Online Appendix

Cities, Heterogeneous Firms, and Trade

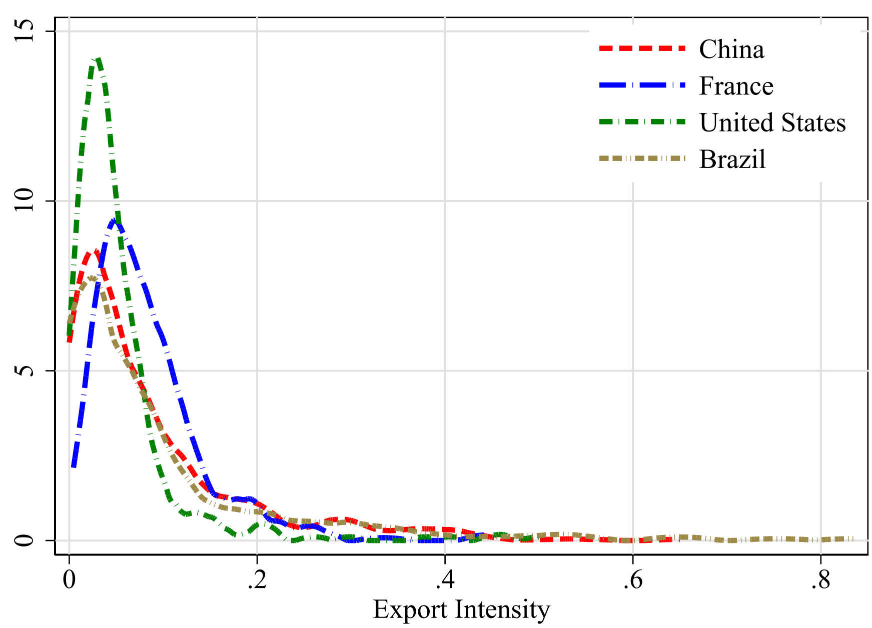
Jan David Bakker Alvaro Garcia-Marin Andrei Potlogea
Nico Voigtländer Yang Yang

A Additional Empirical Results

A.1 Export Intensity Distributions in the Four Countries

Figure A.1 plots the distributions of our main variable of interest – export intensity across cities – for China, France, the United States, and Brazil. The distribution of export intensity is positively skewed for all countries in our sample, with a substantially fatter upper tail in Brazil and China than in France and the United States.

Figure A.1: Distributions of Export Intensity in the Four Countries



Notes: The figure shows the distribution of export intensity across cities in China, France, the United States, and Brazil. See the note to Table 1 in the paper for the definition of cities in the four countries.

A.2 Alternative Measures for Export Intensity and City Size

In this appendix, we document the robustness of our core stylized fact to alternative measures for export intensity and city size.

Per-capita exports. We begin with an alternative variable for export intensity. Our main results (Table 2) use city-level exports relative to *sales*. Table A.1 presents results with an alternative denominator – city *population*, thus using *per capita* exports as a proxy for export activity. All coefficients are statistically significant at the 1 percent level, with magnitudes ranging from 0.23 for China to 0.45 in Brazil.

Table A.1: Per Capita Exports and City Size in the Four Countries

Dependent variable: City-level ln(exports per capita)

	China		France		United States		Brazil	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln City Size	0.318***	0.231***	0.341***	0.324***	0.278***	0.294***	0.470***	0.449***
	(0.062)	(0.051)	(0.057)	(0.054)	(0.043)	(0.046)	(0.132)	(0.134)
Geography Controls		✓		✓		✓		✓
R^2	0.03	0.29	0.12	0.19	0.09	0.20	0.03	0.25
Observations	615	615	304	304	329	329	297	297

Note: The table shows that our main results from Table 2 are robust to using an alternative measure for export intensity: per capita exports. See the note to Table 1 for the definition of cities in the four countries. City-level per capita exports are defined as the ratio between manufacturing exports and population for China, and as the ratio between overall city-level exports and city population for France, the United States, and Brazil. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors in parentheses. Key: ** significant at 1%; ** 5%; * 10%.

Exports over value added. Next, we define export intensity as export sales over *value added* as dependent variable in regression (2). If the 'upstreamness' of firms' production systematically affected our results, we would expect significantly different results under this alternative, value-added based, definition of export intensity. Table A.2 reports these results in columns 2 and 4, showing that they are very similar to our baseline (reported in columns 1 and 3). This suggests that our core stylized fact is not driven by differences in the 'upstreamness' across cities.

Population density. Instead of city population, Panel A of Table A.3 uses city-level population *density* as the main explanatory variable. We compute this measure as population divided by city area. The results in Panel A of Table A.3 show that in all cases, elasticities remain positive, with magnitudes ranging from 0.15 in China to 0.25 in France. Point estimates are actually larger for France and the U.S. (as compared to our baseline results in Table 2), but somewhat smaller for Brazil and China. Panel B of Table A.3 replicates the analysis using ln(per-capita exports) as dependent variable in combination with population density. The estimated coefficients in this

Table A.2: Export Intensity and Exports over Value Added

Dependent variable: As indicated in panel header

	—— China ——		—— France ——	
	(1)	(2)	(3)	(4)
Dep. Var.:	$\ln\left(\frac{\text{exports}_i}{\text{sales}_i}\right)$	$\ln\left(\frac{\text{exports}_i}{\text{Value Added}_i}\right)$	$\ln\left(\frac{\text{exports}_i}{\text{sales}_i}\right)$	$\ln\left(\frac{\text{exports}_i}{\text{Value Added}_i}\right)$
ln(City Size)	0.258***	0.264***	0.185***	0.205***
	(0.046)	(0.046)	(0.041)	(0.043)
Geog. controls	✓	✓	✓	✓
R^2	0.26	0.29	0.18	0.16
Observations	615	614	304	304

Note: This table reports our baseline results in columns 1 and 3 for China and France (where we define export intensity as exports over sales). Columns 2 and 4 use an alternative definition: export sales over *value added*. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors in parentheses. Key: ** significant at 1%; ** 5%; * 10%.

specification are also similar to our baseline, varying from 0.28 in Brazil and China to 0.49 in France.

Table A.3: Export Intensity, Per Capita Exports and City Density in the Four Countries

Dependent variable: As indicated in panel header

	—— China ——		—— France ——		— United States —		—— Brazil ——	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Dependent Variable: ln(City-Level Export Intensity)								
ln(City Density)	0.160**	0.153**	0.266***	0.251***	0.162***	0.183***	0.137*	0.171
	(0.065)	(0.062)	(0.035)	(0.037)	(0.052)	(0.062)	(0.082)	(0.110)
Geog. Controls		✓		✓		✓		✓
R^2	0.01	0.24	0.16	0.25	0.03	0.16	.01	0.18
Observations	615	615	304	304	329	329	297	297
Panel B. Dependent Variable: ln(City-Level Per Capita Exports)								
ln(City Density)	0.324***	0.275***	0.465***	0.491***	0.280***	0.309***	0.233**	0.279**
	(.081)	(.071)	(.038)	(.041)	(.060)	(.070)	(.071)	(.092)
Geog. Controls		✓		✓		✓		✓
R^2	0.02	0.29	0.29	0.36	0.07	0.17	0.02	0.24
Observations	615	615	304	304	329	329	297	297

Note: The table analyzes the relationship between city size and per capita exports. See the note to Table 1 in the paper for the definition of cities in the four countries, the note to Table 2 for the definition of city-level export intensity, and the note to Table A.1 for the definition of per-capita exports. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors in parentheses. Key: ** significant at 1%; ** 5%; * 10%.

Export intensity in manufacturing vs. services. The French data allow us to examine the relationship between export intensity and city size separately for manufacturing, services, and the primary sector. Table A.4 shows the results. For comparison, column 1 repeats the result for firms in all sectors. Columns 2 and 3 show that the gradient of export intensity with city size is remarkably similar when we restrict the sample to firms in manufacturing and services, respectively. Column 4 in Table A.4 instead examines firms in the primary sector, which depends on land abundance and natural resources – and arguably much less on the mechanism in our framework. Accordingly, we find that export activity for primary materials is actually *less* concentrated in larger cities.

Table A.4: Export Intensity and City Size across Sectors in France

Dep. Variable: City-Level ln(Export Intensity), by Sector

	All Industries (1)	Manufacturing (2)	Services (3)	Primary Sector (4)
ln(City Population)	0.185***	0.280***	0.252***	-0.093
	(0.041)	(0.056)	(0.047)	(0.099)
Geog. Controls	✓	✓	✓	✓
Mean Export Intensity:	0.084	0.177	0.037	0.084
R^2	0.18	0.30	0.15	0.10
Observations	304	304	304	269

Note: The table replicates the result for France in Table 2, col 4, using different subsamples. Cities in France are defined in terms of employment zones. The analysis only considers cities with at least 250 firms. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors are in parentheses. Key: *** significant at 1%; ** 5%; * 10%.

A.3 China-Specific Robustness Checks

This appendix section performs China-specific robustness checks of our core stylized fact.

SEZ and CDA cities in China. A distinctive element of the Chinese economy is the existence of Special Economic Zones (SEZ) and Coastal Development Areas (CDA), which are intended to promote exports and overall economic activity in selected areas. We show in row 2 of Table A.5 that our main results are virtually unchanged when we include categorical variables for SEZ and CDA cities (row 1 repeats our baseline results for comparison). Second, in row 3 we define export intensity using information from the Chinese Customs Service, which only includes direct exports but is limited due to poor matching with the Census of Manufacturing. Nevertheless, our results remain very similar. Third, an important body of literature uses prefecture-level Chinese cities as the main unit of analysis (e.g., Au and Henderson, 2006). Row 4 in Table A.5 shows that our main findings are qualitatively unchanged when using prefecture-level cities. Row 5 shows that the positive correlation between export intensity maintains its significance when controlling for the travel time to the nearest port Egger et al. (from 2023), which is a superior proxy for coastal access costs than linear distance as it captures variation in transportation infrastructure across locations. Finally, the results using SEZ and CDA dummies, and using customs data (as in rows 2 and 3) are also very similar when we use prefecture-level cities to compute city size (rows 6 and 7).

'State-involved' and foreign firms. Another distinct feature of the Chinese economy is the large presence of state-involved enterprises and the importance of foreign firms, especially for exporting. We show in Table A.6 that the presence of these firms does not drive our main stylized fact, nor the decomposition into intensive and extensive margins of export intensity. In Panel A, we exclude all enterprises with state involvement. These include purely state-owned enterprises, purely collective-owned enterprises, joint operations between state- and/or collective-owned enterprises, and partly state-invested firms and stock companies (codes 110 - 160 of the Industrial Enterprise Ownership Classification (IEOC) of the Chinese National Bureau of Statistics from 1998). The overall correlation between export intensity and city size becomes even stronger when we exclude these 'state-involved' firms, as compared to our baseline results for China in Table 5. When decomposing the correlation into the intensive and the extensive margin, we find that the result is entirely driven by the latter. Note that the coefficient for the intensive margin is small and insignificant. This suggests that the negative coefficient that we find in our baseline (Table 5) is driven by 'state-involved' firms. We confirm in additional checks (available upon request) that these 'state-involved' firms tend to be i) larger, ii) locate in larger cities, and iii) sell disproportionately domestically.[1] These features can lead to a bias towards a negative intensive margin when these firms are included.

[1] Specifically, on average, state-owned firms are 16% larger, locate in cities that are 6.5% larger and have a domestic share that is 5% higher than the rest of the firms in the economy.

Table A.5: Export Intensity and City Size: Further Robustness Check for China

Dependent variable: City-level ln(export intensity)

		Coeff./ St. Err.	Obs./ R^2	Geographic Controls	SEZ and CDA Dummies	City definition
(1)	Baseline (Urban population, geographical controls)	0.258*** (0.046)	615 0.263	✓		MA
(2)	Urban population, CDA and SEZ dummies	0.269*** (0.047)	615 0.264	✓	✓	MA
(3)	Urban Population, Customs exports	0.224*** (0.053)	576 0.205	✓	✓	MA
(4)	Urban pref. population, geographic controls	0.216** (0.090)	329 0.310	✓		Prefecture
(5)	Urban pref. population, travel time controls	0.282** (0.096)	329 0.268	(✓)		Prefecture
(6)	Urban pref. population, CDA and SEZ dummies	0.219** (0.091)	329 0.313	✓	✓	Prefecture
(7)	Urban pref. population, Customs exports	0.209** (0.102)	322 0.284	✓	✓	Prefecture

Note: The table replicates our baseline results for China from Table 2, col 2, for different definitions of cities and different sources for the trade data. All regressions include the same set of geographical controls as Table 2, except for row 5, which replaces the baseline set of geographical control with the (log) travel time to the nearest port and the average travel time to other prefectures in China. The rows marked with CDA and SEZ dummies additionally include a dummy for whether a city is located in a Special Economic Zones (SEZ) and Coastal Development Areas (CDA), respectively. For all regressions, the dependent variable is the natural logarithm of export intensity.

In Panel B of Table A.6 we exclude foreign-owned firms, using a conservative definition that also counts as 'foreign' firms with ownership from Hong Kong, Tawain and Macao (codes 200 - 340 in the IEOC). The resulting correlations are very similar to our baseline; if anything, they become slightly stronger. Lastly, in Panel C, we use a particularly strong sample restriction, using only domestic-owned private-sector firms (IEOC codes 170 - 174). The resulting correlation between export intensity and city size is similar to the baseline, but the intensive margin in this sample is no longer negative; it is instead marginally increasing with city size, in line with the prediction by our quantitative model (see Table 7). Overall, our results are robust to these different sample restrictions.

