Supplemental Appendix:
The Dynamics of Global Sourcing

Trang Hoang

March 2026

This document presents omitted details from the primary manuscript. Appendix A provides additional firm-level statistics, and Appendix B documents extra estimation results.

Data Appendix

Table A1 provides descriptive statistics for the sample of chemical producers and Table A2 lists all the countries used in the baseline analysis.

Table A1—: Firm-Level Summary Statistics

	Mean	Std. dev.	Min.	Max.
Log domestic revenues	10.534	1.734	-0.169	16.341
Log import values	7.627	3.284	-4.794	15.201
Log export values	8.223	2.199	-4.110	13.722
Import status	0.173	0.378	0	1
Export status	0.312	0.463	0	1
Number of import markets	0.582	1.873	0	23
Number of export markets	2.764	6.523	0	71
State-owned	0.164	0.370	0	1
Private	0.327	0.469	0	1
Foreign	0.115	0.319	0	1
Joint venture	0.205	0.404	0	1

Note: This table provides the firm-year-level descriptive statistics for the main sample. Import values capture the average total values that a firm imports in a year. A firm's import status takes the value of one if a firms imports in that year from any country. The export variables are defined in the same way.

Tables A3-A6 replicate the stylized facts using all firms reported in the customs data. The results demonstrate that the patterns observed in my matched sample extend to the broader universe of Chinese importers.

Table A2—: Country List

Australia	Germany	Malaysia	South Africa
Austria	Hong Kong	Mexico	South Korea
Belgium	Hungary	Netherlands	Spain
Brazil	India	New Zealand	Sweden
Canada	Indonesia	Norway	Switzerland
Chile	Iran	Philippines	Taiwan
Czech Republic	Ireland	Poland	Thailand
Denmark	Israel	Russia	Turkey
Finland	Italy	Saudi Arabia	Ukraine
France	Japan	Singapore	United Kingdom
			United States

Note: This table lists the source countries used in my analysis. Countries are ranked by the total number of importers from 2000 to 2006 and the top 40 are included. (Forty-one countries are included in the list due to a tie.)

Table A3—: Country-Specific Transition Rates - All Firms

Status in year $t-1$	No imports		Imports	
Status in year t	No imports	Imports	No imports	Imports
2000–2001	99.81	0.19	40.93	59.07
2001–2002	99.75	0.25	40.58	59.42
2002–2003	99.71	0.29	39.60	60.40
2003–2004	99.66	0.34	41.07	58.93
2004–2005	99.68	0.32	43.64	56.36
2005–2006	99.57	0.43	40.97	59.03

Note: This table summarizes the transition rates into and out of the average import source. Each column describes the transition from the importing status in a country year $t-1$ to the status in the same country in year t. For example, the last column shows the share of firm-country pairs that have positive import values in year t, among the pairs that also have positive import values in the previous year. The first row shows the transition rates between 2000 and 2001, and the second row shows these rates between 2001 and 2002, and so on.
Source: General Administration of Customs.

Estimation Appendix

B1. Persistence in Import Status - Robustness Check

Table B1 below presents the results of estimating the dynamic discrete choice model in Section 1 using the random effects Probit method proposed by Wooldridge (2005). These columns correspond to columns 2-5 in Table 2 in the main text. For ease of comparison with the Arellano-Bond (1991) GMM estimator, I report the average partial effects (APE).

As can be seen from the two tables, the RE probit APE estimates are very similar to the GMM coefficients. They show that there is substantial persistence at the country-specific level (from Column 4 - preferred specification, Probit APE = 0.24, GMM coefficient = 0.21), suggesting evidence of country-specific sunk cost of importing. The coefficient on global import status, however,

Table A4—: Persistence in Import Status - All Firms

	(1)	(2)	(3)	(4)
Import from country j	0.55	0.33	0.33	0.42
in year $t-1$	(0.0004)	(0.001)	(0.001)	(0.001)
Import in year $t-1$				0.007
				(0.00002)
Observations	201,920,850	168,267,375	168,267,375	201,920,850
Country Dummies	Yes	Yes	Yes	Yes
Firm-Country FE		Yes	Yes	Yes
Year Dummies			Yes	Yes

