Online Appendix for “How Does Skill Distribution Shape Comparative Advantage Across Industries? Theory and Evidence”

A Additional Robustness Checks

A.1 Policy-Driven Initial Conditions

A potential threat to our identification arises from the possibility that pre-existing differences across provinces and industries, especially those induced by government interventions, may be driving our results. Of particular concern is the impact of Deng Xiaoping’s southern tour in 1992, which catalyzed reforms to further open China’s economy to the global market. In the wake of this tour, there was a substantial surge in state-directed investments in specific coastal areas, potentially creating systematic differences in industrial composition and economic conditions across provinces. Even though our baseline specification includes a full suite of province and industry fixed effects, we may still worry that these province-industry-specific initial conditions may be correlated with regional trade patterns and thus confound our estimates.

To evaluate the influence of policy-driven initial conditions, we control for a comprehensive set of province-industry-level variables at key time points: 1990, 1992, 2000, and 2010. These controls fall into four broad categories: (i) enterprise scale indicators, (ii) asset investment indicators, (iii) production and operation indicators, and (iv) profitability indicators. Specifically, we include:

- Enterprise scale indicators: number of industrial enterprises, number of loss-making enterprises, and number of employees at year-end.
- Asset investment indicators: fixed asset investment, original value of fixed assets, net value of fixed assets, and year-end working capital occupation.
- Production and operation indicators: gross industrial output value, net industrial output value, and product sales revenue.
- Profitability indicators: total losses of loss-making enterprises, total profit, total profit and tax, and enterprise retained profits.

Table A1 and Table A2 present the results of repeating our baseline specification augmented with each set of initial condition variables. Across all specifications, our main coefficients of interest remain statistically significant at conventional levels, with point estimates closely aligned with our baseline results. This provides strong evidence that our findings are not driven by pre-existing differences in industrial composition or economic conditions across provinces that may have been influenced by government policies prior to our study period.

A.2 Firm Age Structure

Recent literature has highlighted the influence of industry dynamics and firm age structure on wage dispersion patterns. For instance, rapidly expanding conglomerates in advanced technology sectors may offer substantial promotion opportunities, potentially inducing new recruits to accept lower initial wages in anticipation of future career advancements (Li *et al.*, 2018). Conversely, declining industries with limited growth prospects might exhibit compressed wage structures across tenure levels. These industry-specific characteristics could plausibly affect wage distributions through mechanisms not explicitly accounted for in our baseline specification.

We start our analysis by examining whether our observed patterns of wage dispersion primarily reflect the underlying differences in industry life cycles. To this end, we utilize data from the Chinese Annual Survey of Industrial Firms (ASIF) spanning 1998-2007. For each industry-year observation, we compute the median firm age as a proxy for industry maturity and growth prospects. We then estimate regressions of industrial wage dispersion against these industry-level firm age measures. Table A3 reports the estimated coefficients for each survey year. Our analysis reveals no consistently significant correlation between industrial wage dispersion and firm age structure, with the exception of years 2003 and 2004. Interestingly, despite a lack of statistical significance, the positive coefficients tend to suggest that newer industries are more likely to have a less dispersed distribution of wages. This observation runs contrary to the hypothesis that newer industries offer more promotion opportunities in exchange for low entry salaries, which would typically result in higher wage dispersion.

The results in Table A3 suggest that our measures of skill substitutability, proxied by wage dispersion, are not significantly confounded by industry maturity. However, we cannot definitively rule out the possibility that the age structure of firms within each industry may affect regional trade patterns. To further investigate the potential influence of firm age compositions, we expand our baseline specification to include an interaction term between relative skill dispersion and industry-specific age ranking. Table A4 presents the estimates from this augmented model. Each column in the table corresponds to the firm age ranking for individual years from 1998 to 2007. As shown, the interaction between skill substitutability and relative skill dispersion remains highly significant and has the expected sign. Moreover, the magnitude of the coefficient is comparable to our baseline estimates. Overall, taking industry-specific firm age structure into account, the estimated coefficients do not significantly differ from the benchmark results.

A.3 Alternative Clustering of Standard Errors

Our primary analysis clusters standard errors by exporter-importer pair, which aligns with our focus on trade patterns specific to these bilateral relationships. However, given the possibility of correlations in trade flows within industries, we also investigate the robustness of our results by clustering standard errors at the industry level.

