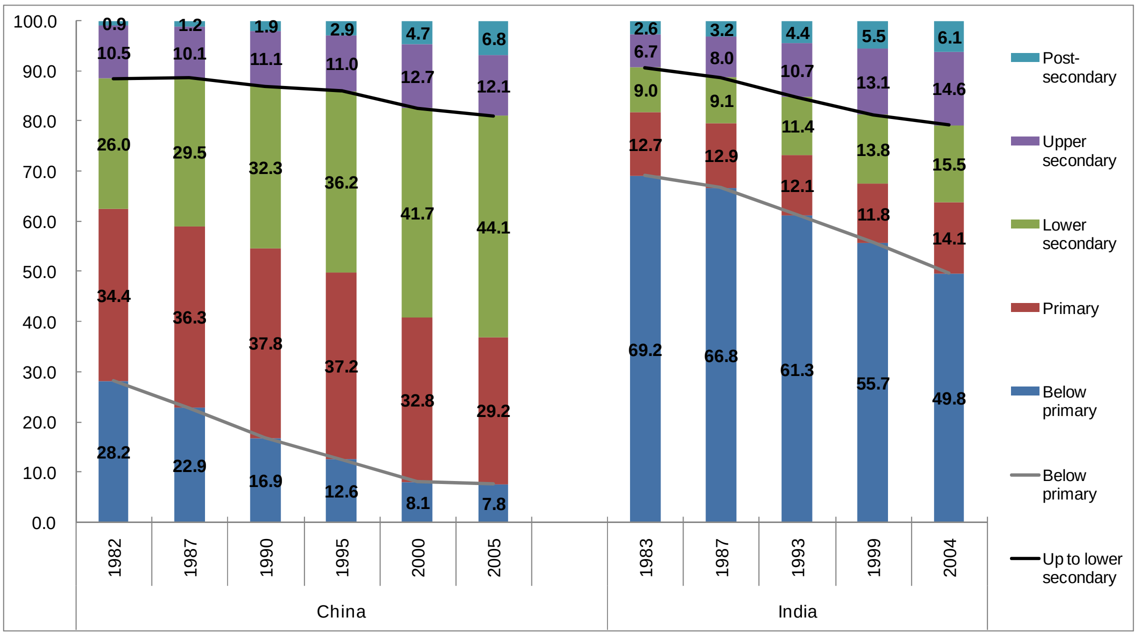
Figure 1. GDP composition by broad sector



Note: "Others" include mining and quarrying; electricity, gas and water supply; and construction.

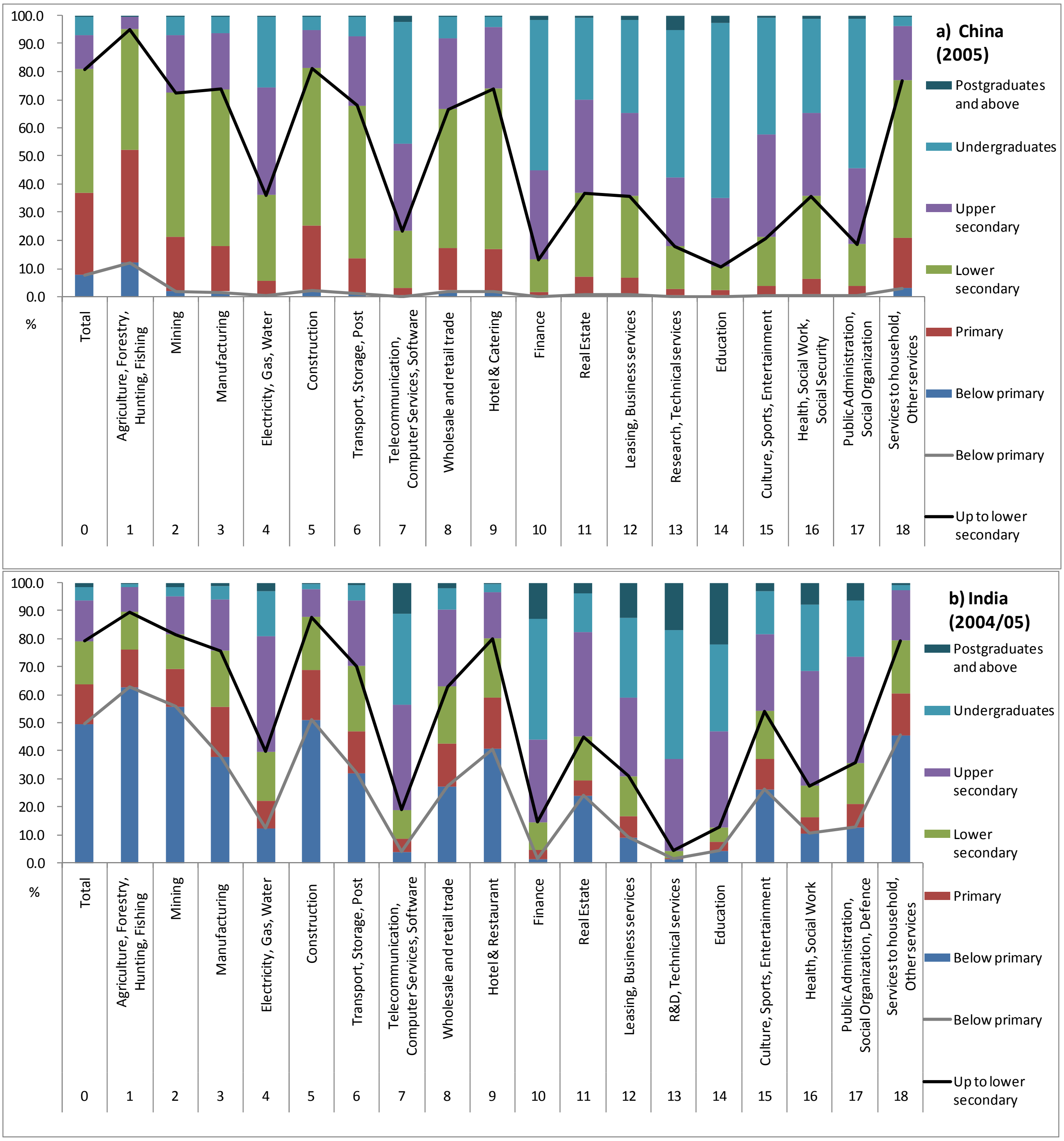
Source: World Bank (2010).