Note: This table reports results on regressing current import status on past import status at the firm-country level. Columns 2 to 4 account for firm-country unobserved heterogeneity using the Arellano-Bond GMM estimator. In the last column, both country-specific and global import status terms are treated as endogenous variables. Robust standard errors are in parentheses.
Source: General Administration of Customs

Table A5—: Country Ranking and Number of Importers - All firms

Country	Rank	2000	2006
Japan	1	12824	30204
United States	2	10999	27367
Taiwan	3	9212	21044
Germany	4	8239	20633
South Korea	5	7993	18841
Hong Kong	6	6307	13851
Italy	7	4660	11632
United Kingdom	8	4436	9946
France	9	4104	8680
Singapore	10	3682	7749

Note: This table presents the top 10 source countries based on the number of unique importers in 2000 and 2006.
Source: General Administration of Customs

remains negligible (Probit APE = 0.002, GMM coefficient = 0.01).

B2. *Alternative Procedure to Predict Sourcing Potential S_{ijt}*

Instead of predicting S_{ijt} through two steps as described in Section IV in the main text, I propose a different procedure to back out S_{ijt} directly through the imported input share χ_{ijt}/χ_{iht}.

I maintain the assumption that $S_{iht} = S_{ht}$–i.e., the domestic sourcing potential is constant across firms but can vary across years. Additionally, S_{ht} is mean independent of S_{ijt}–i.e., $\mathbb{E}(S_{ht}|S_{ijt}) = \mathbb{E}(S_{ht})$. As before, I assume there may be a multiplicative measurement error in the share of imported input over total inputs χ_{ijt}, denoted by ϵ^{x}_{ijt}. We can also assume there is a multiplicative

Table A6—: Share of Chinese firms importing from strings of top 10 countries - All firms

	2000		2006	
	Data	**Random entry**	**Data**	**Random entry**
1	13.34	4.08	13.40	4.61
1-2	1.53	2.47	1.55	2.70
1-2-3	.55	1.15	.47	1.10
1-2-3-4	.24	.45	.21	.41
1-2-3-4-5	.22	.17	.20	.14
1-2-3-4-5-6	.16	.05	.13	.03
1-2-3-4-5-6-7	.15	0	0.09	0
1-2-3-4-5-6-7-8	.15	0	0.11	0
1-2-3-4-5-6-7-8-9	.53	0	0.44	0
1-2-3-4-5-6-7-8-9-10	2.21	0	1.00	0
% following pecking order	19.08	8.38	17.6	8.98

Note: This table presents the shares of firms in a given year that follow a pecking order observed in the data versus predicted by a random entry model. Countries are indexed by their ranks reported in Table **??**. The first row shows the share of firms that import from the top source country (i.e., Japan). The second row shows the share of firms that import only from the top two countries (i.e., Japan and the United States) in a given year. The random entry predictions are based on the assumption that the unconditional probability of importing from country j is the same as the probability of importing from country j conditional on a firm's import decisions in other countries.
Source: General Administration of Customs.

measurement error in the share of domestic inputs χ_{iht}. In that case ϵ^x_{ijt} is treated as the ratio of the two measurement errors.

$$\frac{\chi_{ijt}}{\chi_{iht}} = \frac{S_{ijt}}{S_{ht}}\epsilon^x_{ijt}$$

Next, suppose we run a linear regression of $\log \chi_{ijt} - \log \chi_{iht}$ on the set of independent variables in equation B7:

$$\text{(B1)} \qquad \log \chi_{ijt} - \log \chi_{iht} = \beta_0 + g(X^T_{jt}\beta^T) - \theta h(X^\tau_{ijt}\beta^\tau) - \theta \ln w_{jt} + \lambda_t$$

Under the new specification, the estimated values of the year dummies λ_t will be reduced by $\mathbb{E}(\log S_{ht})$, assuming $\mathbb{E}(\log \epsilon^x_{ijt}) = 0$. If we restrict S_{ht} to be constant across time, then the constant coefficient β_0 is affected. In either case, other coefficient estimates should still be consistent, though the predicted values of $\log S_{ijt}$ will be biased by $\mathbb{E}(\log S_{ht})$.