The results are reported in Table A5. In Columns 1 and 4, we replicate our main analysis with standard errors clustered at the industry level. By doing so, we account for potential within-industry correlation across different exporter-importer pairs. We find that the coefficient on the interaction between relative skill dispersion and skill substitutability remains statistically significant, regardless of the measure of skill dispersion used. This indicates that our key findings are robust even when allowing for correlation of observations within industries.

However, clustering standard errors by industry may lead to biased inference due to the limited number of clusters. To alleviate this concern, we employ the wild cluster bootstrap method proposed by Cameron *et al.* (2008). As indicated by the empirical p-values in Columns 2 and 5, the key point estimates are highly significant even under this more conservative inference approach. We also implement two-way clustering by both exporter-importer pair and industry, following Cameron *et al.* (2011). The results are presented in Columns 3 and 6. Reassuringly, the statistical significance of the estimated coefficient of interest is retained.

B Extension of the Theoretical Model: Trade in Tasks and Firm Heterogeneity

Our baseline model implicitly assumes that all productive tasks within an industry occur in the same location. However, modern global value chains are highly fragmented and operate through geographically dispersed production processes. This naturally raises legitimate concerns about the applicability of our model to industries with internationally distributed production networks. Consider, for instance, the aerospace manufacturing sector. The production of a commercial aircraft, such as the Boeing 787 Dreamliner, involves components sourced from numerous suppliers across various countries. This global sourcing strategy may potentially weaken the relationship between a country's skill distribution and its comparative advantage in the final product. Specifically, the skill complementarity we identify as crucial for the aerospace industry may not apply uniformly across all stages of aircraft production.

To verify the robustness of our main results, we extend our theoretical framework to incorporate two key features of modern international trade: the concept of "trade in tasks" (Grossman and Rossi-Hansberg, 2008; Ottaviano *et al.*, 2015) and firm heterogeneity (Melitz, 2003). In this extended model, we preserve the consumer's utility maximization problem from the baseline setup. This yields the optimal quantity demanded of good ω^s from country i for consumers in country j, consistent with our original formulation:

$$q_{ij}^s(\omega^s) = p_{ij}^s\,(\omega^s)^{-\sigma^s}\, X_j^s P_j^{s\,\sigma^s-1}$$

On the supply side, we conceptualize production as a continuum of tasks that firms can perform domestically or offshore. Tasks in sector s are indexed by $z^s \in [0,1]$ and ordered such that offshoring costs are nondecreasing. The offshoring cost for task z^s is modeled as $\beta t_{ij}^s(z^s) a_L^s$, where β is an offshoring technology parameter, $t_{ij}^s(z^s)$ represents task-specific offshoring cost, and a_L^s is the total domestic labor required to produce one unit of output without offshoring. Firms decide whether to offshore tasks based on the condition: $w_i > w_j \beta t_{ij}^s(0)$, where w_i and w_j are the wage rates in countries i and j, respectively. The marginal task performed domestically, z_*^s, is determined by the condition that wage savings equal offshoring cost: $w_i = w_j \beta t_{ij}^s(z_*^s)$.

We model firm productivity $T_i^s(\omega^s)$ in sector s of exporting country i as a combination of workers' skills according to a CES production function:

$$T_i^s(\omega^s) = \xi(\omega^s)\left(\int_{t\in\mathcal{T}} t^{\delta^s} dS_i(t)\right)^{1/\delta^s}$$

where $\delta^s \in (0,1)$, $1/(1-\delta^s)$ is the elasticity of substitution across workers' skills, t is the skill level of employed workers, $S_i(t)$ is the skill distribution of country i, and $\xi(\omega^s) > 0$ is a firm-specific random productivity shock.

The total cost for a firm with efficiency $T_i^s(\omega^s)$ to deliver its products to destination country j is:

$$c_{ij}^s(\omega^s) = f_{ij}^s + v_{ij}^s \tau_{ij}^s \frac{q_{ij}^s(\omega^s)}{T_i^s(\omega^s)}$$

where f_{ij}^s is the fixed cost of entry into the export market, $v_{ij}^s \tau_{ij}^s$ is the variable marginal cost of exporting with iceberg factor τ_{ij}^s, and $q_{ij}^s(\omega^s)$ is the quantity exported. The variable marginal cost v_{ij}^s is the sum of wages paid to domestic and foreign labor for offshored tasks, given by:

$$v_{ij}^s = w_i a_L^s (1 - z_*^s) + w_j a_L^s \int_0^{z_*^s} \beta t_{ij}^s(z^s)\, dz^s$$