Table A.6: Export Intensity and City Size for Different Subsamples in China

Dependent variable: As indicated in table header

	(1)	(2)	(3)	(4)
Dep. Variable:	$\ln\left[\text{Export intensity}\right]$	$\ln\left[\text{Export intensity}\right]$	$\ln\left[\text{Intensive margin}\right]$	$\ln\left[\text{Extensive margin}\right]$
Panel A. Excluding state-involved firms				
ln(City Size)	0.395***	0.316***	-0.047	0.363***
	(0.063)	(0.056)	(0.032)	(0.050)
Geog. controls		✓	✓	✓
R^2	0.039	0.257	0.066	0.247
Observations	602	602	602	602
Panel B. Excluding foreign firms†				
ln(City Size)	0.372***	0.309***	-0.184***	0.494***
	(0.056)	(0.052)	(0.036)	(0.047)
City controls		✓	✓	✓
Observations	613	613	613	613
Pseudo R^2	0.05	0.21	0.10	0.20
Panel C. Only domestic private firms				
ln(City Size)	0.366***	0.311***	0.059	0.252***
	(0.076)	(0.069)	(0.039)	(0.053)¿
Geog. controls		✓	✓	✓
R^2	0.03	0.22	0.08	0.19
Observations	591	591	591	591

Note: The table replicates the results of the micro-decomposition in Table 5 for different samples of Chinese firms. See the note to Table 5 for details on the decomposition of overall export intensity into the intensive and the (value) extensive margin. In Panel A, we exclude any 'state-involved' enterprises. In Panel B, we exclude foreign firms. In Panel C, we only include domestic private firms (i.e. we exclude state-involved, foreign-owned and other firms). The analysis only considers cities with at least 250 firms. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors are in parentheses. Key: *** significant at 1%; ** 5%; * 10%.
† Foreign firms also comprise those with ownership from Hong Kong, Tawain and Macao

A.4 Accounting for Multi-Location Firms

For our firm-level data (France and China), firm location is only defined at the headquarter-level. This may introduce an upward bias if firms with multiple establishments produce exported goods also in small cities but locate their headquarters in larger cities.

Multi-location firms in China. For China, the Census of Manufacturing contains information on whether a firm has only a single industrial unit. To account for any issues related to multi-location firms, we use this information to create a sample of single-establishment firms that can be clearly linked to one location. Panel A in Table A.7 estimates the correlation between export intensity and city size, as well as the intensive margin and the extensive margin correlations, as defined in equation (3). We find results that are remarkably similar to the full sample (shown in Tables 2 and 5 in the paper).

Multi-location firms in France. For France, we can identify single- and multi-location firms by linking the baseline firm-level data with administrative information from the Annual Declaration of Social Data (DADS). The DADS contains information about employment and wages in every establishment of all firms that pay social security contributions. While this data provides no information on sales at the establishment level, we use this auxiliary dataset in two ways. First, we identify single- vs. multi-location firms. Panel B in Table A.7 shows that our core results are similar when restricting the analysis to firms that are active only in a single employment zone.[2] Second, for firms with establishments in multiple employment zones, we can compute a proxy: assuming that revenues are generated proportionally to the wage bill across each firm's establishments, we can assign both domestic and export revenues to individual establishments of multi-location firms. Thus, we effectively generate establishment-level (and thus city-specific) export intensity in the French data. We use this variable to repeat our core analysis for export intensity in Panel C of Table A.7. All results are fully in line with our baseline estimates.

[2] One difference is the statistically significant (albeit small) coefficient for the intensive margin. However, this coefficient is significantly smaller than the extensive margin, in line with our model (and it disappears again in the next, alternative, analysis for French multi-location plants).

Table A.7: Export Intensity and City Size. Accounting for Multi-City Firms

Dependent variable: As indicated in table header

	(1)	(2)	(3)
Dep. Variable:	ln [Export intensity]	ln [Intensive margin]	ln [Extensive margin]
Panel A. China: Single-establishment firms only			
ln(City Size)	0.272***	-0.134***	0.406***
	(0.047)	(0.031)	(0.043)
Geog. Controls	✓	✓	✓
R^2	0.26	0.09	0.24
Observations	613	613	613
Panel B. France: Single-location firms only			
ln(City Size)	0.206***	0.073***	0.133***
	(0.040)	(0.027)	(0.028)
Geog. Controls	✓	✓	✓
R^2	0.22	0.06	0.26
Observations	303	303	303
Panel C. France: All Establishments, revenue distributed proportional to wage bill			
ln(City Size)	0.146***	0.006	0.140***
	(0.038)	(0.029)	(0.024)
Geog. Controls	✓	✓	✓
R^2	0.18	0.05	0.33
Observations	303	303	303

Note: The table replicates the results of the micro-decomposition in Table 5, accounting for the possible role of multi-establishment firms. Cities in France are defined in terms of employment zones and in terms of MSAs for China. See the note to Table 5 for details on the decomposition of overall export intensity into the intensive and the (value) extensive margin. In Panel B, we restrict the sample for France to firms that only have establishments in one employment zone. In Panel C, we assume that the (domestic and export) revenue of French firms is created proportionally to the wage bill across its establishments. The analysis only considers cities with at least 250 firms. 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors are in parentheses. Key: *** significant at 1%; ** 5%; * 10%.

A.5 2SLS Results Using Agricultural Suitability as Instrument for City Size

This appendix complements Section 3.3 in the paper, describing the underlying data in more detail and providing additional results.

Caloric suitability and historical population. We use data on the caloric suitability index (CSI) for agricultural crops from Galor and Özak (2016). We use the maximum potential caloric yield attainable, given the set of crops that are suitable for cultivation in the post-1500 period. This includes the potato, which was central to the growth of cities after 1500 (Nunn and Qian, 2011). We match the original grid-cell level data to Chinese prefectures and to French employment zones, and we compute the average local caloric yield that is attainable when the maximum-yield crop is planted.[3] Because the suitability measures do not take into account output-enhancing technologies such as fertilizer, we refer to CSI as "historical caloric suitability."

We first verify that historical caloric suitability indeed predicts historical urban populations in our sample. For China, historical population at the MSA level is not available. We thus move to the prefecture level, for which Bai and Jia (2021) provide urban population in the 1580s.[4] We re-compute our geographic control variables using the location and borders of prefecture-level cities, and link them to the prefecture-level dataset. For France, we use population records for 1876 from INSEE at the level of current municipalities, which we aggregate to employment zones.[5] Table A.8 presents the results of regressing log historical population in the two countries on caloric suitability. In order to interpret our results, we normalize CSI. All specifications include both CSI and an interaction of CSI with a dummy for coastal cities. This allows for the effect to differ in coastal areas, where fishing was also an important contributor to overall caloric intake.[6] We obtain statistically highly significant effects of caloric suitability in both countries.[7] Columns 2 and 4 show that the results are similar when we control for log distance to the coast, in addition to coastal

[3]The data are available for download here: https://zenodo.org/records/14714917, and they are available for 5'x5' (arcmin – approx. 10x10 km) grid cells.

[4]More precisely, Bai and Jia (2021) map historical populations into modern prefecture borders and report the historical urban density within each prefecture. We multiply this variable with the prefecture area to obtain the historical urban population. Firms in our data are located in 344 prefectures. However, Bai and Jia (2021) only provide historical population information for the 261 prefectures belonging to China proper – i.e., excluding the remaining 74 prefectures from Inner China provinces, such as the Tibet or Inner Mongolia provinces. See Bai and Jia (2021) for details on the mapping of historical population records to current prefecture geographical areas.

[5]*Historique des populations communales: Recensements de la population 1876-2019*, available at https://www.insee.fr/fr/statistiques/3698339. To aggregate the municipal population records to the employment zone level, we use the official correspondences provided by INSEE: https://www.insee.fr/fr/information/2114596.

[6]We also obtain a strong positive coefficient on CSI when including it alone (without the interaction term). However, the R^2 is higher in the interaction specification.

[7]The interaction term is statistically significant and negative for France, and its magnitude is similar to the level effect of CSI. This implies that in coastal employment zones, urban population was unrelated to CSI. For China, on the other hand, the interaction term is also positive and significant. A likely explanation is that Chinese prefectures are much larger than French employment zones, so that they include areas that were not immediately located on the coast. Higher crop yield in these areas, in turn, may have further accelerated the growth of Chinese cities that are located in relative proximity of the coast.

and border dummies. In terms of magnitude, a one-standard-deviation increase in CSI predicts an approximately 58% higher historical urban population in China and about 16% in France. This is the same order of magnitude as for present population (see the first-stage results in Table 3).

Table A.8: Caloric Suitability and Historical Population

Dependent Variable: ln(Historical Population)

	China		France	
	(1)	(2)	(3)	(4)
CSI	0.568***	0.578***	0.166***	0.160***
	(0.101)	(0.109)	(0.042)	(0.043)
CSI × Coastal	1.157**	1.149**	-0.224**	-0.229**
	(0.451)	(0.451)	(0.104)	(0.096)
Geog. Indicators	✓	✓	✓	✓
Coastal Access Control		✓		✓
Mean Dep. Var.	12.86	12.86	11.43	11.43
R^2	0.262	0.262	0.03	0.05
Observations	261	261	297	297

Notes: This table examines the effect of the historical caloric suitability index (CSI) on historical population size in China (columns 1 to 2) and France (columns 3 to 4). The CSI variable is standardized. The analysis for China is run at the prefecture level, using 1580 prefecture-level population from Bai and Jia (2021). The analysis for France is run at the employment zone level, using the population records for 1876 produced by INSEE. 'Coastal' is a dummy for locations on the coast. 'Geog. Indicators' for both countries include coastal and border dummies. 'Coastal Access Control' is log distance to the coast. Robust standard errors (in parentheses). Key: *** significant at 1%; ** 5%; * 10%.

Table A.9 complements our 2SLS results from Table 3 in the paper, presenting the corresponding reduced-form results that regress export intensity for China and France directly on CSI (and its interaction with a coastal dummy). We find a strong and positive coefficient on CSI in levels, and a negative (but statistically insignificant) coefficient on the interaction CSI×Coastal. When controlling for variation in coastal access (columns 2 and 4), coefficients remain highly significant and similar for the two countries. The point estimates imply that a one-standard-deviation higher CSI results in a 24% (China) and 11% (France) higher export intensity today.

Table A.10 complements our discussion of the exclusion restriction in Section 3.3 in the paper. Here, we examine whether CSI may violate the exclusion restriction by affecting historical or contemporaneous infrastructure. We use data for France, where detailed information on an important historical transportation network is available: Roman Roads. We map Roman Roads into the boundaries of modern-day employment zones in France and then compute the density of roads by employment zone.[8] In addition, we use the same methodology to compute the density of

[8]For the historic Roman Road network we use the data provided by McCormick, Huang, Zambotti, and Lavash (2013), accessible at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TI0KAU.

Table A.9: Reduced-Form Results: Caloric Suitability and Export Intensity

Dependent variable: ln(export intensity)

	— China —		— France —	
	(1)	(2)	(3)	(4)
CSI	0.339***	0.241**	0.116**	0.107**
	(0.116)	(0.117)	(0.053)	(0.051)
CSI × Coastal	-0.138	-0.101	-0.126	-0.133
	(0.266)	(0.314)	(0.108)	(0.107)
Geog. Indicators	✓	✓	✓	✓
Coastal Access Control		✓		✓
R^2	.243	.262	0.06	0.09
Observations	329	329	304	304

Notes: This table shows the reduced-form regressions that correspond to the analysis in Table 3. See note of Table 3 for details. 'Coastal' is a dummy for locations on the coast. 'Geog. Indicators' for both countries include coastal and border dummies. 'Coastal Access Control' is log travel time to the nearest port for China, where this prefecture-level variable is available from Egger et al. (2023); for France we use log distance to the coast as a proxy. Robust standard errors (in parentheses). Robust standard errors (in parentheses). Key: *** significant at 1%; ** 5%; * 10%.

modern-day highways, railways, and canals.[9]

We first show in columns 1 and 2 in Table A.11 that CSI is a strong predictor of urban population in France.[10] Next, columns 3 and 4 show that CSI is not correlated with the density of Roman Roads: The coefficient is negative, minuscule, and statistically insignificant. Note also that the standard errors are small, so that we can be confident of close-to-zero effects. Next, columns 5-10 show that CSI is also uncorrelated with modern highways, railways, and canals. Similar to Roman Roads, the coefficients are small and statistically insignificant. These results render it unlikely that our exclusion restriction in the 2SLS results is violated because of an effect of CSI on export intensity via transport infrastructure.

Finally, we perform an additional consistency check in Table A.11. One concern with the results above could be that Roman Roads or Highway density is merely measured with error and that is why we do not find a relationship with CSI.[11] To directly address this possibility, we show in Table A.11 that Roman Roads strongly predict today's highway network (cols 1-3) as well as contemporaneous population (cols 4-6). On the other hand, adding our instrument CSI to these

[9]The data for modern-day highways, railways, and canals is from OpenStreetMap: https://data.humdata.org/dataset/hotosm_fra_roads (Highways); https://data.humdata.org/dataset/hotosm_fra_railways (Railways); https://data.humdata.org/dataset/hotosm_fra_waterways (Canals). For railways, we only use observations indicated as *rail*; for highways, we use *motorway*, *motorway_link*, *trunk*, *trunk_link*, *primary*, and *primary_link*; and for canals, we use *canal* and *cana*.