Figure 2. Educational attainment of employed population



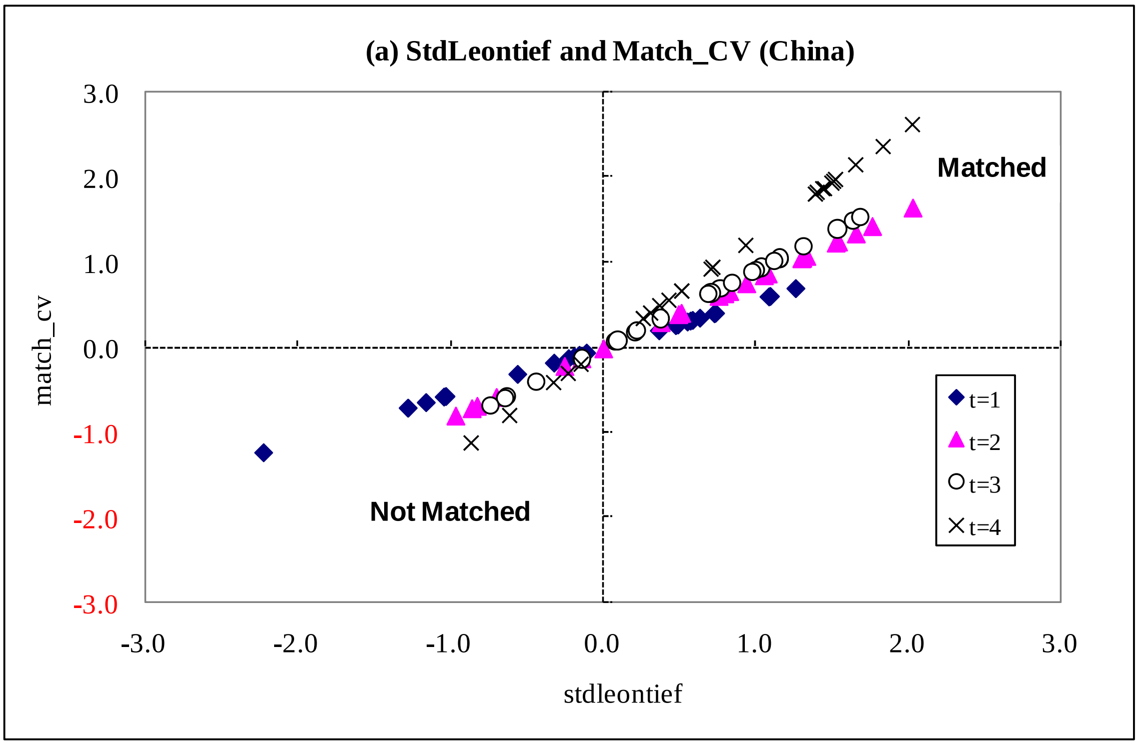
Note: In India, a person is considered working based on the usual principal activity status.
Sources: China: Population census of 1982, 1990, and 2000 (SC and SSB, 1985, 1993; SC and NBS, 2002); 1% population sample survey of 1987, 1995, and 2005 (SSB, 1988; Quan guo ren kou chou yang diao cha ban gong shi, 1997; SC and NBS, 2007). India: NSSO, *Unit-level data of National Sample Survey (NSS), Employment and Unemployment schedule, 1983, 1987-88, 1993-94, 1999-00, 2004-05*.

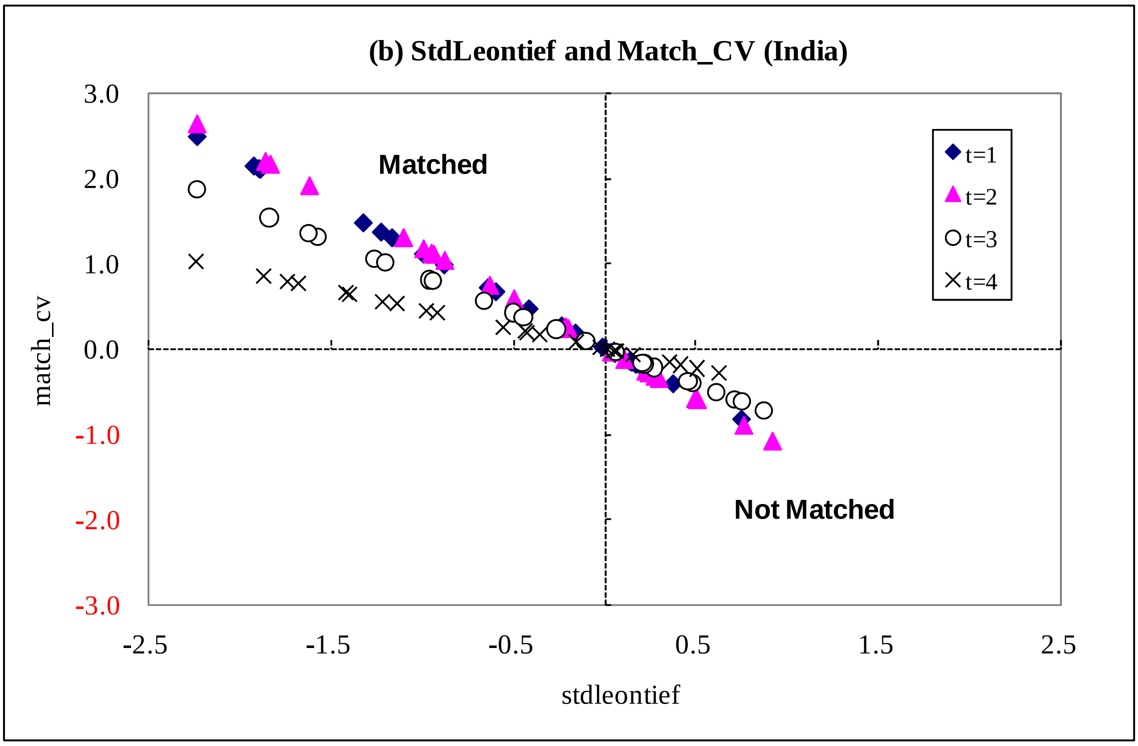
Figure 3. Educational attainment of employed population by industry



Note: In India, a person is considered working based on the usual principal activity status.
Sources: China: 1% population sample survey of 2005 (SC and NBS, 2007); India: NSSO, *Unit-level data of National Sample Survey (NSS), Employment and Unemployment schedule, 2004-05*.

Figure 4. Standardized index for the length of production chains ($Leontief_{xit}$) and match index ($MatchIndex_{xit}$ (CV))





Note: Index results are plotted for all 24 industries including the five service sectors.