Because what we want to obtain is the predicted values for S_{ijt}, the log-linearized model may not be ideal as $\ln \mathbb{E}(S_{ijt}) \neq \mathbb{E}(\ln S_{ijt})$. For that reason, I run a Poisson regression

$$\text{(B2)} \qquad \frac{\chi_{ijt}}{\chi_{iht}} = \exp\left(\beta_0 + g(X^T_{jt}\beta^T) - \theta h(X^\tau_{ijt}\beta^\tau) - \theta \ln w_{jt} + \lambda_t\right)$$

Table B1—: Persistence in Import Status - Random Effects Probit

	(1)	(2)	(3)	(4)
Import from country j	0.22	0.22	0.28	0.24
in year $t-1$	(0.01)	(0.03)	(0.01)	(0.03)
Import in year $t-1$				0.002
				(0.0003)
Observations	1934730	1934730	753912	753912
Country Dummies	Yes	Yes	Yes	Yes
Firm-Country FE	Yes	Yes	Yes	Yes
Year Dummies		Yes	Yes	Yes
Domestic sales			Yes	Yes

Note: This table reports the estimated average partial effects from regressing current import status on past import status at the firm-country level using the random effects probit method proposed by Wooldridge (2005). Standard errors are in parentheses.
Note: Source: General Administration of Customs and National Bureau of Statistics of China

In principle, the Poisson regression allows us to include zeros on the left hand side. That said, recall the definition of sourcing potential: $S_{ijt} = T_j(\tau^m_{ijt} w_{jt})^{-\theta}$. This means $S_{ijt} = 0$ if either country-level technology, variable trade costs, or wages is 0. In practice, this scenario seems implausible that any of these terms is actually zero. For this reason, I exclude observations with zero imported inputs. Note that the Poisson regression is still subject to the previous issue with a predicted value of S_{ijt} being biased, now by a scale of $\mathbb{E}(S_{ht})$.

Table B2 reports results for different methods of estimating country-level sourcing potential. The first two columns are the baseline results reported in Section IV. The next two columns report results for equation B1 under a log-linearized model. As expected, except for the year dummies and constant term, the two sets of estimates are identical.

B3. *New Importer Dynamics*

The sunk-cost model (including my baseline model) tends to predict that new importers would immediately trade a lot upon entering a market, contradicting evidence of gradual growth documented in the literature.[1] Previous studies in the export literature have also shown that new exporters grow slowly in new markets (Ruhl & Willis 2017).

To address this issue, I follow Alessandria, Choi, and Ruhl (2021) and incorporate the ramping up effect through the cost of importing. In particular, I impose that the variable trade cost that firm i pays to imports from country j in year t, τ^m_{ijt}, is a function of firm i's import tenure in country j, denoted by W_{ijt}. Let $W_{ijt} = W(d^m_{ijt-a}, ..., d^m_{ijt-1})$, where $a \geq 1$ and $d^m_{ijt} = 1$ if firm i imports from

[1]Gimenez-Peralesy (2024) documents the gradual growth in import share using data on Colombian manufacturers.

Table B2—: Robustness Check - Predicting S_{ijt}

	Residuals		$\log X_j/X_h$		X_j/X_h	
	OLS	IV	OLS	IV	Poisson	IV Poisson
	(1)	(2)	(3)	(4)	(5)	(6)
log wages	-0.299	-1.985	-0.299	-1.985	0.0137	-0.596
	(0.0639)	(0.478)	(0.0639)	(0.478)	(0.0586)	(1.252)
R&D expenditure	-0.0332	0.643	-0.0332	0.643	-0.0505	0.262
	(0.0469)	(0.196)	(0.0469)	(0.196)	(0.0515)	(0.693)
$\log k$	-0.00168	0.00515	-0.00168	0.00515	0.000996	0.00335
	(0.000380)	(0.00196)	(0.000380)	(0.00196)	(0.000371)	(0.00529)
landlocked	-0.576	0.242	-0.576	0.242	-1.070	-0.793
	(0.161)	(0.284)	(0.161)	(0.284)	(0.274)	(0.580)
GDP	0.0692	0.255	0.0692	0.255	0.0257	0.0967
	(0.0145)	(0.0542)	(0.0145)	(0.0542)	(0.0156)	(0.159)
log distance	-0.683	-0.246	-0.683	-0.246	-0.459	-0.304
	(0.0448)	(0.131)	(0.0448)	(0.131)	(0.0421)	(0.346)
2001	0.0610	-0.117	0.152	-0.0263	-0.154	-0.250
	(0.142)	(0.155)	(0.142)	(0.155)	(0.143)	(0.394)
2002	0.0814	-0.200	0.0846	-0.196	0.241	0.0733
	(0.137)	(0.162)	(0.137)	(0.162)	(0.134)	(0.411)
2003	0.259	0.235	0.216	0.193	-0.0937	-0.169
	(0.133)	(0.137)	(0.133)	(0.137)	(0.137)	(0.427)
2004	0.177	0.0287	0.435	0.286	0.545	0.420
	(0.127)	(0.138)	(0.127)	(0.138)	(0.123)	(0.489)
2005	0.112	0.233	0.0916	0.213	-0.604	-0.650
	(0.130)	(0.138)	(0.130)	(0.138)	(0.141)	(0.382)
2006	0.254	0.449	0.361	0.556	0.102	0.0605
	(0.132)	(0.148)	(0.132)	(0.148)	(0.132)	(0.411)
Constant	5.413	4.956	0.287	-0.170	2.169	1.897
Observations	9341	9341	9341	9341	9341	9341
Adjusted R^2	0.114	0.047	0.115	0.049		
Pseudo R^2					0.117	