[10]We use CSI in levels (without the coastal interaction) throughout these specifications for a more straightforward interpretation of the (potential) correlation of CSI with infrastructure.

[11]Note, however, that the small standard errors in Table A.10 render this unlikely.

Table A.10: Caloric Suitability and Infrastructure in France

Dependent variable: As indicated in table header

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. variable:	$\log(\text{pop}_{i,1999})$		Roman Road dens.		Highway dens.		Railway dens.		Canal dens.	
CSI	0.138***	0.170***	-0.001	-0.002	0.011	0.009	0.019	0.020	0.007	0.009
	(0.043)	(0.051)	(0.002)	(0.002)	(0.007)	(0.008)	(0.014)	(0.013)	(0.005)	(0.006)
Geog. Controls		✓		✓		✓		✓		✓
Mean Dep. Var.	11.61	11.61	0.05	0.05	0.21	0.21	0.16	0.16	0.04	0.04
Observations	304	304	304	304	304	304	304	304	304	304
R-squared	0.02	0.08	0.00	0.04	0.00	0.13	0.00	0.04	0.01	0.10

Notes: This table examines the correlation of Soil Suitability (measured as the optimal crop yield using plants available post-1500) with modern population (columns 1 to 2), local Roman Road density (columns 3 to 4), local Roman Road density (columns 5 to 6), local railway density (columns 7 to 8) and local canal density (columns 9 to 10). The CSI variable is standardized. All densities are defined as the length over area. 'Geog. Controls' are listed in the note to Table 2 in the paper. Robust standard errors (in parentheses). Key: *** significant at 1%; ** 5%; * 10%.

regressions with Roman Roads, we confirm that CSI does not predict modern highways, while it remains a strong predictor of today's urban population. This confirms that i) Roman Roads do have predictive power throughout, while CSI only predicts modern population – exactly in line with our 2SLS strategy and the associated exclusion restriction.

Table A.11: Caloric Suitability, Roman Roads, Modern Outcomes in France

Dependent variable: As indicated in table header

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable:	Highway density			$\log(\text{pop}_{i,1999})$		
Roman Road density	0.971**	0.840**	0.859**	5.170***	5.796***	6.121***
	(0.376)	(0.367)	(0.370)	(1.565)	(1.521)	(1.536)
CSI			0.011			0.183***
			(0.008)			(0.050)
Geog. Controls		✓	✓		✓	✓
Mean Dep. Var.	0.21	0.21	0.21	11.61	11.61	11.61
Observations	304	304	304	304	304	304
R-squared	0.04	0.16	0.16	0.03	0.09	0.12

Notes: This table examines the effects of local Roman Road density and of historical caloric suitability (CSI) on modern population and infrastructure. The CSI variable is standardized. The analysis is run at the French employment zone level. Roman road and highway density is defined as length over area; see Appendix A.5 for detail. 'Geog. Controls' are listed in the note to Table 2 in the paper. Robust standard errors (in parentheses). Key: *** significant at 1%; ** 5%; * 10%.

A.6 Within- and Between-Industries Variation

This appendix complements our results from Section 3.4. We first provide details on the decomposition of export intensity into the underlying within- and between-industries variation (see Table 4 for the results). We then present alternative results using city-industry data in combination with industry fixed effects.

Decomposition into within- and between-industries variation. Our decomposition of export intensity into within- and between-industries variation involves two steps. First, we compute the between-industry component as the counterfactual export intensity measure, ρ_i^B, that would result if city-level export intensity only varied due to differences in industry composition across cities. For each sector s, we first define its national-level export intensity, ρ_s, as the ratio between its national-level exports ($x_s = \sum_i x_{s,i}$) and sales revenues ($r_s = \sum_i r_{s,i}$), so that $\rho_s \equiv x_s/r_s$. We then construct the counterfactual between-industry component by multiplying ρ_s with the revenue share of sector s in city i, and summing over all sectors s:[12]

$$\rho_i^B \equiv \left(\sum_s \rho_s \times \frac{r_{s,i}}{r_i} \right) \tag{A.1}$$

Second, we define the within-industry component of export intensity (in logs) as the residual variation that is not accounted for by differences in the sectoral composition of cities:

$$\ln \rho_i^W \; = \; \ln \rho_i - \ln \rho_i^B \tag{A.2}$$

Specifications with industry fixed effects. Above we decomposed city-level exports into a within-industry and a between-industry component, where the latter picks up differences in industry composition across locations. Another way to gauge the importance of differences in industry composition for our stylized fact is to run regressions at the city-industry level and compare the correlation between export intensity and city size with and without industry fixed effects. Our data allow us to implement these city-industry level regressions with 4-digit industries for China and France (for Brazil and the US, industry-city level data are not available). We run the following specification

[12]We use detailed, 4-digit industries in our decomposition. At this level of detail, the between-component will inevitably pick up variation beyond differences in comparative advantage. For example, a large, highly productive exporting firm in sector s in city i (say, Airbus in Toulouse) may significantly affect ρ_s (i.e., the French aircraft sector is export-intensive) and thereby also this city's ρ_i^B. In other words, some firm- (rather than sector-) specific features may load onto ρ_i^B: Toulouse may appear export-intensive because of its industry composition, even though the export intensity of the aircraft sector is driven by Airbus' high productivity in Toulouse. Thus, our results for the within-component reflect a lower bound.

(with and without industry fixed effect δ_j:

$$f(\rho_{ij}) = \beta \ln(pop_i) + \gamma X_i + \delta_j + \varepsilon_{ij} \tag{A.3}$$

where ρ_{ij} denotes the export intensity of sector j in city i, pop_i is city population, and X_i denotes the standard set of city-level controls (distance to the border and the coast (in logs), a dummy for being at the border, a dummy for being on the coast and the average distance to other domestic locations (in logs)). Finally, the transformation $f(\cdot)$ of the dependent variable is motivated as follows: Since we now run regressions at the more finely-grained city-industry level, we frequently encounter zeros for ρ_{ij} (i.e., cities that do not export in a given 4-digit sector). To address this issue, we present results both using $f(\rho_{ij}) = \ln(\rho_{ij} + 1)$ and the inverse-hyperbolic sin transformation of export intensity as outcome variables: $f(\rho_{ij}) = \sinh^{-1}(\rho_{ij}) = \ln\left(\rho_{ij} + \sqrt{\rho_{ij}^2 + 1}\right)$.

Table A.12 displays the results, which are robust across the two transformations $f(\cdot)$. For China, we find that including industry fixed effects does not affect the correlation of city size and export intensity, while for France, including industry fixed effects reduces the coefficient by roughly one-third. Both of these findings corroborate the results from our main specification (see Table 4 in the paper).

Table A.12: Export Intensity and City Size: Industry Fixed Effects

Dependent variable: As indicated in table header

Dependent Variable:	$\ln(1+\text{Export Intensity})$		$IHS(\text{Export Intensity})$	
	(1)	(2)	(3)	(4)
Panel A. China				
ln(City Population)	0.0109***	0.0111***	0.0119***	0.0121***
	(0.0012)	(0.0012)	(0.0014)	(0.0014)
Implied average coefficient†	*0.213*	*0.218*	*0.221*	*0.225*
Geog. Controls	✓	✓	✓	✓
4-digit industry FE		✓		✓
R^2	0.060	0.062	0.056	0.057
Observations	42,911	42,911	42,911	42,911
Panel B. France				
ln(City Population)	0.0104***	0.0069***	0.0115***	0.0073***
	(0.0004)	(0.0003)	(0.0004)	(0.0004)
Implied average coefficient†	0.301	0.200	0.284	0.181
Geog. Controls	✓	✓	✓	✓
4-digit industry FE		✓		✓
R^2	0.013	0.151	0.0108	0.0141
Observations	83,633	83,633	83,633	83,633

Notes: The table replicates the results from Table 2 for China and France, computing the dependent variable export intensity at the city-industry level (with 4-digit industries). The table shows results for two transformations of the dependent variable that allow to account for observations with zero export intensity for some city-industry observations. Columns 1 and 2 compute the logarithm of 1 plus export intensity, while columns 3 and 4 compute the inverse hyperbolic sine transformation (IHS). 'Geography Controls' are listed in the note to Table 2 in the paper. Robust standard errors are clustered at the city level. Key: *** significant at 1%; ** 5%; * 10%.

† In the case of specifications 1-2, the average coefficient on city size (x) with export intensity as dependent variable (y) is computed as $(\partial \ln y/\partial \ln x) \equiv \hat{\beta} \times (1+\bar{y}/\bar{y})$, where $\bar{y}$ corresponds to the average export intensity. For specifications 3-4, the average coefficient is computed as $(\partial \ln y/\partial \ln x) \equiv \hat{\beta} \times \sqrt{1+\bar{y}^2}/\bar{y}$ (see Bellemare and Wichman, 2020, for details).

A.7 City Size, House Prices, and Wages

Beyond our main object of interest – higher export intensity in larger cities – our simple model also predicts that larger cities have higher house prices and higher wages in equilibrium.[13] To test these predictions, we estimate regression (2) from the paper with house prices and wages (in logs) as dependent variables.

Data for China comes from Egger et al. (2023) for house prices and from the 2004 Chinese Economic Census of Manufacturing for wages. We use wage data for France from FICUS, and house price data comes from the Centre for Studies and Expertise on Risks, the Environment, Mobility and Urban Planning (Cerema).[14] The French house price data is only available for 2010, while population is from 1999, and wages are measured in 2000.

The results are displayed in Table A.13. We find that house prices and wages are significantly higher in larger cities in both China (cols 1 - 4) and France (cols 1 - 4). These results are robust to the inclusion of our standard set of controls. Thus, the data strongly support the auxiliary predictions of our simple model.

Table A.13: City Population, House Prices, and Wages

Dependent variable: As indicated in table header

	China				France			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	$\log(p_i^H)$	$\log(p_i^H)$	$\log(w_i)$	$\log(w_i)$	$\log(p_i^H)$	$\log(p_i^H)$	$\log(w_i)$	$\log(w_i)$
ln(City Size)	0.305***	0.273***	0.057***	0.073***	0.247***	0.236***	0.043***	0.043***
	(0.036)	(0.032)	(0.021)	(0.021)	(0.025)	(0.023)	(0.006)	(0.007)
Geog. controls		✓		✓		✓		✓
R^2	0.266	0.490	0.030	0.130	0.26	0.43	0.13	0.16
Observations	255	255	329	329	293	293	304	304

Notes: This table examines the relationship between city size and two key economic variables: house prices and wages, for China and France. The dependent variables are $\log(p_i^H)$, the log of the house prices in city i, and $\log(w_i)$, the log of the wages in city i. The regressions are run on the employment zone level for France and the prefecture level for China. See the text in Appendix A.7 for data sources. 'Geog. Controls' are listed in the note to Table 2 in the paper. Robust standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

A.8 Additional Model Validation Results

This appendix complements Section 4.4 in the paper, where we provide empirical evidence supporting our model's novel prediction that the local trade elasticity varies by city size (Proposition

[13]Note that these predictions would also follow from differences in the Pareto *scale* parameter and are thus not a distinguishing feature of our model. In contrast, our novel predictions for export intensity depend on the Pareto *shape* parameter.

[14]Cerema: https://cerema.app.box.com/v/dv3f-indicateurs/folder/210388649273. FICUS: https://www.casd.eu/source/statistique-structurelle-annuelle-dentreprises-issue-du-dispositif-suse/. Note that house price data is missing for some prefectures and some employment zones, so that the corresponding regressions have fewer observations.

5). Here, we use the PNTR shock to Chinese industries to shed empirical light on Proposition 2 (trade shifting population towards larger cities) and Proposition 3 (trade raising nominal wages disproportionately in larger cities). We examine how migration flows, employment, and wages react heterogeneously across city sizes following the PNTR trade shock.

We use data on wages and employment from the Chinese Annual Survey of Manufacturing, which provides data at the prefecture level (as in Section 4.4). To evaluate migration flows we use the 2005 National 1% Population Sampling Survey that contains information on whether an individual was living in a different province five years ago. Note that provinces are significantly larger than prefectures. Migration between provinces accounts for 32.4% of all migration in 2005, with the remainder occurring within provinces (see Table 9 in Chengrong, Lidan, and Mengyao, 2018). Thus, we underestimate the size of migration flows, which will bias our coefficients downwards. Since our outcomes of interest refer to the local labor market, our unit of analysis is a location (prefecture i) rather than prefecture-industry cells (which we used in the analysis of the trade elasticity in Section 4.4).