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

Exporter	Year	Raw variables						Standardized variables		
		Average years of education	Percentage of population with no schooling	Percentage of population with post-secondary education	CV	Gini	MID	CV	Gini	MID
China	1975	4.380	40.2	0.9	0.925	0.502	58.9	-0.552	-0.436	0.438
	1980	4.760	34.0	0.9	0.829	0.452	65.0	-0.815	-0.771	0.788
	1985	4.940	31.5	1.2	0.792	0.431	67.4	-0.916	-0.909	0.925
	1990	5.850	22.2	1.9	0.653	0.352	75.9	-1.292	-1.444	1.413
India	1975	2.700	62.6	2.1	1.535	0.706	35.3	1.108	0.946	-0.915
	1980	3.270	66.6	2.4	1.561	0.731	31.1	1.178	1.114	-1.156
	1985	3.640	61.6	2.8	1.436	0.699	35.5	0.837	0.897	-0.904
	1990	4.100	55.8	3.3	1.295	0.656	41.0	0.454	0.603	-0.588

Notes: For the definitions of CV, Gini, and MID index, see Section 4.2.1 in the main text and Appendix II.
Sources: Average years of education: Barro and Lee (2000); other variables: computed by author from Barro and Lee (2000).

Table 2. Index for the length of industry production chains ($Leontief_{xit}$)

Industry \ Exporter / Period / Year of IO Table	China t1 1985	China t2 1990	China t3 1995	China t4 2000	China Average	India t1 1983	India t2 1989	India t3 1993	India t4 1998	India Average
1 Paddy		1.568	1.664	1.829	1.687	1.515	1.665	1.653	1.455	1.572
2 Other agricultural products	1.394	1.524	1.659	1.792	1.592	1.419	1.469	1.404	1.339	1.408
3 Livestock and poultry	1.693	1.867	2.013	2.057	1.907	2.239	1.839	1.719	1.697	1.873
4 Forestry	1.443	1.582	1.619	1.567	1.553	1.144	1.155	1.164	1.153	1.154
5 Fishery	1.497	1.635	1.742	1.865	1.685	1.129	1.255	1.275	1.230	1.223
6 Crude petroleum and natural gas	1.492	2.130	1.868	1.674	1.791	1.140	1.167	1.253	1.206	1.192
7 Other mining	1.881	2.315	2.405	2.140	2.185	1.378	1.533	1.536	1.347	1.449
8 Food, beverage and tobacco	2.131	2.240	2.245	2.314	2.233	1.951	1.837	1.817	1.918	1.881
9 Textile, leather, and the products thereof	2.452	2.468	2.358	2.553	2.458	2.024	2.308	2.224	2.187	2.186
10 Timber and wooden products	2.079	2.269	2.343	2.560	2.313	1.861	1.720	1.743	1.746	1.768
11 Pulp, paper and printing	2.230	2.559	2.332	2.222	2.336	1.989	2.132	2.127	1.955	2.051
12 Chemical products	2.190	2.465	2.277	2.528	2.365	2.014	2.137	2.117	1.948	2.054
13 Petroleum and petro products	1.862	2.374	2.220	2.082	2.134	1.662	1.515	1.528	1.548	1.563
14 Rubber products	2.171	2.138	2.084	2.547	2.235	1.999	2.027	2.239	2.139	2.101
15 Non-metallic mineral products	2.125	2.565	2.564	2.511	2.441	1.680	2.050	1.884	1.751	1.841
16 Metal products	2.157	2.768	2.607	2.689	2.555	2.083	2.242	2.289	1.932	2.137
17 Machinery	2.233	2.614	2.471	2.506	2.456	1.830	1.971	2.041	1.864	1.926
18 Transport equipment	2.382	2.479	2.407	2.769	2.509	1.923	2.040	2.183	2.073	2.055
19 Other manufacturing products	2.165	2.363	2.392	2.525	2.361	1.755	1.844	2.017	2.100	1.929
20 Electricity, gas, and water supply	1.793	2.256	2.213	2.227	2.122	1.985	2.019	2.013	1.992	2.002
21 Construction	2.379	2.658	2.626	2.616	2.570	1.946	1.941	1.951	1.779	1.904
22 Trade and transport	1.848	2.083	1.959	2.105	1.999	1.562	1.538	1.536	1.422	1.515
23 Services	1.833	1.926	2.017	2.036	1.953	1.444	1.563	1.429	1.524	1.490
24 Public administration	1.000	1.820	1.965	2.141	1.732	1.000	1.000	1.000	1.000	1.000
Average	1.932	2.194	2.169	2.244	2.132	1.695	1.749	1.756	1.679	1.720

Notes: Each figure indicates the column sum of the Leontief inverse coefficient of each industry computed from the input-output tables of China and India (See Section 4.2.2 in the main text and Appendix II). Only domestic inputs are used in the calculation. Cells are highlighted when the length of the production chains is larger than 1.927, which is the average length for all 24 industries in both countries in all 4 time periods.

Table 3. Ranking of the four-period average index for the length of industry production chains ($Leontief_{xit}$)

Total rank	Rank in own country	Exporter	Sector	Industry (code and description)		4-period average Leontief	StdLeontief
1	1	China	Service	21	Construction	2.570	1.548
2	2	China	Manufacturing	16	Metal products	2.555	1.513
3	3	China	Manufacturing	18	Transport equipment	2.509	1.403
4	4	China	Manufacturing	9	Textile, leather, and the products thereof	2.458	1.278
5	5	China	Manufacturing	17	Machinery	2.456	1.274
6	6	China	Manufacturing	15	Non-metallic mineral products	2.441	1.238
7	7	China	Manufacturing	12	Chemical products	2.365	1.055
8	8	China	Manufacturing	19	Other manufacturing products	2.361	1.046
9	9	China	Manufacturing	11	Pulp, paper and printing	2.336	0.985
10	10	China	Manufacturing	10	Timber and wooden products	2.313	0.929
11	11	China	Manufacturing	14	Rubber products	2.235	0.741
12	12	China	Manufacturing	8	Food, beverage and tobacco	2.233	0.736
13	1	India	Manufacturing	9	Textile, leather, and the products thereof	2.186	0.623
14	13	China	Mining	7	Other mining	2.185	0.622
15	2	India	Manufacturing	16	Metal products	2.137	0.505
16	14	China	Manufacturing	13	Petroleum and petro products	2.134	0.499
17	15	China	Service	20	Electricity, gas, and water supply	2.122	0.469
18	3	India	Manufacturing	14	Rubber products	2.101	0.418
19	4	India	Manufacturing	18	Transport equipment	2.055	0.307
20	5	India	Manufacturing	12	Chemical products	2.054	0.305
21	6	India	Manufacturing	11	Pulp, paper and printing	2.051	0.298
22	7	India	Service	20	Electricity, gas, and water supply	2.002	0.181
23	16	China	Service	22	Trade and transport	1.999	0.172
24	17	China	Service	23	Services	1.953	0.062
25	8	India	Manufacturing	19	Other manufacturing products	1.929	0.005
26	9	India	Manufacturing	17	Machinery	1.926	-0.002
27	18	China	Agriculture	3	Livestock and poultry	1.907	-0.048
28	10	India	Service	21	Construction	1.904	-0.055
29	11	India	Manufacturing	8	Food, beverage and tobacco	1.881	-0.112
30	12	India	Agriculture	3	Livestock and poultry	1.873	-0.129
31	13	India	Manufacturing	15	Non-metallic mineral products	1.841	-0.207
32	19	China	Mining	6	Crude petroleum and natural gas	1.791	-0.328
33	14	India	Manufacturing	10	Timber and wooden products	1.768	-0.384
34	20	China	Service	24	Public administration	1.732	-0.471
35	21	China	Agriculture	1	Paddy	1.687	-0.578
36	22	China	Agriculture	5	Fishery	1.685	-0.584
37	23	China	Agriculture	2	Other agricultural products	1.592	-0.807
38	15	India	Agriculture	1	Paddy	1.572	-0.856
39	16	India	Manufacturing	13	Petroleum and petro products	1.563	-0.877
40	24	China	Agriculture	4	Forestry	1.553	-0.902
41	17	India	Service	22	Trade and transport	1.515	-0.994
42	18	India	Service	23	Services	1.490	-1.053
43	19	India	Mining	7	Other mining	1.449	-1.153
44	20	India	Agriculture	2	Other agricultural products	1.408	-1.251
45	21	India	Agriculture	5	Fishery	1.223	-1.697
46	22	India	Mining	6	Crude petroleum and natural gas	1.192	-1.772
47	23	India	Agriculture	4	Forestry	1.154	-1.863
48	24	India	Service	24	Public administration	1.000	-2.234