Note: This table provides estimation results for the country-level sourcing potential equation under different specifications. Columns 1 and 2 report the baseline results. Columns 3 and 4 report results for the log-linearized model with $\log(\chi_{ijt}/\chi_{iht})$ on the left hand side. Finally, columns 5 and 6 report the estimation results for a Poisson regression with χ_{ijt}/χ_{iht} as the dependent variable. The independent variables are the same in all regressions. In columns 2, 4, and 6, log population is used as IV for log wages. The last equation is estimated via generalized method of moments.

Standard errors in parentheses.

country j in year t and 0 otherwise. Effectively, the term W_{ijt} summarizes the firm's history of importing in country j between time $t-a$ and time $t-1$. The larger a is, the longer the history of import affects the firm's variable trade cost. In other words, the term a captures the length of the ramping up process.

As the variable trade costs affect a firm's variable profit, the static profit from importing from a set b_{it} of import sources becomes

$$\pi_{it}(b_{it-a},...,b_{it-1},b_{it}) = \sigma^{-1}r_{iht}(b_{it-a},...,b_{it-1},b_{it}) + f_{it}(b_{it}) + s_{it}(b_{it},b_{it-1}) \tag{B3}$$

In the original model, the firm's static profit in year t depends only $t-1$ history through the sunk costs. This version allows for a longer history through the firm's variable profit.

The firm's dynamic profit for a given set b_{it} is then

$$\mathbb{E}[\sum_{\tau=t}^{t+L_{it}} \delta^{\tau-t}\pi_{i\tau}(b_{i\tau-a},...,b_{i\tau-1},b_{i\tau})|b_{it},\Omega_{it}] \tag{B4}$$

where Ω_{it} is the firm's information set at year t, which includes the firm's history $(b_{i\tau-a},...,b_{i\tau-1})$. Under Bellman's optimality principle, the optimal set of import sources satisfies:

$$V_{it}(\Omega_{it}) = \max_{b} \overline{\pi}_{it}(b,b_{it-1},...,b_{it-a}) + \delta\mathbb{E}[V_{it+1}(\Omega_{it+1})|b,\Omega_{it}] \tag{B5}$$

where $\overline{\pi}_{it}(.)$ is the expected value of equation B3. The choice-specific value function for set b is $V_{it}(b,\Omega_{it}) = \overline{\pi}(b,b_{it-1},...,b_{it-a}) + \delta\mathbb{E}[V_{it+1}(\Omega_{it+1})|b,\Omega_{it}]$. If the firm chooses set b over set b', it must follow that $V_{it}(b,\Omega_{it}) \geq V_{it}(b',\Omega_{it})$. Plugging in the expression for the firm's static profit, we can rewrite this condition in terms of differences in current profits, fixed costs, sunk costs, and future profits as follows:

$$\begin{aligned} &\underbrace{\sigma^{-1}\mathbb{E}[r_{iht}(b) - r_{iht}(b')|\Omega_{it}]}_{(1)} + \underbrace{\{\delta\mathbb{E}[V_{it+1}(\Omega_{it+1})|b,\Omega_{it}] - \delta\mathbb{E}[V_{it+1}(\Omega_{it+1})|b',\Omega_{it}]\}}_{(2)} \\ &\geq \underbrace{\mathbb{E}[\sum_{j\in b} f_{ijt} - \sum_{j\in b'} f_{ijt}|\Omega_{it}]}_{(3)} + \underbrace{\mathbb{E}[\sum_{\substack{j\in b\\ j\notin b_{it-1}}} s_{ijt} - \sum_{\substack{j\in b'\\ j\notin b_{it-1}}} s_{ijt}|\Omega_{it}]}_{(4)} \end{aligned} \tag{B6}$$