Migration. To test for heterogeneous effects of the PNTR shocks across city sizes in line with Proposition 2, we estimate the following regression:

$$MigrantShare_{i,2005} = \beta_1 NTRGap_i + \beta_2 \left[NTRGap_i \times ln(pop_i)\right] + \beta_3 ln(pop_i) + \beta_4 X_i + \varepsilon_i \quad \text{(A.4)}$$

where $MigrantShare_{i,2005}$ is the migration inflow into prefecture i over the period 2000-2005, divided by the population of prefecture i in 2005; pop_i is the urban population in prefecture i in 2000, and X_i is a vector of city controls (capital-labor ratio, average distance to other domestic cities, distance to the border, distance to the coast, (all in logs), as well as border and coastal dummies). $NTRGap_i$ is the city-specific exposure to the NTR gap, which is defined as:

$$NTRGap_i = \sum_j \frac{L_{ji}}{L_i} NTRGap_j \quad \text{(A.5)}$$

where $NTRGap_j$ is the NTR gap for sector j, L_{ji} is employment in industry j in location i, and L_i is total employment in location i in 2000. Since we do not have a panel dimension in this setting, we cannot include time or location fixed effects. Note that our dependent variable is a flow variable rather than a stock variable, and our main explanatory variable is the *change* in expected future tariffs. Thus, regression (A.4) can be thought of as an equation in first differences, where β_1 identifies the effect of a decrease in (expected) tariffs on immigration to a location. Since this reflects a trade-induced positive labor demand shock, any Rosen-Roback type model would predict $\beta_1 > 0$. The key coefficient of interest related to Proposition 2 is β_2, where a decrease in variable trade costs should increase immigration particularly strongly into larger cities $\beta_2 > 0$. Intuitively, larger cities have a higher export intensity prior to the reduction in trade costs, such that labor demand increases more following a trade shock, which in turn attracts in-migration.

Table A.14 (Panel A) reports the results from regression (A.4). In columns 1 and 2 we focus exclusively on the level effect (β_1) in order to verify that the standard trade-induced labor-demand shock is present in our data. In line with our expectations, we find a positive and statistically significant effect of trade on immigration flows ($\beta_1 > 0$). These results hold across different specifications: In column 1 we include only the baseline coefficients of interest and in column 2, we add $ln(pop_i)$ together with additional city-level controls. The coefficient becomes smaller in the more restrictive specification but remains highly statistically significant throughout. Next, in columns 3 and 4 we turn to our coefficient of interest. We find that the effect of trade on migration is significantly stronger for larger cities ($\beta_2 > 0$), confirming the prediction from Proposition 2. The coefficient is statistically highly significant and economically meaningful: For a prefecture in the 10th population percentile, a one-standard-deviation higher PNTR gap raises the migrant share by 6%, while this number is 13% for a prefecture in the 90th percentile.

In Panels B and C of Table A.14, we distinguish between skilled (migrants with more than a high school degree) and unskilled migrants (with a high school degree or less). We find that both the main coefficient on the level effect of the trade shock (β_1) as well as the interaction term with city size (β_2) are consistently larger for unskilled migration than for skilled migration. This highlights that the skill composition of cities exposed to trade, and particularly so for larger cities, shifts towards more unskilled labor. While this is beyond the scope of our model, it becomes relevant when we study the wage effects of the PNTR shock below.

Employment. Our results on migration showed that the PNTR trade shock led to a higher (gross) *flow* of labor into larger cities. In what follows, we test the prediction of Proposition 2 with respect to the *stock* of labor, using the evolution of employment around the PNTR shock. In contrast to migration, we observe prefecture employment at various points in time, so that we can use the following panel estimation equation:

$$L_{it} = \beta_1 \left[\text{NTR Gap}_i \times PostPNTR_t\right] + \beta_2 \left[\textbf{\textit{NTRGap}}_i \times PostPNTR_t \times ln(pop_i)\right] + \gamma_i + \gamma_t + \varepsilon_{it} \tag{A.6}$$

where L_{it} is total employment in prefecture i at time $t = 1998, 2000, 2003, 2006,$ and 2007. $PostPNTR_t$ is a dummy that takes on value one for $t \geq 2001$, and *NTRGap*$_i$ is defined as in equation (A.5). Given the panel structure, γ_i and γ_t are prefecture and time fixed effects. The coefficient β_1 identifies the effect of a reduction in expected tariff on employment in the local labor market. As above, we expect this effect to be positive ($\beta_1 > 0$). The distinguishing feature of our model, as formalized in Proposition 2, is the prediction that $\beta_2 > 0$, i.e., a reduction in trade costs leads to a reallocation of population to larger cities.

Table A.14: Effect of Trade Shock on Migration to Cities

Dependent variable: As indicated in panel header

	(1)	(2)	(3)	(4)
Panel A. Dependent variable: Share of migrants$_i$				
$\beta_1 : PNTR\,Gap_j$	0.049***	0.029***	0.047***	0.035***
	(0.010)	(0.009)	(0.010)	(0.009)
$\beta_2 : PNTR\,Gap_j \times ln(pop_i)$			0.044***	0.029***
			(0.010)	(0.009)
Mean Dep. Var.	0.060	0.060	0.060	0.060
Observations	299	299	299	299
Pseudo R^2	0.15	0.37	0.29	0.42
Panel B. Dependent variable: Share of skilled migrants$_i$				
$\beta_1 : PNTR\,Gap_j$	0.0013**	0.0005	0.0005	0.0008
	(0.0006)	(0.0006)	(0.0006)	(0.0005)
$\beta_2 : PNTR\,Gap_j \times ln(pop_i)$			0.0017***	0.0014**
			(0.0006)	(0.0006)
Mean Dep. Var.	0.0026	0.0026	0.0026	0.0026
Observations	299	299	299	299
Pseudo R^2	0.03	0.19	0.17	0.23
Panel C. Dependent variable: Share of unskilled migrants$_i$				
$\beta_1 : PNTR\,Gap_j$	0.047***	0.028***	0.047***	0.034***
	(0.010)	(0.008)	(0.010)	(0.009)
$\beta_2 : PNTR\,Gap_j \times ln(pop_i)$			0.042***	0.028***
			(0.010)	(0.009)
Mean Dep. Var.	0.059	0.059	0.059	0.059
Observations	299	299	299	299
Pseudo R^2	0.15	0.37	0.29	0.42
$ln(pop_i)$		✓	✓	✓
City controls		✓		✓

Note: This table provides evidence for the prediction of Proposition 2 that a trade shock translates into a larger labor demand shock in larger cities ($\beta_2 > 0$). The dependent variable in Panel A is the share of new migrants in Chinese prefectures' total population in 2005, where a migrant is defined as someone who moved from a different province in the last 5 years. Panel B focuses on the share of skilled migrants (with more than a high school degree), and Panel C considers unskilled migrants (with a high school degree or less). Migration data are from the 2005 National 1% Population Sampling Survey. The sample is restricted to prefectures with at least 250 firms. The sample of prefectures is slightly smaller than in other specifications, as a couple of prefectures miss data on migration. All variables are defined in the context of regression (A.4). The regression uses the standardized $NTRGap_i$, with mean zero and standard deviation one. Robust standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

The first two columns in Table A.15 show our results. In column 1, we first estimate a specification without the interaction term, since this level effect on employment is necessary for us to find any meaningful heterogeneity in this effect across city sizes. We find a strong effect of the reduction in expected tariffs on local employment ($\beta_1 > 0$). Introducing the interaction in column 2, we find that these effects are statistically and economically stronger in larger cities, in line with the Prediction of Proposition 2.

Table A.15: Effects of PNTR Gap on Labor Market Outcomes

Dependent variable: As indicated in table header

Dependent Variable:	(1) log(L_i)	(2) log(L_i)	(3) log(w_i)	(4) log(w_i)
$\beta_1 : [Post\ PNTR_t \times NTR\ Gap_i]$	0.139***	0.162***	-0.015	-0.007
	(0.042)	(0.049)	(0.017)	(0.019)
$\beta_2 : [Post\ PNTR_t \times NTR\ Gap_i] \times ln(pop_i)$		0.055***		0.010
		(0.019)		(0012012)
Year FE	✓	✓	✓	✓
City FE	✓	✓	✓	✓
Observations	1,516	1,516	1,516	1,516
Pseudo R^2	0.97	0.97	0.91	0.91

Notes: This table examines the effect of the PNTR trade shock on employment (L_i) in cols 1-2, and one average wages (w_i) across Chinese prefectures in cols 3-4. The sample is restricted to prefectures with at least 250 firms. All variables are defined in the context of regression (A.6). The regression uses the standardized $NTRGap_i$, with mean zero and std one. Robust standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Wages. Lastly, we test whether we can use the exogenous variation from the PNTR shock to provide supporting evidence for Proposition 3 from our model: A reduction in trade cost increases nominal wages in large cities relative to small cities. To do so, we estimate equation (A.6) with average wages in prefecture i at time t as the dependent variable. Columns 3 and 4 in Table A.15 display our results. In column 3, we find that there is no effect of the trade shock on wages. Therefore, it is not surprising that we do not find any heterogeneity in this effect in column 4. The positive and significant effect of the trade shock on employment (col 1) and the lack of an effect on wages (col 3) could be indicative of a highly elastic labor supply (as documented by Imbert, Monras, Seror, and Zylberberg (2025)). This is consistent with the large domestic migration flows in China during this period (Imbert, Seror, Zhang, and Zylberberg, 2022). Our results on migration suggest that an additional mechanism may also have suppressed wage effects of the trade shock: As we showed in Table A.14, immigration into cities was largely unskilled, which will tend to reduce the average wage.

A.9 Additional Robustness for the PNTR Regressions for China

One concern with the PNTR shock is that it led to large inflows of foreign (especially US) firms investing in China for export purposes (Pierce and Schott, 2016). To make sure that this does not drive our results, we estimate regression (23) excluding foreign firms (as classified above in Appendix A.3). We report the results in Panel A in Table A.16, showing that they are very similar to our baseline results from Table 6 in the paper.[15] Panel B shows that the same is true when we exclude also 'state-involved' firms (classified above in Appendix A.3), thus focusing on the particularly restrictive sample that uses only domestic-owned private-sector firms.

Table A.16: Additional Robustness for PNTR Regressions

Dependent variable: $ln(exports_{cjt})$

	(1)	(2)	(3)	(4)
Panel A. Excluding foreign firms†				
$\beta_1 : [Post\ PNTR_t \times PNTR\ Gap_j]$	0.166***	0.154**	0.164*	0.143*
	(0.058)	(0.061)	(0.089)	(0.078)
$\beta_2 : [Post\ PNTR_t \times PNTR\ Gap_j] \times ln(pop_c)$	-0.144***	-0.146***	-0.109*	-0.127**
	(0.046)	(0.046)	(0.058)	(0.053)
Observations	30,953	30,953	30,953	30,948
Pseudo R^2	0.50	0.51	0.52	0.59
Panel B. Only domestic private firms				
$\beta_1 : [Post\ PNTR_t \times PNTR\ Gap_j]$	0.282***	0.282***	0.400**	0.177
	(0.104)	(0.104)	(0.195)	(0.152)
$\beta_2 : [Post\ PNTR_t \times PNTR\ Gap_j] \times ln(pop_c)$	-0.172**	-0.183**	-0.203**	-0.183**
	(0.078)	(0.075)	(0.094)	(0.084)
Observations	14,436	14,436	14,436	14,436
Pseudo R^2	0.61	0.62	0.63	0.68
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
$ln(pop_c)$	✓	✓	✓	(✓)
$ln(pop_c) \times Post\ PNTR_t$	✓	✓	✓	✓
$ln(pop_c) \times PNTR\ Gap_j$	✓	✓	✓	✓
City controls		✓	(✓)	(✓)
City controls $\times Post\ PNTR_t$			✓	✓
City controls $\times PNTR\ Gap_j$			✓	✓
City controls $\times [Post\ PNTR_t \times PNTR\ Gap_j]$			✓	✓
City FE				✓

Notes: The table presents additional evidence for Proposition 5, which predicts that the local trade elasticity is smaller in larger cities. This is captured by the negative coefficient β_2 on the triple interaction. See the note to Table 6 in the paper for the definition of variables. In Panel A, we exclude foreign firms. In Panel B, we only include domestic private firms (i.e., we exclude state-involved, foreign-owned and other firms). The regression uses the standardized $NTRGap_j$, with mean zero and standard deviation one. 'City Controls' include distance to the coast, distance to the border, average distance to other domestic cities, border dummies, coastal dummies, and capital-to-labor ratio (all in logs and standardized). Standard errors clustered at the city and at the industry level in parentheses. Key: *** significant at 1%; ** 5%; * 10%.
† Foreign firms also comprise those with ownership from Hong Kong, Taiwan, and Macao

[15] In the Annual Survey of Manufacturing, we cannot identify processing trade firms (recall that we exclude these in our baseline results, as we can identify them in the Manufacturing Census). Excluding foreign firms allows to ensure that our results are not driven by the rise in processing trade due to the increased offshoring of foreign-owned firms.

B Detailed Derivation of the Simple Model

Consumer problem. To solve the simple model, we begin with the optimization problem of workers/consumers:

$$c_i(x) = \left(\frac{p_i(x)}{P_i}\right)^{-\sigma} c_i \tag{A.7}$$

Since trade is costless within countries, the prices of varieties are equal across locations and so is the price index, which we choose as numeraire:

$$P_i = \left[\int_x p_i(x)^{1-\sigma} dx\right]^{\frac{1}{1-\sigma}} = \left[\int_x p(x)^{1-\sigma} dx\right]^{\frac{1}{1-\sigma}} = P = 1 \tag{A.8}$$

Hence, consumption of each variety by the representative consumer at each location i becomes:

$$c_i(x) = p(x)^{-\sigma} c_i \tag{A.9}$$

Consumer expenditure on housing and the consumption composite is given by

$$c_i = \beta v_i \tag{A.10}$$

$$h_i = (1-\beta)\frac{v_i}{p_i^h} \tag{A.11}$$

where v_i is defined in equation (16).