Table 4. Four-period average $MatchIndex_{xit}$ (CV)

Industry	4-period average			
	China	(Rank)	India	(Rank)
1 Paddy	-0.530	(22)	0.728	(10)
2 Other agricultural products	-0.628	(23)	1.088	(5)
3 Livestock and poultry	0.041	(18)	0.022	(15)
4 Forestry	-0.781	(24)	1.666	(2)
5 Fishery	-0.437	(21)	1.528	(4)
6 Crude petroleum and natural gas	-0.275	(20)	1.601	(3)
7 Other mining	0.604	(13)	1.002	(6)
8 Food, beverage and tobacco	0.698	(12)	0.106	(13)
9 Textile, leather, and the products thereof	1.164	(6)	-0.555	(24)
10 Timber and wooden products	0.940	(9)	0.333	(11)
11 Pulp, paper and printing	0.864	(10)	-0.295	(20)
12 Chemical products	1.013	(8)	-0.308	(21)
13 Petroleum and petro products	0.479	(15)	0.774	(9)
14 Rubber products	0.754	(11)	-0.333	(22)
15 Non-metallic mineral products	1.185	(5)	0.148	(12)
16 Metal products	1.457	(1)	-0.512	(23)
17 Machinery	1.189	(4)	-0.006	(17)
18 Transport equipment	1.342	(3)	-0.246	(19)
19 Other manufacturing products	1.015	(7)	0.081	(14)
20 Electricity, gas, and water supply	0.507	(14)	-0.165	(18)
21 Construction	1.431	(2)	0.006	(16)
22 Trade and transport	0.206	(16)	0.855	(8)
23 Services	0.102	(17)	0.942	(7)
24 Public administration	-0.173	(19)	1.997	(1)
Average	0.507		0.436	

Note: See Section 4.2.3 in the main text.

Table 5. Determinants of comparative advantage (baseline results)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, Manu
MatchIndex (Skill)	CV	CV	CV	CV	CV	CV
MatchIndex_x*i	0.310 ***	0.287 ***	0.103 *	0.238 ***	0.224 ***	0.319 ***
	(0.030)	(0.051)	(0.055)	(0.061)	(0.081)	(0.107)
ImportRatio_i	-0.004	0.028 ***	-0.010 ***	0.011 *	-0.005	-0.018 **
	(0.004)	(0.005)	(0.004)	(0.006)	(0.009)	(0.008)
Kintensity_x		-0.001		0.006 *		
		(0.003)		(0.004)		
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		-0.006		-0.139		
		(0.116)		(0.131)		
Skillintensity_i					-3.013 **	-1.304
					(1.182)	(1.164)
Skillintensity_x*i					0.162 ***	0.072 *
					(0.033)	(0.038)
ln(Distance)_xm	-0.868 ***		-0.787 ***			
	(0.145)		(0.170)			
Contiguous_xm	0.796 ***		0.624 **			
	(0.216)		(0.255)			
Legalsystem_xm	0.113		0.112			
	(0.069)		(0.076)			
Colonial ties_xm	0.184		0.497 **			
	(0.311)		(0.249)			
Language_xm	0.056		0.113			
	(0.094)		(0.102)			
Religion_xm	0.000 ***	0.002 ***	0.000 ***	0.003 ***	0.001 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
GATT_WTO_xm	-0.141 *	-0.104	-0.112	-0.081	-0.280 **	-0.250 **
	(0.085)	(0.078)	(0.089)	(0.083)	(0.118)	(0.122)
Exporter-Time FE	Yes		Yes			
Importer-Industry FE	Yes		Yes		Yes	Yes
Time FE		Yes		Yes	Yes	Yes
Exporter-Importer-Industry FE		Yes		Yes		
Number of observations	11017	11017	8555	8555	5110	3886
R-squared	0.757	0.857	0.752	0.854	0.872	0.874
F-statistics	82.27	100.17	86.14	125.84	63.95	80.86

Notes: The dependent variable is the logarithm of three-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 importer-industry pair in (1), (3), (5), and (6), or clustered by exporter-importer-industry pair in (2) and (4) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table 6. Estimated coefficients of the three match indices (baseline results)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, Manu
MatchIndex_x*i (CV)	0.310 ***	0.287 ***	0.103 *	0.238 ***	0.224 ***	0.319 ***
	(0.030)	(0.051)	(0.055)	(0.061)	(0.081)	(0.107)
MatchIndex_x*i (Gini)	0.302 ***	0.259 ***	0.103 **	0.176 ***	0.111	0.329 ***
	(0.029)	(0.049)	(0.052)	(0.057)	(0.098)	(0.114)
MatchIndex_x*i (MID)	0.300 ***	0.254 ***	0.106 **	0.190 ***	0.109	0.315 ***
	(0.029)	(0.049)	(0.052)	(0.058)	(0.092)	(0.110)

Notes: The dependent variable is the logarithm of three-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 importer-industry pair in (1), (3), (5), and (6), or clustered by exporter-importer-industry pair in (2) and (4) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table 7. Tests for endogeneity