As before, under the revealed preferences assumption, we can create one-period deviations from the firm's observed choices to generate the inequalities that identify the bounds for cost parameters using equation B6. Intuitively, by altering only the firm's current choice but keeping its past and future choices the same, we can identify the fixed cost using continuing and exiting importers and the sum of the fixed and sunk costs using new importers and firms that never import. This identification strategy is the same as in the baseline model.

The longer history dependence in the firm's revenue, however, introduces two changes to the estimation procedure. The first term in equation B6, which captures the change in expected current revenue when we alter the observed choice, will now depend on the firm's import tenure in the country that is either added or dropped to the firm's observed set.

Furthermore, in the original model, deviations to choices in year t affect expected future profits (second term in equation B6) only through changes in sunk costs in year $t+1$. However, in the presence of the ramping up effect, if firms correctly expect growth in revenues, the second term would also capture changes to future revenues from year $t+1$ to year $t+a$.

The longer history, unfortunately, requires substantial more information. Suppose the ramping up effect takes three years to fully materialize (i.e., $a = 3$). In this case, for each one-period deviation, we would need three years of history and three additional years to capture future growth, if assuming that firms have the correct belief about their future profits. Given that my sample only covers 7 years, I assume that firms have incorrect beliefs and do not account for future growth when choosing their current import sets. This assumption allows me to incorporate the ramping up effect in the estimation without much loss of data.[2] I allow the firm's history to go back three years.

Estimation details

As previously mentioned, we would need to account for the ramping up effect when estimating the marginal revenues of importing, or the change in current revenue when adding or dropping an import source. A key part of this step is to estimate the sourcing potential. Recall that sourcing potential is defined as $S_{ijt} = T_j(\tau^m_{ijt} w_{jt})^{-\theta}$, where T_j is country j's technological level and w_{jt} is country j's labor wage in year t. As the variable trade cost τ^m_{ijt} is now a function of firm i's import tenure in country j, W_{ijt}, S_{ijt} is also a function of W_{ijt}.

The modified estimating equation is

$$\hat{\xi}_{ijt} = \beta_0 + g(X^T_{jt}\beta^T) - \theta h(X^\tau_{ijt}\beta^\tau, W_{ijt}\beta^W) - \theta \ln w_{jt} + \lambda_t + \nu_{ijt} \quad \text{(B7)}$$

[2]Under correct belief and $a = 3$, I can use deviations from only one single year. Reducing the ramping effect to two years increases the information I can use but still does not suffice to generate meaningful bounds for the cost parameters. Fortunately, the moment inequalities allow for flexible assumptions on the firm's belief.

where $\hat{\xi}_{ijt}$ is estimated following the procedure in Appendix C1 and captures the sourcing potential S_{ijt} and unobserved firm-country-year-specific shock in the firm's imported input share (e.g., measurement error), assumed to be independent of S_{ijt}. As before, X^T_{jt} is a set of technology proxies, X^τ_{ijt} is a set of controls to proxy for variable trade costs (excluding import tenure), and $\ln w_{jt}$ is the log of human capital-adjusted hourly wages.