Firm problem. To produce q units of output a firm of productivity ψ requires the following amount of labor:

$$\ell(\psi) = F_d + \frac{q(\psi)}{\psi} \tag{A.12}$$

Firms face downward-sloping demand:

$$q_i(\psi) = p_i(\psi)^{-\sigma} R = p_i(\psi)^{-\sigma} c \tag{A.13}$$

where R is national level expenditure and $c = \sum_{i \in N} c_i L_i$ is national level consumption. The profit maximization problem for serving the domestic market (π_i^d) is given by:

$$\max \pi_i^d = p_i^d q_i^d(\psi) - w_i l_i(\psi) = p_i^d q_i^d(\psi) - w_i(F_d + \frac{q_i(\psi)}{\psi}) \tag{A.14}$$

which, after substituting in equation (A.13), yields the optimal price as a constant mark-up over marginal costs:

$$p_i^d(\psi) = \frac{\sigma}{\sigma-1}\frac{w_i}{\psi} = \frac{w_i}{\psi\rho} \tag{A.15}$$

where we defined $\rho = {}^{\sigma}/_{\sigma-1}$. Using the expression for optimal prices we get firm-level domestic revenues and profits:

$$r_i^d(\psi) = p_i^d(\psi)q_i^d(\psi) = p_i^d(\psi)^{1-\sigma}R = R\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} \tag{A.16}$$

$$\pi_i^d(\psi) = \frac{r_i^d(\psi)}{\sigma} - w_iF_d = \frac{R}{\sigma}\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} - w_iF_d \tag{A.17}$$

which is (10) in the main text. The formulas for export revenues and export profits can be obtained analogously, noting that the marginal cost faced by firms to serve the foreign market is the same as the marginal cost applicable to domestic sales scaled up by the variable cost of exporting τ, and foreign demand is the same as domestic demand:

$$\max \pi_i^x = p_i^x q_i^x(\psi) - w_i(F_d + \tau\frac{q_i^x(\psi)}{\psi}) \tag{A.18}$$

$$p_i^x(\psi) = \frac{\sigma}{\sigma-1}\frac{\tau w_i}{\psi} = \frac{\tau w_i}{\psi\rho} \tag{A.19}$$

$$r_i^x(\psi) = R\left(\frac{\psi\rho}{w_i\tau}\right)^{\sigma-1} = \tau^{1-\sigma}r_i^d(\psi) \tag{A.20}$$

$$\pi_i^x(\psi) = \frac{r_i^x(\psi)}{\sigma} - w_iF_x = \tau^{1-\sigma}\frac{R}{\sigma}\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} - w_iF_x \tag{A.21}$$

which yields (11) in the main text. Total firm revenues for a firm of productivity ψ at a location i are thus given by

$$r_i(\psi) = \begin{cases} r_i^d(\psi) = R\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} & \text{if firm does not export} \\ r_i^d(\psi) + r_i^x(\psi) = R\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} + R\left(\frac{\psi\rho}{w_i\tau}\right)^{\sigma-1} & \text{if firm exports} \end{cases} \tag{A.22}$$

Industry Equilibrium. In each location, the equilibrium set of active firms is determined by the local zero-profit cut-off conditions for serving the domestic market and the export market, as well as the free entry condition. Together with the labor market clearing condition, these pin down the productivity cut-offs for serving the domestic market (ψ_i^{d*}), the productivity cut-off of serving the export market (ψ_i^{x*}), and the mass of entrants (M_i^e):

$$\pi_i^d(\psi_i^{d*}) = 0 \leftrightarrow \frac{R}{\sigma}\left(\frac{\psi_i^{d*}\rho}{w_i}\right)^{\sigma-1} = w_i F_d \tag{A.23}$$

$$\pi_i^x(\psi_i^{x*}) = 0 \leftrightarrow \tau^{1-\sigma}\frac{R}{\sigma}\left(\frac{\psi_i^{x*}\rho}{w_i}\right)^{\sigma-1} = w_i F_x \tag{A.24}$$

$$\bar{\pi}_i = \int_{\psi_i^{d*}}^{\infty} \pi_i^d(\psi) g_i(\psi) d\psi + \int_{\psi_i^{x*}}^{\infty} \pi_i^x(\psi) g_i(\psi) d\psi = w_i F_e \tag{A.25}$$

$$M_i^e * \bar{l}_i = M_i^e \frac{\bar{r}_i}{w_i} = L_i \tag{A.26}$$

where in equation (A.26) we used the fact that expected profits are zero and hence revenue is equal to total labor payments. (A.26) can be simplified to yield equation (14) as follows: $\bar{r}_i$ is the average revenues of a firm that enters at location i and is given by

$$\bar{r}_i = \int_{\psi_i^{d*}}^{\infty} r_i^d(\psi) g_i(\psi) d\psi + \int_{\psi_i^{x*}}^{\infty} r_i^x(\psi) g_i(\psi) d\psi \tag{A.27}$$

We define a notion of average productivity among firms serving the domestic market and among firms serving the export market, respectively:

$$\tilde{\psi}_i^d = \left[\frac{1}{1 - G_i(\psi_i^{d*})} \int_{\psi_i^{d*}} \psi^{\sigma-1} g_i(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - \sigma + 1}\right)^{\frac{1}{\sigma-1}} \psi_i^{d*} \quad \forall i \tag{A.28}$$

$$\tilde{\psi}_i^x = \left[\frac{1}{1 - G_i(\psi_i^{x*})} \int_{\psi_i^{x*}} \psi^{\sigma-1} g_i(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - \sigma + 1}\right)^{\frac{1}{\sigma-1}} \psi_i^{x} \quad \forall i \tag{A.29}$$

where the second equality imposes the assumption that the productivity distribution in each location follows a standard Pareto with shape parameters α_i.

With the above definitions in place we can rewrite the free entry condition (equation (13)) and the expression for average revenues (equation (A.27)) as

$$\bar{\pi}_i = p_i^d \pi_i^d(\tilde{\psi}_i^d) + p_i^x \pi_i^x(\tilde{\psi}_i^x) = w_i F_e \tag{A.30}$$

$$\bar{r}_i = p_i^d r_i^d(\tilde{\psi}_i^d) + p_i^x r_i^x(\tilde{\psi}_i^x) \tag{A.31}$$

where $p_i^d = 1 - G_i(\psi_i^{d*})$ and $p_i^x = 1 - G_i(\psi_i^{x*})$ represent the probability of serving the domestic and the foreign market, respectively.

Note that with constant mark-ups over marginal costs we have that

$$\pi_i^d(\psi) = \frac{r_i^d(\psi)}{\sigma} - w_i F_d \qquad \forall \psi \tag{A.32}$$

$$\pi_i^x(\psi) = \frac{r_i^x(\psi)}{\sigma} - w_i F_x \qquad \forall \psi \tag{A.33}$$

Combining equations (A.30), (A.31), (A.32) and (A.33) yields the following expression for average firm revenues:

$$\bar{r}_i = \sigma \left[\bar{\pi}_i + p_i^d w_i F_d + p_i^x w_i F_x \right] = \sigma \left[w_i F_e + p_i^d w_i F_d + p_i^x w_i F_x \right] \tag{A.34}$$

Plugging (A.34) into (A.26) yields the expression for M_i^e outlined in equation (A.109). Further, noting that $1 - G_i(\psi) = \psi^{-\alpha_i}$ we obtain the expression for the mass of entrants at each location from equation (14).

As in Melitz (2003), the zero profit conditions outlined in (12) can be rewritten in terms of the profit levels of average productivity firms

$$\pi_i^d(\tilde{\psi}_i^d) = w_i F_d k_i(\psi_i^{d*}) \tag{A.35}$$

$$\pi_i^x(\tilde{\psi}_i^x) = w_i F_x k_i(\psi_i^{x*}) \tag{A.36}$$

Where the function $k_i(.)$ is given by

$$k_i(\psi_i^{d*}) = \left(\frac{\tilde{\psi}_i^d}{\psi_i^*} \right)^{\sigma-1} - 1 = \frac{\sigma - 1}{\alpha_i - \sigma + 1} \tag{A.37}$$

$$k_i(\psi_i^{x*}) = \left(\frac{\tilde{\psi}_i^x}{\psi_i^{x*}} \right)^{\sigma-1} - 1 = \frac{\sigma - 1}{\alpha_i - \sigma + 1} \tag{A.38}$$

For the purposes of the proofs below, and also to keep closer to Melitz (2003) we introduce notation for the measure of "active" firms (i.e. firms that have not only entered by paying the fixed entry cost F_e but also serve at least one market) and the fraction of active firms that engage in exporting. These are given by:

$$M_i = [1 - G_i(\psi_i^{d*})] M_i^e \tag{A.39}$$

$$\chi_i^* = \frac{p_i^x}{p_i^d} = \frac{1 - G_i(\psi_i^{x*})}{1 - G_i(\psi_i^d *)} = \left(\frac{\psi_i^{d*}}{\psi_i^{x*}} \right)^{\alpha_i} = \left[\frac{1}{\tau} \left(\frac{F_d}{F_x} \right)^{\frac{1}{\sigma-1}} \right]^{\alpha_i} \tag{A.40}$$

where it is important to note that our parametric restriction $\tau^{\sigma-1} F_x > F_d$ ensures that there is no firm that exports without serving the domestic market such that exporting firms are a strict subset of firms serving the domestic market, as in Melitz (2003). It is also important to note that the fraction of firms engaged in exporting is decreasing in the shape parameter of the city-specific Pareto productivity distribution: In equation (A.40), the expression in square brackets is smaller than 1 so that higher α_i values are associated with a smaller fraction of active firms exporting.

Housing and land markets. Market clearing in the housing market requires that expenditure on housing equals the revenues of perfectly competitive housing developers

$$(1-\beta)v_i L_i = p_i^h H_i \tag{A.41}$$

where $v_i = \frac{w_i}{\beta}$ is as derived in equation (16). In turn, land market clearing requires spending on land by housing developers to equal the income of landowners (which is then fully taxed by local governments):

$$\gamma p_i^h H_i = r_i N \tag{A.42}$$

Substituting for $p_i^h H_i$ from equation (A.41) into equation (A.42), substituting for $v_i = \frac{w_i}{\beta}$ and rearranging gives us the expression for land rents in equation (18).

C Model with Endogenous Agglomeration

In this appendix, we describe a version of our model that includes endogenous agglomeration in the tail of the firm productivity distribution. We proceed in three steps. First we set up a version of our simple model, augmented to include agglomeration forces that depend on the size of the local population. We show that our key theoretical result linking export intensity and city size also holds in this setup. We then provide a microfoundation for these agglomeration forces that affect the upper tail of the firm productivity distribution. We show how a thicker productivity upper tail in large cities can emerge from embedding a learning mechanism in a standard firm growth process. Finally, we discuss the calibration process for the model featuring agglomeration forces in the tail of the firm productivity distribution, providing the underlying details for the results presented in Table 8 in the main text.

C.1 Introducing Agglomeration Forces in our Simple Model

For reasons of analytical tractability, in our simple model (Section 4) we model the productivity differences across locations as fully exogenous locational fundamentals. A large literature, however, emphasizes the important role of agglomeration economies in driving differences in productivity across locations (e.g. Bleakley and Lin, 2012). To link our theoretical work to this wider literature, we develop a model extension where differences in the shape of firm productivity distributions across locations arise due to agglomeration economies as well as differences in locational fundamentals. We show that our key theoretical result, which predicts a positive relationship between city size and export intensity, holds in this extended framework.

We follow the setup of our simple model described in Section 4.1. Recall that in each location, productivity is distributed Pareto with the corresponding cumulative distribution function $G_i(\psi) = 1 - \left(\frac{A_i}{\psi}\right)^{\alpha_i}$, for $\psi \geq A_i$, where α_i is the location-specific productivity shape parameter and A_i is the scale parameter. We now allow both location-specific natural advantage and local agglomeration economies to affect A_i and α_i:

$$A_i(L_i) = \bar{A}_i + a(L_i) \tag{A.43}$$

$$\alpha_i(L_i) = \bar{\alpha}_i + d(L_i) \tag{A.44}$$

where the parameters $\bar{A}_i$ and $\bar{\alpha}_i$ capture location-specific natural advantage operating on the productivity shifters and dilation parameters respectively, while the functions $a(\cdot)$ and $d(\cdot)$ capture agglomeration economies operating on the shifters. We assume that both $a(\cdot)$ and $d(\cdot)$ are monotonically increasing and concave ($a'(\cdot) > 0$, $d'(\cdot) > 0$, $a''(\cdot) < 0$, $d''(\cdot) < 0$).

This specification nests our simple model from Section 4 with $a(\cdot) = d(\cdot) = 0$. We now consider the polar opposite case with no differences in exogenous productivity and only agglomeration economies: $\bar{A}_i = \bar{\alpha}_i = 0$ and $a(L_i), d(L_i)$ increasing functions of L_i and nonzero for any positive L_i (a setup more closely related to the one studied by Combes and Overman (2004)). It is possible to show that our key theoretical prediction, which rationalizes our stylized fact, holds in this alternative setup. Formally:

Proposition A.1. *In a model featuring agglomeration economies of the form $A_i = a(L_i)$ and $\alpha_i = d(L_i)$, in any equilibrium featuring cities of asymmetric sizes it will be the case that*

(i) Larger cities have higher average productivity and higher export intensity.

(ii) The higher export intensity of larger cities is driven by the proportion of exporters among local firms (i.e., by the extensive as opposed to the intensive margin of exporting).

Proof: See Appendix D.7.