	(1)	(2)	(3)	(4)
	First stage	Endogeneity test	First stage	Endogeneity test
Dependent variable	MatchIndex_x*i	ln(Export)_xmi	MatchIndex_x*i	ln(Export)_xmi
Sample	NonServ	NonServ	NonServ	NonServ
MatchIndex (Skill)	CV	CV	CV	CV
MatchIndex_x*i		0.256 ***		0.297
		(0.070)		(0.278)
MatchIndex_iv	0.291 ***		0.100 ***	
	(0.006)		(0.003)	
ImportRatio_i	0.005 ***	-0.004	-0.004 ***	0.028 ***
	(0.001)	(0.004)	(0.001)	(0.005)
Kintensity_x			-0.028 ***	-0.001
			(0.000)	(0.008)
Kintensity_x*i	0.000 ***	0.000	0.000 ***	0.000 ***
	(0.000)	(0.000)	(0.000)	(0.000)
stdAvgEdu_x			-0.119 ***	-0.007
			(0.024)	(0.127)
ln(Distance)_xm	0.000	-0.868 ***		
	(0.051)	(0.145)		
Contiguous_xm	0.000	0.796 ***		
	(0.080)	(0.216)		
Legalsystem_xm	0.000	0.113 *		
	(0.034)	(0.069)		
Colonial ties_xm	0.000	0.182		
	(0.177)	(0.312)		
Language_xm	0.000	0.056		
	(0.048)	(0.094)		
Religion_xm	0.000	0.000 ***	0.000 **	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)
GATT_WTO_xm	0.000	-0.140 *	-0.001	-0.104
	(0.016)	(0.085)	(0.014)	(0.078)
1st stage residuals		0.063		-0.010
		(0.075)		(0.279)
Exporter-Time FE	Yes	Yes		
Importer-Industry FE	Yes	Yes		
Time FE			Yes	Yes
Exporter-Importer-Industry FE			Yes	Yes
Number of observations	23716	11017	23716	11017
R-squared	0.440	0.757	0.866	0.857
F-statistics	1622.88	60.66	1087.60	56.34

Notes: The dependent variables in (1) and (3) are the match index of exporter *x* in industry *i* using CV as the exporter skill inequality measure. The dependent variables in (2) and (4) are the logarithm of three-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 importer-industry pair in (1) and (2), or clustered by exporter-importer-industry pair in (3) and (4) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table 8. Selectioncorrected estimates I (Poisson regression)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, Manu
MatchIndex (Skill)	CV	CV	CV	CV	CV	CV
MatchIndex_x*i	0.163 **	0.313 **	0.174 *	-0.017	0.291 **	0.202
	(0.075)	(0.160)	(0.099)	(0.126)	(0.145)	(0.187)
ImportRatio_i	0.013	0.052 ***	0.004	0.021	0.003	-0.018
	(0.008)	(0.013)	(0.009)	(0.013)	(0.016)	(0.015)
Kintensity_x		0.008		0.025 ***		
		(0.009)		(0.009)		
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		0.592		-0.105		
		(0.405)		(0.334)		
Skillintensity_i					-0.988	0.958
					(1.729)	(1.334)
Skillintensity_x*i					0.013	-0.048
					(0.043)	(0.052)
ln(Distance)_xm	-1.179 ***		-1.001 ***			
	(0.284)		(0.281)			
Contiguous_xm	0.943 ***		1.062 ***			
	(0.341)		(0.358)			
Legalsystem_xm	0.085		0.106			
	(0.145)		(0.151)			
Colonial ties_xm	0.832 **		0.901 **			
	(0.389)		(0.457)			
Language_xm	0.088		-0.073			
	(0.224)		(0.267)			
Religion_xm	0.000 **	0.001 ***	0.000 **	0.001 ***	0.002 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GATT_WTO_xm	0.009	-0.040	-0.009	-0.082	-0.029	0.058
	(0.176)	(0.166)	(0.167)	(0.175)	(0.166)	(0.157)
Exporter-Time FE	Yes		Yes			
Importer-Industry FE	Yes		Yes		Yes	Yes
Time FE		Yes		Yes	Yes	Yes
Exporter-Importer-Industry FE		Yes		Yes		
Number of observations	17182	14978	12752	11424	7036	5230

Notes: The dependent variable in each fixed effects Poisson regression is the three-year average exports (divided by [GDP of exporter x * GDP of importer m]) from exporter x to importer m in industry i. Bootstrap standard errors with 300 replications, clustered by importer-industry pair in (1), (3), (5), and (6), or clustered by exporter-importer-industry pair in (2) and (4) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table 9. Selection corrected estimates II (proposed by Wooldridge, 1995)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Excluded variables	Language_xm	Religion_xm	Religion_xm	Language_xm	Religion_xm	Religion_xm	Religion_xm	Religion_xm
Sample	NonServ	NonServ	NonServ	Manu	Manu	Manu	India, NonServ	India, Manu
MatchIndex (Skill)	CV	CV	CV	CV	CV	CV	CV	CV
MatchIndex_x*i	0.326 ***	0.324 ***	0.315 ***	0.103 *	0.111 **	0.248 ***	0.211 ***	0.247 **
	(0.031)	(0.031)	(0.050)	(0.055)	(0.052)	(0.062)	(0.081)	(0.098)
ImportRatio_i	-0.006	-0.005	0.028 ***	-0.012 ***	-0.012 ***	0.008	-0.007	-0.019 **
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)	(0.008)	(0.007)
Kintensity_x			0.001			0.009 **		
			(0.003)			(0.004)		
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)	(0.000)	(0.000)	(0.000)	(0.000)
stdAvgEdu_x			-0.092			-0.242		
			(0.119)			(0.150)		
Skillintensity_i							-3.087 ***	-1.576
							(1.125)	(1.124)
Skillintensity_x*i							0.168 ***	0.077 **
							(0.032)	(0.037)
ln(Distance)_xm	-0.890 ***	-1.316 ***		-0.792 ***	-1.227 ***			
	(0.143)	(0.127)		(0.171)	(0.141)			
Contiguous_xm	0.695 ***	0.823 ***		0.530 **	0.642 **			
	(0.238)	(0.232)		(0.259)	(0.259)			
Legalsystem_xm	0.140 **	0.131 *		0.163 **	0.141			
	(0.066)	(0.075)		(0.071)	(0.086)			
Colonial ties_xm	0.233	-0.005		0.556 **	0.280			
	(0.325)	(0.308)		(0.242)	(0.245)			
Language_xm		0.175 *			0.223 **			
		(0.091)			(0.099)			
Religion_xm	0.000 ***			0.000 ***				
	(0.000)			(0.000)				
GATT_WTO_xm	-0.117	-0.107	-0.061	-0.097	-0.085	-0.052	-0.175	-0.110
	(0.084)	(0.086)	(0.080)	(0.093)	(0.092)	(0.079)	(0.113)	(0.118)
Tobit_residuals	23902 ***	24640 ***	34507 ***	22759 ***	23564 ***	30202 **	63437 **	107738 ***
	(6466)	(7661)	(9157)	(6580)	(7130)	(13230)	(28595)	(32714)
Exporter-Time FE	Yes	Yes		Yes	Yes			
Importer-Industry FE	Yes	Yes		Yes	Yes		Yes	Yes
Time FE			Yes			Yes	Yes	Yes
Exporter-Importer-Industry FE			Yes			Yes		
Number of observations	11017	11017	11017	8555	8555	8555	5110	3886