We now determine the equilibrium prices that a representative firm with productivity $T_i^s(\omega^s)$ sets. The optimization problem is:

$$\max_{\left\{p_{ij}^s(\omega^s)\right\}_{j\in\mathcal{N}}} \sum_{j\in\mathcal{N}} \left(p_{ij}^s(\omega^s)\, q_{ij}^s(\omega^s) - f_{ij}^s - v_{ij}^s \tau_{ij}^s \frac{q_{ij}^s(\omega^s)}{T_i^s(\omega^s)}\right)$$

The first order condition implies that a firm from $i \in \mathcal{N}$ with productivity $T_i^s(\omega^s)$ selling to destination j will charge a price:

$$p_{ij}^s(\omega^s) = \frac{\sigma^s}{\sigma^s - 1} \frac{v_{ij}^s \tau_{ij}^s}{T_i^s(\omega^s)}$$

Then we can write the average prices charged by all firms in $i \in \mathcal{N}$ selling to $j \in \mathcal{N}$ as:

$$\int_{\omega^s\in\Omega^s} p_{ij}^s(\omega^s)^{1-\sigma^s} d\omega^s = \left(\frac{\sigma^s}{\sigma^s - 1} \frac{v_{ij}^s \tau_{ij}^s}{\Xi^s \tilde{\mu}}\right)^{1-\sigma^s} \exp\left(\frac{\sigma^s - 1}{2}(\delta^s - 1)\,\theta_i\right)$$

where $\Xi^s = \left(\int_{\omega^s\in\Omega^s} \xi(\omega^s)^{\sigma^s - 1} d\omega^s\right)^{1/(\sigma^s - 1)}$ captures the aggregation of random productivity shocks of sector s, and $\tilde{\mu}$ and θ_i are parameters related to the skill distribution defined in the main text.

Substituting the expression of average prices into the equation for export flows, we obtain

$$X_{ij}^{s}=\alpha_{j}^{s}X_{j}\left(1+\tau_{ij}^{s\,\sigma^{s}-1}\left(\frac{w_{i}}{w_{j}}\left(1-z_{*}^{s}\right)+\int_{0}^{z_{*}^{s}}\beta t_{ij}^{s}\left(z^{s}\right)dz^{s}\right)^{\sigma^{s}-1}\exp\left(\frac{1-\sigma^{s}}{2}\left(\delta^{s}-1\right)\Delta_{ij}\right)\right)^{-1}$$

which closely resembles the expression for export flows in our baseline model.

The extended model, while accounting for the fragmentation of production processes and firm heterogeneity, preserves our key prediction: countries with relatively more dispersed skill distributions export more in sectors characterized by higher degrees of skill substitutability across workers.

References

[1] Cameron, A.C., Gelbach, J.B., and Miller, D.L. 2008. "Bootstrap-based improvements for inference with clustered errors", *Review of Economics and Statistics*, 90(3): 414-427.

[2] Cameron, A.C., Gelbach, J.B., and Miller, D.L. 2011. "Robust inference with multiway clustering", *Journal of Business & Economic Statistics*, 29: 238-249.

[3] Grossman, G.M., and Rossi-Hansberg, E. 2008. "Trading tasks: A simple theory of offshoring", *American Economic Review*, 98(5): 1978-1997.