C.2 Microfoundation

In this section we provide a microfoundation for the agglomeration forces introduced in the previous section. We show that a learning mechanism – by which firms in larger cities benefit from a greater pool of local experts – naturally yields a Pareto steady state firm distribution where the location parameter is inversely related to city size. Such a mechanism can produce agglomeration forces of the type described in Section C.1. Our (static) spatial equilibrium framework can be understood as describing the steady state of the process described below.

Learning is widely recognized as a fundamental source of agglomeration economies. For instance, Glaeser et al. (1992) document that the proximity of firms in cities facilitates knowledge spillovers, while Rosenthal and Strange (2004) show that local interactions boost productivity and innovation. Similarly, Duranton and Puga (2004) emphasize that localized knowledge spillovers are a critical driver of urban productivity. Our framework builds on these insights and demonstrates that enhanced learning opportunities in larger cities not only raise *average* productivity but can also deliver a thicker upper tail in firm performance.

We follow Gabaix (2009) and build on a simple model of firm growth obeying Gibrat's Law.[16] In this setting, firm productivity grows in each period by a deterministic component and a random component that are both proportional to initial productivity. We adapt this framework to our urban economics setting by introducing an agglomeration force into this otherwise standard growth process. Formally, we assume that the law of motion of firm productivity φ_t in city i, is given by:

$$d\varphi_t = \left(a_{i,z} + \eta \cdot a_{L_i}(N_{i,t})\right) \varphi_t dt + \nu \varphi_t dB_t \qquad \text{with } \nu > 0 \text{ and } \eta \in \{0,1\} \tag{A.45}$$

[16] Gibrat's Law states the empirical regularity that the distribution of the percentage growth rate of firms is independent of their size.

where B_t is a Brownian motion and $a_{i,z} \geq 0$ is a (potentially) city-specific exogenous growth factor; $a_{L_i}(N_{i,t})$ is an endogenous growth component that depends on the number of experts $N_{i,t}$ that each firm in city i at time t chooses to meet and learn from. $N_{i,t}$ comprises individuals in city i with technology and business expertise that can help to improve a firm's productivity. The term $a_{L_i}(N_{i,t})$ thus formalizes the learning externality that underpins agglomeration in our setting. The parameter η governs the relative importance of the exogenous productivity differences vis-à-vis the agglomeration force. In line with Gabaix (2009), we assume that firms die with an exogenous exit hazard $\delta > 0$ and are born at the minimum possible value of $\varphi_{Min} > 0$, which acts as the lower bound of the productivity growth process. This assumption ensures that the productivity process does not explode.

Our approach nests the original Gabaix (2009) growth process with no agglomeration forces ($\eta = 0$) and no differences across cities ($a_{i,z} = a_z$), as well as the case of purely exogenous differences in productivity across cities ($\eta = 0$ and $a_{i,z} > 0$ varying across cities), as in our simple model in Section 4.[17] In what follows, we focus on the case with $\eta > 0$, where agglomeration forces give rise to an endogenous local growth component.

As outlined above, a firm in city i can improve its productivity through meeting and learning from experts. However, meetings with experts are costly. In every period t, firms optimally choose how many experts $N_{i,t}$ to meet, trading off the costs and benefits of meetings. Firms choose the number of meetings that maximizes productivity growth in each period t:[18]

$$\max_{N_{i,t}} a_{L_i}(N_{i,t}) = g(N_{i,t}) - N_{i,t}\frac{c}{L_{i,t}} \tag{A.46}$$

The productivity gains associated with meetings are given by $g(N_{i,t}) \in C^2(\mathbb{R})$, where $g(0) = 0$ $g' > 0$ and $g'' < 0$, such that there are decreasing marginal returns to meetings. This formalizes the intuition that there is some overlap between what can be learned from different experts. Importantly, the cost of meetings decreases in local population $L_{i,t}$ and is given by: $\frac{c}{L_{i,t}}$ with $c > 0$. This formalizes the idea that it is easier to find experts in larger cities. Each firm takes city size $L_{i,t}$ as given. Given these differences in the cost of meetings between larger and smaller cities, the optimal choice of meetings, and hence the strength of the learning externality, increases with city size. In turn, this stronger learning externality in larger cities yields steady state firm productivity distributions that feature thicker upper tails. Formally, we can show that:

Proposition A.2. *If the endogenous growth rate of firm productivity is bounded above by firms'*

[17] Higher $a_{i,z}$ in equation (A.45) translates into a thicker upper tail in the stationary distribution. This follows from a simplified version of the proof outlined below, with $\eta = 0$.

[18] Note that since firm profits are a monotone function of productivity, maximizing productivity growth is isomorphic to maximizing profits. Further note that the cost of meetings is also paid in terms of productivity growth, e.g., time spent on meeting outside experts cannot be spent on in-house productivity improvements.

death probability (i.e., $a_{L_i}(N_{i,t}) < \delta$)[19] *there exists a steady state of the process described by equations (A.45) and (A.46) which is characterized by a Pareto firm productivity distribution with shape parameter $\alpha_i > 1$. Moreover, in steady state, the shape parameter of the stable Pareto distribution is negatively related to city size.*

Proof: See Appendix D.8.

Intuitively, with a higher number and density of experts in larger cities, meetings become less costly. In turn, cheaper meetings in larger cities promote faster learning by firms operating in these cities, which results in these locations featuring a higher number of very productive firms (i.e., a thicker upper tail of firm productivity) in steady state. This mechanism provides a microfoundation for the core assumption in our model from Section C.1 which posits agglomeration economies that affect the dilation terms of the city-level firm productivity distributions.[20]

C.3 Calibration of the Model with Endogenous Shape Parameters

In this section, we describe the calibration of our quantitative model with endogenous, city-specific productivity shape parameters. This differs from the baseline quantitative model, which keeps the values of the productivity Pareto shape parameter fixed for each city in the counterfactual exercises. In contrast, we now allow the location-specific shape parameters to change as cities grow or shrink, as formalized in Section C.1.

The learning and matching mechanism presented above in Appendix C.2 predicts a stable Pareto distribution of firm productivity with shape parameter $\alpha_i > 1$, which declines with city size. In the calibration, we treat the observed firm productivity distribution in each city as the stable distribution that emerges from the growth process in Appendix C.2. Note that the microfoundation above does not yield a closed-form solution for the resulting productivity shape parameter and hence does not provide a functional form for α_i declining in city size. We therefore follow the large literature on agglomeration economies and estimate a log-linear specification:[21]

$$\tilde{\alpha}_i = \beta_0 + \beta_1 \ln L_i + \varepsilon_i \tag{A.47}$$

where $\tilde{\alpha}_i$ is the city-specific shape parameter estimated through the QQ regressions described in Appendix (E.3), and L_i is the population in city i. This specification provides a good fit: As shown in Figure 2 in the paper, the slope coefficient β_1 is statistically highly significant and negative.

[19] Since $\bar{L}$ bounds the number of meetings a sufficient condition is $a_{L_i}(\bar{L}) < \delta$

[20] In principle, one could also integrate this productivity process in a dynamic version of our model. We leave this to future work, given that our core contribution in this paper is empirical in nature.

[21] As Combes and Gobillon (2015) point out in their review article, "most articles estimate a log-linear relationship between local outcome and local characteristics." We follow this approach and apply it to agglomeration affecting the tail rather than the mean of the firm productivity distribution. For recent papers using both reduced-form and structural approaches for the estimation of agglomeration elasticities, see, for example, Allen and Arkolakis (2014), Kline and Moretti (2014), or Roca and Puga (2017).

According to our microfoundation in Appendix C.2, the resulting city-specific shape parameter has two components: i) differences across cities that are exogenous to our learning-based microfoundation, such as sorting of firms into locations (linked to $a_{i,z}$ in equation (A.45)), and ii) differences in the learning-based agglomeration component (linked to a_{L_i}). Note that both components are reflected in the estimated correlation β_1 between the shape parameter and city size. For our calibration, we thus need to choose how much of β_1 is due to the endogenous agglomeration component. We build on the work by Gaubert (2018) suggesting that half of the correlation between productivity and city size is driven by sorting, while the other half is driven by agglomeration.[22] Accordingly, our calibration adjusts the shape parameter as follows when city size changes from $L_{i,0}$ to $L_{i,1}$:

$$\hat{\alpha}_{i,1} = \tilde{\alpha}_{i,0} + 0.5 \times \hat{\beta}_1 \triangle \ln L_i \tag{A.48}$$

where $\triangle \ln L_i = \ln L_{i,1} - \ln L_{i,0}$ is the change in city size (e.g., due to trade liberalization) and $\tilde{\alpha}_{i,0}$ is the original shape parameter in the equilibrium before the counterfactual.

The corresponding results for the trade counterfactual are shown in Table 8, row 6 in the paper. There, the city-specific shape parameters change with city size according to the "50% endogenous-agglomeration-rule" in equation (A.48). Alternative approaches with lower (higher) endogenous agglomeration percentages yield somewhat smaller (larger) welfare and inequality effects of trade liberalization. However, the magnitude of results is broadly conserved (results available upon request).

D Proofs

D.1 Lemma 1

Substituting equations (A.35), (A.36) into (A.30), noting equations (A.37), (A.38) and (22) and solving for the entry threshold for serving the domestic market we can obtain a closed form solution for these thresholds for all locations i:

$$\psi_i^{d*} = \left\{ \frac{1}{F_e} \frac{\sigma - 1}{\alpha_i - \sigma + 1} \left\{ F_d + \left[\frac{1}{\tau} \left(\frac{F_d}{F_x} \right)^{\frac{1}{\sigma - 1}} \right]^{\alpha_i} F_x \right\} \right\}^{\frac{1}{\alpha_i}} \tag{A.49}$$

It is straightforward to check that the RHS expressions of equation (A.49) is strictly decreasing in α_i which yields the result that cities with lower shape parameters have higher productivity thresholds for entry into serving the domestic market. Moreover, noting the relationship between

[22] Gaubert (2018) develops a structural model of firm sorting and agglomeration in order to structurally disentangle the role of these two forces for the productivity differences between smaller and larger cities. She estimates the model on French firm-level data using a simulated method of moments approach and finds that both forces account for roughly half of the productivity advantage of large cities.

the domestic entry threshold and average productivity revealed by (A.28), as well as the similar relationship for exporters (see (A.29)) and the relationship between the entry thresholds for the domestic and export markers (in equation (22)) completes our proof that locations with lower shape parameters will have higher average firm productivity in equilibrium (as it has higher productivity of both domestic firms and exporters, and a higher share of exporters).

To prove the population claim in Lemma 1, let us consider an arbitrary pair of cities i and i' with city i assumed more productive than city i'. Let us first consider the zero-profit conditions of firms serving the domestic market at each location (see equation (12)). Combining these zero-profit conditions with the expression of domestic profits outlined in equation (A.32) yields the following expressions for the revenues of firms found at the entry threshold of the domestic market in each location:

$$r_i^d(\psi_i^{d*}) = \sigma w_i F_d \quad \text{and} \quad r_{i'}^d(\psi_{i'}^{d*}) = \sigma w_{i'} F_d \tag{A.50}$$

Using the relationship

$$r_i^d(\psi) = R\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1},$$

we can derive the following expressions for locations i and i'

$$R(\psi_i^{d*}\rho)^{\sigma-1} = \sigma w_i^{\sigma} F_d \tag{A.51}$$
$$R(\psi_{i'}^{d*}\rho)^{\sigma-1} = \sigma w_{i'}^{\sigma} F_d \tag{A.52}$$

where again, ρ is given by $\rho = (\sigma-1)/\sigma$. Dividing (A.51) by (A.52) yields after some rearranging

$$\frac{w_i}{w_{i'}} = \left(\frac{\psi_i^{d*}}{\psi_{i'}^{d*}}\right)^{\frac{\sigma-1}{\sigma}} \tag{A.53}$$

Given our previous results, this establishes that more productive cities have higher entry productivity thresholds, which further leads to the conclusion that more productive cities display higher wages in equilibrium. Moreover, from the spatial equilibrium conditions we can write:

$$\frac{v_i}{v_{i'}} = \frac{w_i}{w_{i'}} = \left(\frac{r_i}{r_{i'}}\right)^{\gamma(1-\beta)} \tag{A.54}$$

Substituting for r_i via equation (18) yields after some manipulation

$$\frac{L_i}{L_{i'}} = \left(\frac{w_i}{w_{i'}}\right)^{\frac{1-\gamma(1-\beta)}{\gamma(1-\beta)}} = \left(\frac{\psi_i^{d*}}{\psi_{i'}^{d*}}\right)^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}} \tag{A.55}$$

which, given the result established above that more productive cities will have higher wages, also implies that more productive cities will feature larger populations in equilibrium. From equations (17) and (18), it follows immediately that higher wages and populations imply higher house prices:

$$\frac{p_i^H}{p_{i'}^H} = \left(\frac{w_i L_i}{w_{i'} L_{i'}}\right)^{\gamma} \tag{A.56}$$

D.2 Proposition 1

Regarding the first claim of Proposition 1, we can write the export intensity of a given location i as

$$Expint_i = \frac{\int_{\psi_i^{x*}}^{\infty} r_i^x(\psi) M_i \mu_i(\psi) d\psi}{\int_{\psi_i^{d*}}^{\infty} r_i^d(\psi) M_i \mu_i(\psi) d\psi + \int_{\psi_i^{x*}}^{\infty} r_i^x(\psi) M_i \mu_i(\psi) d\psi} \tag{A.57}$$

where $\mu_i(.)$ denotes the endogenous productivity distribution given by

$$\mu_i(\psi) = \begin{cases} 0 \quad if \quad \psi \leq \psi_i^{d*} \\ \frac{1}{1-G_i(\psi_i^{d*})} g_i(\psi) \quad if \quad \psi > \psi_i^{d*} \end{cases} \tag{A.58}$$

Substituting for $r_i^x(\psi)$ and $r_i^d(\psi)$ in equation (A.57) and simplifying yields:

$$Expint_i = \frac{\tau^{1-\sigma} \chi_i^* \tilde{\psi}_i^x}{\tilde{\psi}_i^d + \tau^{1-\sigma} \chi_i^* \tilde{\psi}_i^x} \tag{A.59}$$

where χ_i^* is the share of firms that export and is given by equation (A.40).