Notes: This table reports the estimation results from the panel selection model proposed by Wooldridge (1995), Procedure 3.1 and 4.1.1. The dependent variable in each regression is the logarithm of three-year average exports (divided by [GDP of exporter *x* * GDP of importer *m*]) from exporter *x* to importer *m* in industry *i*. Bootstrap standard errors with 300 replications, clustered by importer-industry pair in (1), (2), (4), (5), (7), and (8), or clustered by exporter-importer-industry pair in (3) and (6) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table 10. Infrastructure match index added (Poisson regression)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Infrastructure variables	Powerloss	Road	Powerloss	Road	Powerloss	Road	Powerloss	Road
Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, NonServ	India, Manu	India, Manu
MatchIndex (Skill)	CV	CV	CV	CV	CV	CV	CV	CV
MatchIndex_x*i (Skill)	0.604 ***	0.476 **	0.519 ***	0.296 *	0.215	0.136	0.287	0.277
	(0.208)	(0.187)	(0.194)	(0.170)	(0.147)	(0.140)	(0.178)	(0.190)
MatchIndex_x*i (Powerloss)	-0.559 ***		-0.682 ***		-0.544 ***		-0.319 *	
	(0.132)		(0.194)		(0.179)		(0.189)	
MatchIndex_x*i (Road)		0.605 ***		0.489 **		0.584 ***		0.319 *
		(0.170)		(0.201)		(0.122)		(0.175)
ImportRatio_i	0.044 ***	0.034 ***	0.023 *	0.020	0.019	0.022	-0.003	-0.002
	(0.013)	(0.013)	(0.013)	(0.014)	(0.017)	(0.015)	(0.018)	(0.017)
Kintensity_x	0.013	0.014 *	0.025 ***	0.024 ***				
	(0.009)	(0.008)	(0.008)	(0.008)				
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)	(0.000)	(0.000)	(0.000)	(0.000)
stdAvgEdu_x	0.437	0.227	0.026	-0.208				
	(0.348)	(0.411)	(0.314)	(0.335)				
Skillintensity_i					-2.034	-2.234	0.126	0.003
					(1.509)	(1.448)	(1.240)	(1.315)
Skillintensity_x*i					0.004	-0.010	-0.032	-0.029
					(0.042)	(0.043)	(0.051)	(0.049)
Religion_xm	0.001 ***	0.001 ***	0.001 ***	0.001 ***	0.002 ***	0.002 ***	0.002 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GATT_WTO_xm	-0.050	-0.052	-0.093	-0.086	-0.057	-0.043	0.047	0.056
	(0.174)	(0.182)	(0.179)	(0.180)	(0.143)	(0.144)	(0.150)	(0.159)
Importer-Industry FE					Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter-Importer-Industry FE	Yes	Yes	Yes	Yes				
Number of observations	14978	14978	11424	11424	7036	7036	5230	5230

Notes: The dependent variable in each fixed effects Poisson regression is the three-year average exports (divided by [GDP of exporter *x* * GDP of importer *m*]) from exporter *x* to importer *m* in industry *i*. Bootstrap standard errors with 300 replications, clustered by exporter-importer-industry pair in (1)-(4), or clustered by importer-industry pair in (5)-(8) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Appendix I. Regression Tables

Table A.1. Estimated coefficients of the three match indices (Selection corrected estimates I with Poisson regression)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, Manu
MatchIndex_x*i (CV)	0.163 **	0.313 **	0.174 *	-0.017	0.291 **	0.202
	(0.075)	(0.160)	(0.099)	(0.126)	(0.145)	(0.187)
MatchIndex_x*i (Gini)	0.176 **	0.271	0.167 *	-0.096	0.087	0.165
	(0.076)	(0.167)	(0.088)	(0.108)	(0.199)	(0.212)
MatchIndex_x*i (MID)	0.173 **	0.279 *	0.167 **	-0.083	0.071	0.149
	(0.074)	(0.170)	(0.083)	(0.108)	(0.204)	(0.198)

Notes: The dependent variable in each fixed effects Poisson regression is the three-year average exports (divided by [GDP of exporter *x* * GDP of importer *m*]) from exporter *x* to importer *m* in industry *i*. Bootstrap standard errors with 300 replications, clustered by importer-industry pair in (1), (3), (5), and (6), or clustered by exporter-importer-industry pair in (2) and (4) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table A.2 Estimated coefficients of the three match indices (Selection corrected estimates II proposed by Wooldridge, 1995)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Excluded variables	Language_xm	Religion_xm	Religion_xm	Language_xm	Religion_xm	Religion_xm	Religion_xm	Religion_xm
Sample	NonServ	NonServ	NonServ	Manu	Manu	Manu	India, NonServ	India, Manu
MatchIndex_x*i (CV)	0.326 ***	0.324 ***	0.315 ***	0.103 *	0.111 **	0.248 ***	0.211 ***	0.247 **
	(0.031)	(0.031)	(0.050)	(0.055)	(0.052)	(0.062)	(0.081)	(0.098)
MatchIndex_x*i (Gini)	0.318 ***	0.316 ***	0.294 ***	0.103 **	0.111 **	0.190 ***	0.083	0.245 **
	(0.031)	(0.030)	(0.049)	(0.050)	(0.050)	(0.056)	(0.096)	(0.112)
MatchIndex_x*i (MID)	0.316 ***	0.314 ***	0.288 ***	0.106 **	0.114 **	0.203 ***	0.081	0.234 **
	(0.030)	(0.030)	(0.051)	(0.052)	(0.048)	(0.056)	(0.091)	(0.113)

Notes: This table reports the estimation results from the panel selection model proposed by Wooldridge (1995), Procedure 3.1 and 4.1.1. The dependent variable in each regression is the logarithm of three-year average exports (divided by [GDP of exporter *x* * GDP of importer *m*]) from exporter *x* to importer *m* in industry *i*. Bootstrap standard errors with 300 replications, clustered by importer-industry pair in (1), (2), (4), (5), (7), and (8), or clustered by exporter-importer-industry pair in (3) and (6) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Table A.3 Estimated coefficients of the three match indices (Infrastructure match index added in Poisson regression)