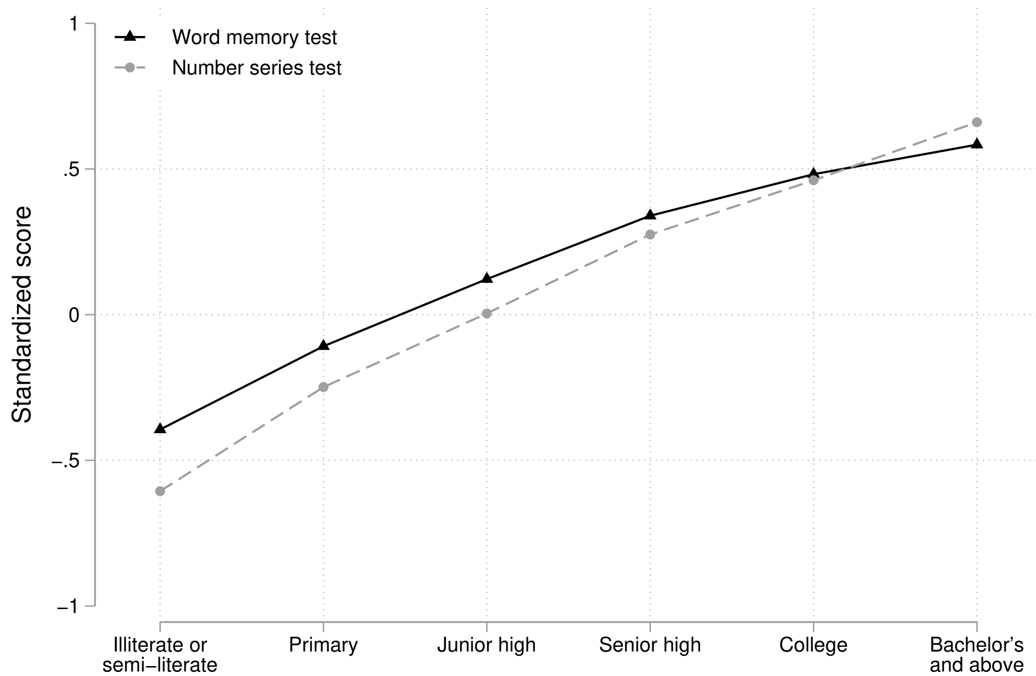
[4] Li, J., Powell, M., and Ke, R. 2018. "Firm growth and promotion opportunities", Mimeo.

[5] Melitz, M.J. 2003. "The impact of trade on intra-industry reallocations and aggregate industry productivity". *Econometrica*, 71(6), 1695-1725.

[6] Ottaviano, G.I.P., Peri, G., and Wright, G.C. 2015. "Immigration, trade and productivity in services: Evidence from U.K. firms", NBER Working Paper.

C Additional Figures

Figure A1 Correlations between Standardized Scores for Word Memory and Number Series Tests and Education Level



Notes: This figure shows the mean performance on word memory and number series tests across educational attainment levels. All test scores are standardized by age.

D Additional Tables

Table A1 Accounting for Region-Specific Initial Industrial Conditions: Enterprise Scale and Asset Investment Indicators

	Dependent variable: Log $export_{pcg}$						
	Enterprise scale indicators			Asset investment indicators			
	Number of industrial enterprises	Number of loss-making enterprises	Number of employees at year-end	Fixed asset investment	Original value of fixed assets	Net value of fixed assets	Year-end working capital occupation
A. Substitutability proxied by wage dispersion rankings:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Substitutability_g$	0.015***	0.015***	0.015***	0.015***	0.014***	0.015***	0.016***
× Relative skill $dispersion_{pc}$	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Region specific industrial initial condition	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9038	9038	9038	9038	9038	9038	9038
R^2	0.670	0.669	0.673	0.669	0.672	0.672	0.675
*B. Substitutability proxied by O*NET rankings:*							
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$Substitutability_g$	-0.014***	-0.014***	-0.015***	-0.015***	-0.014***	-0.014***	-0.014***
× Relative skill $dispersion_{pc}$	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Region specific industrial initial condition	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9038	9038	9038	9038	9038	9038	9038
R^2	0.670	0.669	0.673	0.669	0.672	0.672	0.675

Notes: The dependent variable is the natural logarithm of exports from province p to country c in industry g. Panels A and B present results using alternative proxies for skill substitutability. Panel A employs the standard deviation of residual wages, while Panel B uses the aggregate index of O*NET rankings. Standardized beta coefficients are reported. Bootstrap standard errors clustered by exporter-importer pair in parentheses (50 replications).

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table A2 Accounting for Region-Specific Initial Industrial Conditions: Production and Operation Indicators, and Profitability Indicators

	Dependent variable: Log export_{pcg}						
	Production and operation indicators			Profitability indicators			
	Gross industrial output value	Net industrial output value	Product sales revenue	Total losses of loss-making enterprises	Total profit	Total profit and tax	Enterprise retained profits
A. Substitutability proxied by wage dispersion rankings:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Substitutability}_g$	0.016***	0.016***	0.016***	0.015***	0.014***	0.016***	0.013**
× Relative skill dispersion_{pc}	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Region specific industrial initial condition	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9038	9038	9038	9038	9038	9038	9038
R^2	0.673	0.675	0.674	0.670	0.669	0.672	0.669
*B. Substitutability proxied by O*NET rankings:*							
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$\text{Substitutability}_g$	-0.015***	-0.016***	-0.015***	-0.013***	-0.016***	-0.017***	-0.016***
× Relative skill dispersion_{pc}	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Region specific industrial initial condition	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9038	9038	9038	9038	9038	9038	9038
R^2	0.673	0.675	0.674	0.669	0.669	0.672	0.669

Notes: The dependent variable is the natural logarithm of exports from province p to country c in industry g. Panels A and B present results using alternative proxies for skill substitutability. Panel A employs the standard deviation of residual wages, while Panel B uses the aggregate index of O*NET rankings. Standardized beta coefficients are reported. Bootstrap standard errors clustered by exporter-importer pair in parentheses (50 replications).