Substituting $\tilde{\psi}_i^d$ and $\tilde{\psi}_i^x$ from (A.28) and (A.29) and using equation (22), equation (A.59) further simplifies to

$$Expint_i = \frac{\chi_i^* F_x}{F_d + \chi_i^* F_x} \tag{A.60}$$

The expression on the RHS of (A.60) can be shown to be increasing in the fraction of exporting firms located at each location i (χ_i^*), which is defined in (A.40). Hence, given that more productive cities have already been shown to have a higher share of firms that export, they will also have higher export intensities.

Moreover, the expression in (A.60) shows that the fraction of exporting firms χ_i^* fully drives the cross-city variation in export intensity, which establishes the second claim of Proposition 1. This completes our proof of Proposition 1.

D.3 Proposition 2

To prove this proposition it will again be useful to consider the case of two arbitrary locations i and i' with location i assumed more productive $\alpha_i < \alpha_{i'}$.

Population. First let us consider Proposition 2's prediction regarding movements in relative population. From equation (A.55) we can write that:

$$sgn\left(\frac{\partial}{\partial \tau}\frac{L_i}{L_{i'}}\right) = sgn\left(\frac{\partial}{\partial \tau}\frac{\psi_i^{d^*}}{\psi_{i'}^{d^*}}\right) \tag{A.61}$$

Substituting for $\psi_i^{d^*}$ and $\psi_{i'}^{d^*}$ via equation (A.49) and differentiation yields:

$$\frac{\partial}{\partial \tau}\frac{\psi_i^{d^*}}{\psi_{i'}^{d^*}} = \frac{\psi_i^{d^*}\psi_{i'}^{d^*}F_xF_d}{(\psi_{i'}^{d^*})^2\tau(F_d + p_i^xF_x)(F_d + p_{i'}^xF_x)}(p_{i'}^x - p_i^x) < 0 \tag{A.62}$$

This establishes that, as trade liberalizes (i.e., τ falls), the relative productivity threshold for successful entry shifts in favor of more productive cities; consequently, their relative population shares also rise. Given that this result was established for an arbitrary pair of cities, this implies that international trade liberalization will be associated with a general shift in population to more productive (and larger) cities.

Aggregrate TFP and wages. Moving on to the implications of trade liberalization to aggregate (country-level TFP), we define aggregate TFP as

$$TFP_{Agg} = \frac{Q}{L} = \frac{R}{L} \tag{A.63}$$

Where Q denotes total output of the tradable composite good, $R = \sum_k w_kL_k$ denotes aggregate revenues/ expenditure, and the second equality results from our choice of the tradable composite good as the numeraire. Hence, proving that trade liberalization results in an increase in aggregate TFP is equivalent to showing that it results in an increase in aggregate revenues. To show this, we first consider the effect of trade liberalization on wages at each location i. From equation (A.55) and the national labor market clearing condition in equation (21), we can derive the following expression for equilibrium population at location i

$$L_i = \frac{\left(\psi_i^{d^*}\right)^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}{\sum_{k=1}^{N}\left(\psi_k^{d^*}\right)^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}L \tag{A.64}$$

Further, using equation (A.51), substituting $R = \sum_k w_kL_k$ and then each L_k using equation (A.64) and (A.53), we can derive the following expression for wages at each location i

$$w_i = \left(\frac{L}{\sigma F_d}\right)^{\frac{1}{\sigma-1}} \rho(\psi_i^{d*})^{\frac{\sigma-1}{\sigma}} \left(\frac{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}}{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}\right)^{\frac{1}{\sigma-1}} \tag{A.65}$$

Differentiation with respect to τ yields:

$$\begin{aligned}\frac{\partial w_i}{\partial \tau} &= \left(\frac{L}{\sigma F_d}\right)^{\frac{1}{\sigma-1}} \rho\frac{\sigma-1}{\sigma}(\psi_i^{d*})^{\frac{-1}{\sigma}}\frac{\partial \psi_i^{d*}}{\partial \tau} \left(\frac{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}}{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}\right)^{\frac{1}{\sigma-1}} \\ &+ \left(\frac{L}{\sigma F_d}\right)^{\frac{1}{\sigma-1}} \rho(\psi_i^{d*})^{\frac{\sigma-1}{\sigma}} \left(\frac{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}}{\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}\right)^{\frac{1}{\sigma-1}-1} \frac{1}{\sigma-1} \\ &\times \frac{\left[\sum_{i=1}^{N}\sum_{j=1}^{N}(\psi_i^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}(\psi_j^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}\left(\frac{1}{\psi_j^{d*}}\frac{\sigma-1}{\sigma}\frac{\partial \psi_j^{d*}}{\partial \tau}\right)\right]}{\left[\sum_{k=1}^{N}(\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}\right]^2} < 0\end{aligned}$$

where the final inequality follows as the partial derivatives of the productivity thresholds are negative. We have shown that trade liberalization increases wages in all locations, while also redistributing labor towards the most productive, highest-wage cities. We can therefore write, for any trade liberalization:

$$TFP_{Agg}^{BL} = \frac{\sum_{k=1}^{N} w_k^{BL} L_k^{BL}}{L} \leq \frac{\sum_{k=1}^{N} w_k^{BL} L_k^{AL}}{L} < \frac{\sum_{k=1}^{N} w_k^{AL} L_k^{AL}}{L} = TFP_{Agg}^{AL} \tag{A.66}$$

where the superscripts BL and AL denote variables before and after liberalization, the first inequality uses our labor redistribution results, keeping wages constant, and the second uses our result about the impact of liberalization on wages in all locations. We thus conclude that trade liberalizations are associated with an increase in aggregate productivity.

Welfare. To analyze the effect of trade liberalization on welfare, note that solving for the indirect utility function at each location (by combining equations (15), (17) and (18) with the fact that $v_i = \frac{w_i}{\beta}$) yields:

$$U_i = \bar{U} = \left[\frac{(1-\beta)(1-\gamma)}{\beta}\right]^{(1-\gamma)(1-\beta)} N^{\gamma(1-\beta)} \frac{w_i^{1-\gamma(1-\beta)}}{L_i^{\gamma(1-\beta)}} \tag{A.67}$$

Let us consider then the welfare of workers in the lowest productivity city

$$U_{min} = \left[\frac{(1-\beta)(1-\gamma)}{\beta}\right]^{(1-\gamma)(1-\beta)} N^{\gamma(1-\beta)} \frac{w_{min}^{1-\gamma(1-\beta)}}{L_{min}^{\gamma(1-\beta)}} = \bar{U} \tag{A.68}$$

We know that for this city, trade liberalization is associated with an increase in wages, as wages in all locations increase, and a decrease in population (as this city shrinks relative to all other cities). Hence it must be the case that welfare of workers in this city rises after trade liberalization. Given the requirements of spatial equilibrium, that must also imply that welfare rises everywhere in the aftermath of trade liberalization.

D.4 Proposition 3

As established in equation (A.53), relative wages in equilibrium between two locations are given by

$$\frac{w_i}{w_{i'}} = \left(\frac{\psi_i^{d*}}{\psi_{i'}^{d*}}\right)^{\frac{(\sigma-1)}{\sigma}}$$

Moreover, trade liberalization moves the relative entry thresholds in the direction of the initially more productive city (i.e. the city with the lower shape parameter $\alpha_i < \alpha_{i'}$):

$$\frac{\partial}{\partial \tau} \frac{\psi_i^{d*}}{\psi_{i'}^{d*}} = \frac{\psi_i^{d*} \psi_{i'}^{d*} F_x F_d}{(\psi_{i'}^{d*})^2 \tau (F_d + p_i^x F_x)(F_d + p_{i'}^x F_x)} (p_{i'}^x - p_i^x) < 0$$

Taken together, these two results indicate that the nominal wage inequality between any pair of two cities increases in the aftermath of trade liberalization (as we know from Proposition 1 that the larger city starts out with higher wages and from the two equations reproduced in this section that nominal wages in larger and more productive cities grow further as a result of trade liberalization). In turn, this implies an increase in regional inequality in terms of nominal wages, as initially high-wage locations grow their nominal wages further relative to low nominal wage locations. Note that in our model this increase in regional wage inequality is inconsequential from a welfare perspective, as spatial equilibrium ensures the absence of welfare inequality across space both before and after any trade liberalization.

Changes in aggregate nominal wage inequality will be driven by these changes in regional inequality and population movements and are ambiguous. Aggregate welfare inequality is guaranteed to remain unchanged in response to trade liberalization in our setting.

D.5 Proposition 4

To see the effects of an increase in housing supply elasticity on export intensity and aggregate TFP, it is important to note from equations (A.49), (A.40) and (A.60) that changing the housing supply elasticities will not affect the entry thresholds, probability of exporting or export intensities of any locations. Thus any effect on aggregate export intensity or aggregate productivity will result simply from reallocating population, production and expenditure across domestic locations.

Country-level export intensity. We can write country-level export intensity as:

$$Expint = \sum_{i=1}^{N} \frac{w_i L_i}{\sum_{k=1}^{N} w_k L_k} Expint_i \tag{A.69}$$

Note that using (A.65) and (A.64) we can write:

$$w_i L_i = \left(\frac{L}{\sigma F_d}\right)^{\frac{1}{\sigma-1}} \rho L (\psi_i^{d*})^{\frac{\sigma-1}{\gamma\sigma(1-\beta)}} \frac{\left[\sum_{k=1}^{N} (\psi_k^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}\right]^{\frac{1}{\sigma-1}}}{\left[\sum_{k=1}^{N} (\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}\right]^{\frac{\sigma}{\sigma-1}}} \tag{A.70}$$

For any pair of cities i and i', with i more productive, we can therefore obtain:

$$\frac{\partial}{\partial \gamma}\left(\frac{w_i L_i}{w_{i'} L_{i'}}\right) = \frac{\ln(\psi_{i'}^{d*}) - \ln(\psi_{i'}^{d*})}{[(\psi_{i'}^{d*})^{\frac{\sigma-1}{\gamma\sigma(1-\beta)}}]^2} < 0 \tag{A.71}$$

Which implies that relaxing housing supply constraints (i.e., reducing γ) increases the revenue share of more productive, high export-intensity cities, and thus increases aggregate (i.e., country-level) export intensity.

Aggregate TFP. As in Proposition 2, to determine the impact of changing housing supply elasticity on aggregate TFP, it suffices to determine the impact of relaxing housebuilding constraints (i.e., lowering γ) on aggregate revenues. In turn, we can write aggregate revenues using equation (A.70):

$$R = \sum_{k=1}^{N} w_k L_k = \left(\frac{L}{\sigma F_d}\right)^{\frac{1}{\sigma-1}} \rho L \left[\underbrace{\frac{\sum_{k=1}^{N} (\psi_k^{d*})^{\frac{(\sigma-1)}{\gamma\sigma(1-\beta)}}}{\sum_{k=1}^{N} (\psi_k^{d*})^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}}}_{E}\right]^{\frac{\sigma}{\sigma-1}} \tag{A.72}$$

It is easy to see that

$$sgn(\frac{\partial R}{\partial \gamma}) = sgn(\frac{\partial E}{\partial \gamma}) \tag{A.73}$$

By differentiating E with respect to γ, we can show that

$$\frac{\partial E}{\partial \gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}[\ln(\psi_i^{d*}) - \ln(\psi_j^{d*})][(\psi_i^{d*})(\psi_j^{d*})]^{\frac{(\sigma-1)[1-\gamma(1-\beta)]}{\gamma\sigma(1-\beta)}}[(\psi_j^{d*})^{\frac{\sigma-1}{\sigma}} - (\psi_i^{d*})^{\frac{\sigma-1}{\sigma}}] \times EXP < 0 \tag{A.74}$$

where EXP is a strictly positive expression. This establishes that decreases in γ - i.e., increases in housing supply elasticity $(1-\gamma)/\gamma$ - result in an increase in aggregate revenues and hence an increase in aggregate TFP. This completes our proof of Proposition 4.