The new term in equation B7 is firm i's import tenure in country j at year t, $W_{ijt} = W(d^m_{ijt-1}, d^m_{ijt-2}, d^m_{ijt-3})$. I allow for three different specifications of $W(.)$:

$$W^{(1)}(d^m_{ijt-1}, d^m_{ijt-2}, d^m_{ijt-3}) = \sum_{s=1}^{3} d^m_{ijt-s}$$

$$W^{(2)}(d^m_{ijt-1}, d^m_{ijt-2}, d^m_{ijt-3}) = \iota_1 1\Big\{\sum_{s=1}^{3} d^m_{ijt-s} = 1\Big\} + \iota_2 1\Big\{\sum_{s=1}^{3} d^m_{ijt-s} = 2\Big\} + \iota_3 1\Big\{\sum_{s=1}^{3} d^m_{ijt-s} = 3\Big\}$$

$$W^{(3)}(d^m_{ijt-1}, d^m_{ijt-2}, d^m_{ijt-3}) = \kappa_1 1\Big\{d^m_{ijt-1} = 1\Big\} + \kappa_2 1\Big\{d^m_{ijt-2} = 1\Big\} + \kappa_3 1\Big\{d^m_{ijt-3} = 1\Big\}$$

The first two specifications imply that experience from each of the past three years are perfect substitutes for one another, but the second specification allows for non-linear effects. The third specification, however, differentiates between recent and distant history.

Table B3 reports the regression results based on equation B7 for all three specifications of $W(.)$. As predicted, columns 1 and 2 show that sourcing potential increases with a firm's import tenure. Column 3 shows that the gain is concentrated in year $t-1$, and the magnitude of the lag coefficients drops significantly in $t-2$ and $t-3$. This result suggests that firms learn most from the more recent experience, and firms that import two or three years ago enjoy a smaller benefit. This result is consistent with the findings documented in Roberts and Tybout (1997) and Clerides et al. (1998) in the context of Colombian exporters.

Figure B1 shows the 95% confidence sets for the cost parameters under the baseline model (red region) and the model with new importer dynamics (black region).[3] While the costs to existing importers stay somewhat in the same range, new importers pay substantially lower costs when accounting for the ramping up effect compared with the baseline estimate (the new upper bound is only about 60% of the baseline upper bound). This is because when profits grow more slowly, a new import source becomes less valuable compared to the baseline model, and as a result, the estimated cost that can rationalize entering a new market becomes smaller. The result is consistent with Ruhl and Willis (2017), which found that export entry costs fall significantly when augmenting the sunk-cost model with gradual growth in export volumes.

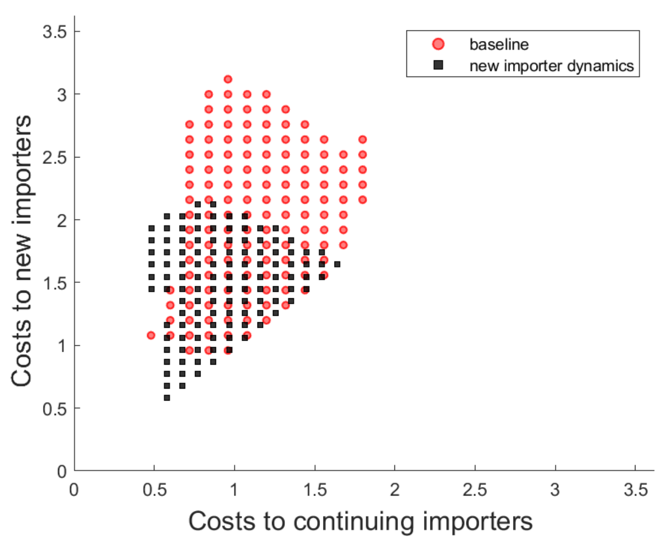
[3] I use results from column 1 in Table B3 to compute the bounds. The results are similar using the other specifications.

Table B3—: Predicting sourcing potential

	(1)	(2)	(3)
Import tenure	0.505		
	(0.032)		
Import tenure = 1		0.395	
		(0.098)	
Import tenure = 2		0.535	
		(0.120)	
Import tenure = 3		1.375	
		(0.115)	
Import in $t-1$			0.785
			(0.090)
Import in $t-2$			0.117
			(0.111)
Import in $t-3$			0.350
			(0.118)
Observations	6669	5096	5096

Note: This table reports the regression results based on equation B7. Columns 1, 2, 3 correspond to specifications $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$ listed in the text. Standard errors in parentheses.
Source: General Administration of Customs and National Bureau of Statistics

Figure B1. : 95% confidence set



Note: This figure illustrates the 95% confidence sets of the total costs to continuing versus new importers. The red region depicts the CI under the baseline model, whereas the blue region depicts the CI when accounting for new importer dynamics. Monetary values are in million of 1998 RMB.