Match Index (Skill)	Added Control variables / Sample	(1) Powerloss	(2) Road	(3) Powerloss	(4) Road	(5) Powerloss	(6) Road	(7) Powerloss	(8) Road
	Sample	NonServ	NonServ	Manu	Manu	India, NonServ	India, NonServ	India, Manu	India, Manu
CV	MatchIndex_x*i (CV)	0.604 ***	0.476 **	0.519 ***	0.296 *	0.215	0.136	0.287	0.277
		(0.208)	(0.187)	(0.194)	(0.170)	(0.147)	(0.140)	(0.178)	(0.190)
	MatchIndex_x*i (Powerloss)	-0.559 ***		-0.682 ***		-0.544 ***		-0.319 *	
		(0.132)		(0.194)		(0.179)		(0.189)	
	MatchIndex_x*i (Road)		0.605 ***		0.489 **		0.584 ***		0.319 *
			(0.170)		(0.201)		(0.122)		(0.175)
Gini	MatchIndex_x*i (Gini)	0.640 ***	0.460 **	0.426 *	0.124	0.128	0.149	0.295	0.308
		(0.233)	(0.192)	(0.232)	(0.162)	(0.175)	(0.170)	(0.235)	(0.231)
	MatchIndex_x*i (Powerloss)	-0.677 ***		-0.682 ***		-0.561 ***		-0.342 *	
		(0.185)		(0.251)		(0.171)		(0.197)	
	MatchIndex_x*i (Road)		0.646 ***		0.388 *		0.609 ***		0.362 *
			(0.170)		(0.234)		(0.120)		(0.197)
MID	MatchIndex_x*i (MID)	0.640 ***	0.476 **	0.451 **	0.159	0.100	0.124	0.266	0.278
		(0.240)	(0.201)	(0.207)	(0.168)	(0.170)	(0.172)	(0.218)	(0.194)
	MatchIndex_x*i (Powerloss)	-0.662 ***		-0.690 ***		-0.559 ***		-0.333 *	
		(0.178)		(0.229)		(0.167)		(0.187)	
	MatchIndex_x*i (Road)		0.651 ***		0.412 *		0.608 ***		0.352 *
			(0.179)		(0.226)		(0.113)		(0.194)

Notes: The dependent variable in each fixed effects Poisson regression is the three-year average exports (divided by [GDP of exporter x * GDP of importer m]) from exporter x to importer m in industry i. Bootstrap standard errors with 300 replications, clustered by exporter-importer-industry pair in (1)-(4), or clustered by importer-industry pair in (5)-(8) are reported in parentheses. *, **, and *** indicate 10%, 5%, and 1% significance level, respectively.

Appendix II. Data Sources and Construction of Variables Used in Analysis

Data / Variable	Sources / Construction method
Export	Source: *NBER-United Nations Trade Data, 1962-2000* constructed by Feenstra and Lipsey (see Feenstra et al., 2005). For the construction, see Section 4.1 in the main text. The implicit GDP deflator is calculated from nominal and real GDP of China and India (China: DCS-NBS (2005) until 1977 and NBS (2009) onwards; India: CSO (2007) until 1998 and CSO (2008) onwards.) As exchange rates, the annual average official exchange rates from World Bank (2010) are used. GDPs of importing countries (real 2000 USD prices) are primarily taken from World Bank (2010). For the former USSR, the former Czechoslovakia, the former Yugoslavia, the former German FR, the former Yemen Dm, and the former Yemen AR, GDP data are taken from UN, *National Accounts Main Aggregates Database*, and UN (1993), p.234. GDP data on Taiwan is from IMF (2010).
Skill dispersion indices (CV_{xt} , $Gini_{xt}$, and MID_{xt})	Source: *International Data on Educational Attainment: Updates and Implications* constructed by Barro and Lee (Barro and Lee, 2000). For the construction of index, see Section 4.2.1 in the main text.
Index for the length of industry production chains ($Leontief_{xit}$)	Sources: For China, *Asian International I/O Table* of 1985, 1990, 1995, and 2000, constructed by IDE-JETRO. For India, *Input-Output Transaction Table* of 1983-84, 1989-90, 1993-94, and 1998-99 (CSO, 1990, 1997, 2000, 2005). $Leontief_{xit}$, which is the column sum of the Leontief inverse coefficient of each industry, is computed from the input-output (IO) tables of China and India as follows: China: The original IO tables distinguish domestic and imported inputs as in the transaction table below. Let subscripts *k* and *i* denote row and column respectively.

<table>
<tr><td colspan="2">i
k</td><td>1</td><td>2</td><td>..</td><td>n</td><td>Final demand</td><td>Import</td><td>Output</td></tr>
<tr><td rowspan="4">Domestic inputs</td><td>1</td><td>x_{11}^{d}</td><td>x_{12}^{d}</td><td>..</td><td>x_{1n}^{d}</td><td>F_1^d</td><td>-</td><td>X_1</td></tr>
<tr><td>2</td><td>x_{21}^{d}</td><td>x_{22}^{d}</td><td>..</td><td>x_{2n}^{d}</td><td>F_2^d</td><td>-</td><td>X_2</td></tr>
<tr><td>:</td><td>:</td><td>:</td><td>:</td><td>:</td><td>:</td><td>-:</td><td>:</td></tr>
<tr><td>n</td><td>x_{n1}^{d}</td><td>x_{n2}^{d}</td><td>..</td><td>x_{nn}^{d}</td><td>F_n^d</td><td>-</td><td>X_n</td></tr>
<tr><td colspan="2">Imported inputs</td><td>X_1^m</td><td>X_2^m</td><td>..</td><td>X_n^m</td><td>F^m</td><td>$-M$</td><td>-</td></tr>
<tr><td colspan="2">Value added</td><td>V_1</td><td>V_2</td><td>..</td><td>V_n</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Output</td><td>X_1</td><td>X_2</td><td>..</td><td>X_n</td><td></td><td></td><td></td></tr>
</table>

1) Input coefficient matrix A_d is computed for domestic inputs, as follows:

$$A_d = \begin{pmatrix} a^d_{11} & a^d_{12} & .. & a^d_{1n} \\ a^d_{21} & a^d_{22} & .. & a^d_{2n} \\ : & : & : & : \\ a^d_{n1} & a^d_{n2} & .. & a^d_{nn} \end{pmatrix}, \text{ where } a^d_{ki} = x^d_{ki} / X_i .$$

2) $Leontief_{xit} = \sum_k l_{ki}$, where l_{ki} is the Leontief inverse coefficient, which is computed as follows:

$$\begin{pmatrix} l_{11} & l_{12} & .. & l_{1n} \\ l_{21} & l_{22} & .. & l_{2n} \\ : & : & : & : \\ l_{n1} & l_{n2} & .. & l_{nn} \end{pmatrix} = L = (I - A_d)^{-1}, \text{ where } I \text{ is the identity matrix.}$$

India: Since the original IO tables do not distinguish domestic and imported inputs, input coefficient matrix A_d is computed by using import flow matrix (commodity-by-industry U_m table), U table (commodity-by-industry Use table), V table (industry-by-commodity Make table), and X table (commodity-by-commodity transaction table) which are interrelated as in the table below.