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table A3 Relationship between Industrial Skill Substitutability and Firm Age Structure (By Year)

Year	Marginal effect	Year	Marginal effect
1998	0.002	2003	0.004*
	(0.001)		(0.002)
1999	0.001	2004	0.005*
	(0.001)		(0.002)
2000	0.001	2005	0.002
	(0.001)		(0.003)
2001	0.001	2006	0.001
	(0.001)		(0.004)
2002	0.001	2007	0.001
	(0.001)		(0.004)

Notes: The dependent variable is the degree of industrial wage dispersion. The independent variable is the median firm age within each industry. We report the estimated coefficient from a separate cross-sectional regression for the indicated year. Standard errors are in parentheses.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table A4 Accounting for Firm Age Structure at the Industry Level (By Year)

	Dependent variable: Log export$_{pcg}$									
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A. Substitutability proxied by wage dispersion rankings:										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Substitutability$_g$	0.019***	0.019***	0.018***	0.018***	0.019***	0.021***	0.025***	0.021***	0.021***	0.022***
$\times$ Relative skill dispersion$_{pc}$	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)
Firm age structure$_g$	-0.003	-0.002	-0.000	-0.001	-0.004	-0.008	-0.012*	-0.010	-0.013*	-0.014**
$\times$ Relative skill dispersion$_{pc}$	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11056	11056	11056	11056	11056	11056	11056	11056	11056	11056
R^2	0.661	0.661	0.661	0.661	0.661	0.661	0.662	0.662	0.662	0.662
*B. Substitutability proxied by O*NET rankings:*										
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Substitutability$_g$	-0.017***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***
$\times$ Relative skill dispersion$_{pc}$	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)
Firm age structure$_g$	0.004	0.004	0.007*	0.005	0.006	0.002	0.000	-0.000	-0.003	-0.004
$\times$ Relative skill dispersion$_{pc}$	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11056	11056	11056	11056	11056	11056	11056	11056	11056	11056
R^2	0.662	0.662	0.662	0.662	0.661	0.661	0.661	0.661	0.661	0.661

Notes: The dependent variable is the natural logarithm of exports from province p to country c in industry g. Panels A and B present results using alternative proxies for skill substitutability. Panel A employs the standard deviation of residual wages, while Panel B uses the aggregate index of O*NET rankings. Standardized beta coefficients are reported. Bootstrap standard errors clustered by exporter-importer pair in parentheses (50 replications).

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table A5 Alternative Clustering of Standard Errors

	Dependent variable: Log export$_{pcg}$					
	Wage dispersion rankings			O*NET rankings		
	Industry-level clustering	Wild cluster bootstrap	Two-way clustering	Industry-level clustering	Wild cluster bootstrap	Two-way clustering
	(1)	(2)	(3)	(4)	(5)	(6)
Substitutability$_g$	0.018**	0.018**	0.018*	-0.018**	-0.018**	-0.018**
$\times$ ΔSkill dispersion$_{pc}$	(0.008)	[0.046]	(0.009)	(0.009)	[0.045]	(0.009)
Exporter FE	Yes	Yes	Yes	Yes	Yes	Yes
Importer FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11056	11056	11056	11056	11056	11056
R^2	0.661	0.661	0.661	0.661	0.661	0.661

Notes: The dependent variable is the natural logarithm of exports from province p to country c in industry g. The degree of skill dispersion is measured by the standard deviation of residual scores. Substitutability in Columns 1-3 is proxied by the standard deviation of residual wage dispersion, while substitutability in Columns 4-6 is proxied by the aggregate index of O*NET rankings. Standardized beta coefficients are reported. In Columns 1 and 4, bootstrap standard errors clustered at the industry level in parentheses (50 replications). In Columns 2 and 5, p-values obtained via wild bootstrap as in Cameron et al. (2008) with clustering at the industry level, based on 999 repetitions, appear in brackets. In Columns 3 and 6, two-way clustering by both exporter-importer pair and industry to account for potential correlation along both dimensions in parentheses.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.