	Commodities	Industries	Final demand	Output
Commodities	X table $(= X_d + X_m)$	U table $(= U_d + U_m)$	e	q
Industries	V table			g
Final demand		y'		
Output	q'	g'		

Note: Subscripts, *d* and *m* indicate domestic and imported inputs, respectively.

1) For the IO data of 1989-90 and 1993-94, in which import flow matrix (U_m) is available, input coefficient matrix A_d is computed as follows:

$$A_d = B_d D,$$

where $B_d = U_d \hat{g}^{-1}$, which is the input coefficient matrix calculated from U_d table ($= U - U_m$), and $\hat{g}$ indicates the diagonal matrix with vector g as the diagonal elements; and

$D = V\hat{q}^{-1}$, which is the matrix of the proportions in which industries produce total output of a particular commodity, calculated from V table, and $\hat{q}$ indicates the diagonal matrix with vector q as the diagonal elements.

2) For the IO data of 1983-84 and 1998-99, import flow matrices are estimated since they are not available. First, X table for imported inputs, X_m for 1989-90 and 1993-94 is constructed respectively as [X- X_d], where $X_d = A_d X$. Then, X_m table for 1983-84 [1998-99] is estimated by assuming that the share of imported inputs

	for each column *i* in the total import is unchanged from 1989-90 [1993-94] for all rows. Then, A_d for 1983-84 and 1998-99 is calculated from $X_d (= X - X_m)$. 3) Finally, $Leontief_{xit}$ is computed similarly to the case of China.
MatchIndex_*x*i* (Skill)	For the construction of index, see Section 4.2.3 in the main text.
MatchIndex_*iv*	For the construction of index, see Section 5.2.1 in the main text. School enrollment data (1964-66, 1969-71, 1974-76, 1979-81) are from DCS-NBS (2005), p.80 for China; MHRD and NIC, *Compilation on 50 years of Indian Education: 1947-1997* and MHRD (2008), for India. Population data are from World Bank (2010). $Leontief_{xit}$ of Thailand is computed from the IO tables of Thailand taken from IDE-JETRO's *Asian International I/O Table* of 1985, 1990, 1995, and 2000.
ImportRatio_*i*	Sources: The same as $Leontief_{xit}$. ImportRatio_*i* is calculated for each industry as the percentage of the value of imported inputs to the value of total inputs, using the IO tables of China and India.
Kintensity_*x*	Capital intensity of China and India is the capital stock divided by GDP times 100(%). Both capital stock and GDP are measured in terms of real 2000 prices and three-year average (1983-85, 1988-90, 1993-95, 1998-2000). Capital stock data are from: Holtz (2006), p.170, Table 6, BC3, for China, and CSO (2007) for India. Regarding the sources for GDP data, see the above explanation on Export.
Kintensity_*x*i*	Interaction of Kintensity_*x* and capital intensity of industry *i* (Kintensity_*i*), which is defined as the percentage of capital stock to the industry's gross value added (GVA). Both capital stock and GVA are measured in terms of 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 capital stock deflator computed from Holz (2006), p.178, Table 8 (columns (2)/(4)) and implicit GDP deflator (see the explanation on Export) are applied. 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, 2002; ASI website of Ministry of Statistics and Programme Implementation, Government of India), which cover only registered manufacturing are used. Implicit capital stock (fixed asset) deflator computed from the net fixed capital stock (NFCS) data on manufacturing industry from CSO (2007, 2008) is applied. Similarly, implicit GVA deflator is computed from CSO (2007, 2008) and applied. For the non-manufacturing industries, NFCS and GDP data from CSO (2007, 2008) are used. Kintensity_*i* of agriculture is applied for three industries (No. 1, 2, and 3), and that of mining and quarrying is applied for two industries (No. 6 and 7), due to broad industry classification of NFCS and GDP data.

stdAvgEdu_*x*	Source: Barro and Lee (2000). Average years of education of the population over age 15 for China and India are taken from Barro and Lee (2000) and have been standardized.
Skillintensity_i (only for India)	Source: NSSO, *Unit-level data of National Sample Survey, Employment and Unemployment schedule, 1983, 1987-88, 1993-94, 1999-00*. Skill intensity of industry, defined as the ratio of working population with post-secondary level education, is constructed by using the unit-level data of the four-round Employment and Unemployment schedules of the National Sample Surveys. A person is considered working based on the usual principal activity status.
Skillintensity_*x*i* (only for India)	Interaction of Skill intensity_*i* and skill intensity of exporting country, which is defined as the ratio of population over age 15 with post-secondary level education. This ratio is calculated from Barro and Lee (2000).
ln(Distance)_*xm*	Source: CEPII's data (see Mayer and Zignago, 2006). Logarithm of the distance between the capital cities of exporting and importing countries.
Contiguous_*xm*	Source: CEPII's data (see Mayer and Zignago, 2006). A binary variable that equals one if exporting and importing countries are contiguous, and zero otherwise.
Legalsystem_*xm*	Source: Data from Helpman et al. (2008). A binary variable that equals one if exporting and importing countries share the same legal origin, and zero otherwise.
Colonial ties_*xm*	Source: Data from Helpman et al. (2008). A binary variable that equals one if importing country ever colonized exporting country or vice versa, and zero otherwise.
Language_*xm*	Source: CEPII's data (see Mayer and Zignago, 2006). A binary variable that equals one if a language is spoken by at least 9% of the population in both exporting and importing countries.
Religion_*xm*	Source: "Religion Adherence Data" constructed by Barro and McCleary (2005). Index for the degree of shared religion 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.
GATT_WTO_*xm*	Source: Data from Helpman et al. (2008) and WTO website from 1990 onwards. Number of exporter/importer who are members of GATT or WTO.

MatchIndex_*x***i* (Powerloss)	Source: World Bank (2010). For the construction of index, see Section 5.2.3 in the main text. As a measure of Powerloss, figures for electric power transmission and distribution losses (as percentage of output) of China and India are used.
MatchIndex_*x***i* (Road)	Source: World Bank (2010), and Ghosh and De (2005), p.754, Table 6.11, for the road length of India until 1995. For construction of the index, see Section 5.2.3 in the main text. As a measure of Road, figures for road density (kilometers of total road network divided by 100 sq. km of land area) of China or India are used.