D.6 Proposition 5

Note that we can write the city-level exports of location i as:

$$Exp_i = \int_{\psi_i^x}^{\infty} r_i^x(\psi) M_i \mu_i(\psi) d\psi \tag{A.75}$$

Similarly to Chaney (2008) we can write

$$-\frac{\tau}{Exp_i}\frac{\partial Exp_i}{\partial \tau} = -\frac{\tau}{Exp_i}\left(\int_{\psi_i^x}^{\infty}\frac{\partial r_i^x(\psi)}{\partial \tau} M_i \mu_i(\psi) d\psi\right) + \frac{\tau}{Exp_i}\left(M_i r_i^x(\psi_i^{x*})\mu_i^x(\psi_i^{x*})\right) \tag{A.76}$$

where we have a separation of the intensive and extensive margin responses to a reduction in trade costs like in Chaney (2008). Note that we can write

$$r_i^x(\psi) = R\left(\frac{\psi\rho}{w_i\tau}\right)^{\sigma-1} \tag{A.77}$$

$$\frac{\partial r_i^x(\psi)}{\partial \tau} = R\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1}[-(\sigma-1)]\tau^{\sigma} = -(\sigma-1)\frac{r_i^x(\psi)}{\tau} \tag{A.78}$$

Where when moving from (A.77) to (A.78), we made use of the small open economy assumption (each city and country is too small for its trade relationship to affect aggregate outcomes such as aggregate revenues at the country level or wages for any location). Plugging (A.78) into the first term of (A.76) yields that the intensive margin elasticity is, as in Chaney (2008) given by $(\sigma-1)$. Moreover, similarly to Chaney (2008) we have that

$$\frac{\partial \psi_i^{x*}}{\partial \tau} = \frac{\psi_i^{x*}}{\tau} \tag{A.79}$$

and under the assumption that ψ_i^{d*} does not change, the second term of equation (A.76) collapses to $\alpha_i - (\sigma - 1)$. Thus the trade elasticity for each city is given by:

$$-\frac{\partial log(Exp_i)}{\partial log\tau} = \alpha_i \tag{A.80}$$

Since we established in Lemma 1 and Proposition 1 that in equilibrium, locations with lower tail parameters α_i will feature higher populations, it follows that the trade elasticity is decreasing with population size.

Decomposing trade flows into the number of exporters and the average export value of exporting firms:

$$log(Exp_i) = log(M_i^e(1 - G(\psi_i^x)) + log(\underbrace{\int_{\psi_i^x} r_i^x(\psi)\mu_i(\psi)d\psi}_{\bar{r}_i^x})$$

and substituting in for equation (A.77), the respective elasticities are given by:

$$-\frac{\partial log(M_i^e(1 - G(\psi_i^x))}{\partial log\tau} = \alpha_i \tag{A.81}$$

$$-\frac{\partial log(\bar{r}_i^x)}{\partial log\tau} = 0 \tag{A.82}$$

where we have relied on the assumptions that $\frac{\partial \psi_i^{d*}}{\partial \tau_{ix}} = \frac{\partial w_i}{\partial \tau_{ix}} = 0$. The first equation above shows that the change in the log number of exporters is positive ($\alpha_i > 0$) and equal to α_i. Since larger cities have lower shape parameters (note: the negative correlation between city size and the city-specific shape parameters is established by the combination of Lemma 1 and Proposition $1(i)$), they will also have a smaller increase in the percent of exporting firms following a reduction in trade costs. Note that there is no change in the average exports per exporter, as the increase in the exports of existing exporters is exactly offset by the fact that new exporters export small amounts. The second equation shows that the intensive trade margin of the trade elasticity is zero. This proves the first two statements of Proposition 5. It also directly links to our regression equation (23). Since $\bar{r}_i^x$ does not depend on τ_{ix}, the implied prediction for the estimated coefficients is that $\beta_1 = \beta_2 = 0$ for the intensive margin regression. For the extensive margin regression the model predicts $\beta_1 > 0$ as $\alpha_i > 0$ and $\beta_2 < 0$ as α_i is decreasing with city size.

Moving on to proving the third statement of Proposition 5, note that we can write aggregate exports at the country level as

$$Exp = \sum_i Exp_i \tag{A.83}$$

Taking logs and differentiating with respect to τ yields

$$\frac{1}{Exp}\frac{\partial Exp}{\partial \tau} = \sum_i \frac{1}{Exp}\frac{\partial Exp_i}{\partial \tau} = \sum_i \frac{Exp_i}{Exp}\frac{\partial Exp_i}{\partial \tau}\frac{1}{Exp_i} \tag{A.84}$$

Which yields:

$$\epsilon_c = \sum_i \frac{Exp_i}{Exp}\alpha_i \tag{A.85}$$

which completes the proof of this statement.

Elasticities with fixed trade costs. Finally, while our proof of the heterogenous trade elasticities across cities relies on variable trade costs (τ_x), we obtain qualitatively similar heterogeneous effects when using fixed costs of exporting (F_x). These effects are also driven by the extensive margin:

$$-\frac{\partial log(Exp_i)}{\partial log F_x} = \frac{\alpha_i}{\sigma - 1} - 1 \tag{A.86}$$

$$-\frac{\partial log(M_i^e(1 - G(\psi_i^x))}{\partial log F_x} = \frac{\alpha_i}{\sigma - 1} - 1 \tag{A.87}$$

$$-\frac{\partial log(\bar{r}_i^x)}{\partial log F_x} = 0 \tag{A.88}$$

where (A.86) shows that the trade elasticity is decreasing with city size, as the Pareto shape parameter α_i is smaller for larger cities. (A.87) shows that this heterogeneity is solely driven by the extensive margin, while the intensive margin in (A.88) is zero. Comparing these equations to (A.80), (A.81), and (A.82) reveals that we obtain the same qualitative predictions for variable and fixed trade costs.

D.7 Proposition A.1

Following the steps in the proof of proposition 1 it is straightforward to show that:

- The fraction of firms in a city of size L in equilibrium that is engaged in exporting is given by

$$\chi_L^* = \frac{1 - G_L(\psi_L^{x*})}{1 - G_L(\psi_L^d *)} = \left(\frac{\psi_L^{d*}}{\psi_L^{x*}}\right)^{d(L)} = \left[\frac{1}{\tau}\left(\frac{F}{F_x}\right)^{\frac{1}{\sigma-1}}\right]^{d(L)} \tag{A.89}$$

- The domestic entry productivity thresholds for a city of size L will be given by

$$\psi_L^{d*} = \left\{\frac{1}{F_e}\frac{\sigma - 1}{d(L) - \sigma + 1}\left\{F_d + \left[\frac{1}{\tau}\left(\frac{F_d}{F_x}\right)^{\frac{1}{\sigma-1}}\right]^{d(L)} F_x\right\}\right\}^{d(L)^{-1}} A(L) \tag{A.90}$$

- Average productivities for domestic and exporting firms in a city of size L is given by

$$\tilde{\psi}_L^d = \left[\frac{1}{1-G_L(\psi_L^{d*})}\int_{\psi_L^{d*}}\psi^{\sigma-1}g_L(\psi)d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{d(L)}{d(L)-\sigma+1}\right)^{\frac{1}{\sigma-1}}\psi_L^{d*} \quad \forall L \quad \text{(A.91)}$$

$$\tilde{\psi}_L^x = \left[\frac{1}{1-G_L(\psi_L^{x*})}\int_{\psi_L^{x*}}\psi^{\sigma-1}g_L(\psi)d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{d(L)}{d(L)-\sigma+1}\right)^{\frac{1}{\sigma-1}}\psi_L^{x*} \quad \forall L \quad \text{(A.92)}$$

- City-level export intensity for a city of size L is given by

$$Expint_L = \frac{\chi_L^* F_x}{F_d + \chi_L^* F_x} \quad \text{(A.93)}$$

Taken together, equations (A.89) and (A.93) show that larger cities will have higher export intensities, and that this will be fully driven by the extensive margin of exporting.

Moreover, equation (A.90) shows that the domestic market entry thresholds are higher for larger cities. Given that equation (22) still holds this further implies that larger cities will feature higher entry into exporting productivity thresholds. Coupling these findings with equations (A.91) and (A.92) leads to the conclusion that larger cities will display higher average productivity for both domestic and exporting firms. Finally noting that the average productivity of exporters is higher in all cities, and that larger cities will have a higher fraction of exporting firms, this leads to the conclusion that larger cities have higher overall average productivity.

D.8 Proposition A.2

The proof proceeds in multiple steps. First, we will show that our learning externality will endogenously lead to a higher productivity growth rate in larger cities. We will then show that a Pareto distribution is a candidate solution for the stable firm productivity distribution of the stochastic growth process from Section C.2. Lastly, we will combine these steps to show that stable firm productivity distributions in larger cities feature lower tail parameters.

The optimal number of experts a firm decides to meet in each period is given by the solution to equation (A.46):

$$\frac{da_{L_i}(N_{i,t})}{dN_{i,t}} = g'(N_{i,t}) - \frac{c}{L_{i,t}} = 0 \quad \Rightarrow N_{i,t}^* = (g')^{-1}\left(\frac{c}{L_{i,t}}\right)$$

which yields the following productivity growth term as a function of the number of meetings:

$$a_{L_i}(N_{i,t}^*) = g\left((g')^{-1}\left(\frac{c}{L_{i,t}}\right)\right) - (g')^{-1}\left(\frac{c}{L_{i,t}}\right)\frac{c}{L_{i,t}}$$

which we write as $a_{L_i}(N_{i,t}^*(L_{i,t})) = a_{L_i}(L_{i,t})$ for notational convenience. To show that the drift increases in city size, we first show that the optimal number of meetings is an increasing function of city size:

$$\frac{dN_{i,t}^*}{dL_{i,t}} = \frac{-\frac{c}{L_{i,t}^2}}{g''((g')^{-1}\left(\frac{c}{L_{i,t}}\right))} > 0 \tag{A.94}$$

It follows that the drift term in (A.45) increases with city size:

$$\begin{aligned}\left.\frac{da_{L_i}(L_{i,t})}{dL_{i,t}}\right|_{N_{i,t}=N_{i,t}^*} &= g'(N_{i,t}^*)\frac{dN_{i,t}^*}{dL_{i,t}} - \frac{dN_{i,t}^*}{dL_{i,t}}\frac{c}{L_{i,t}} + N_{i,t}^*\frac{c}{L_{i,t}^2} = \frac{dN_{i,t}^*}{dL_{i,t}}(g'(N_{i,t}^*) - \frac{c}{L_{i,t}}) + N_{i,t}^*\frac{c}{L_{i,t}^2} \\ &= N_{i,t}^*\frac{c}{L_{i,t}^2} > 0\end{aligned}$$

A result we will use in equation (A.99) below.

Next, following the approach reviewed by Gabaix (2009), we show that a Pareto distribution is a viable candidate solution for the stable distribution of the firm growth process from Section C.2, employing a guess and verify procedure. The Fokker-Planck equation, which describes the motion of the underlying firm productivity distribution ($\mathcal{L}_{i,t}$), that corresponds to this process is given by:

$$\partial_t\mathcal{L}_{i,t}(\varphi) = -\partial_\varphi(a_i(L_{i,t})\varphi\mathcal{L}_{i,t}(\varphi)) + \frac{1}{2}\partial_{\varphi\varphi}(\nu^2\varphi^2\mathcal{L}_{i,t}(\varphi)) - \delta\mathcal{L}_{i,t}(\varphi) + \delta\mathbf{1}_{\{\varphi=\varphi_{Min}\}} \tag{A.95}$$

Substituting in a Pareto candidate distribution ($\mathcal{L}_i(\varphi) = \varphi_{Min}^{\alpha_i}\alpha_i\varphi^{-\alpha_i-1}$), yields for $\varphi > \varphi_{Min}$:

$$\begin{aligned}\partial_t\mathcal{L}_{i,t}(\varphi) = 0 = -\partial_\varphi(a_i(L_i)\varphi_{Min}^{\alpha_i}\alpha_i\varphi^{-\alpha_i}) + \frac{1}{2}\partial_{\varphi\varphi}(\nu^2\varphi_{Min}^{\alpha_i}\alpha_i\varphi^{-\alpha_i+1}) - \delta\varphi_{Min}^{\alpha_i}\alpha_i\varphi^{-\alpha_i-1} \\ \Rightarrow 0 = (a_i(L_i)\alpha_i + \frac{\nu^2}{2}\alpha_i(\alpha_i-1) - \delta)\varphi^{-\alpha_i-1}\end{aligned}$$

Since $\varphi^{-\alpha_i-1} > 0$, it follows that a Pareto distribution is a stationary solution, if the following equation holds:

$$0 = a_i(L_i)\alpha_i + \frac{\nu^2}{2}\alpha_i(\alpha_i-1) - \delta \tag{A.96}$$

Note that as L_i is a constant in steady state and so is $a_i(L_i)$. Now, consider the function $h(\alpha_i) :=$

$a_i(L_i)\alpha_i + \frac{\nu^2}{2}\alpha_i(\alpha_i - 1) - \delta$ which is continuous. Then we have that:

$$\lim_{\alpha_i \to 1_+} h(\alpha_i) = a_i(L_i) - \delta < 0 \tag{A.97}$$

$$\lim_{\alpha_i \to +\infty} h(\alpha_i) = +\infty \tag{A.98}$$

since we assumed $a_i(L_i) \in [0, \delta)$. By the intermediate value theorem, there exists $\alpha_i > 1$ such that $h(\alpha_i) = 0$. Therefore, given exogenous ν and δ, we obtain that for each $a_i(L_i) \in [0, \delta)$, there exists $\alpha_i > 1$ such that the corresponding Pareto distribution $\mathcal{L}_i(\varphi) = \varphi_{Min}\alpha_i\varphi^{-\alpha_i - 1}$ is a stationary solution of the process.

Lastly, using the implicit function theorem and the result from equation (A.95), it is straightforward to show that:

$$\frac{d\alpha_i}{dL_i} = -\frac{\alpha_i a_i'(L_i)}{a_i(L_i) + \nu^2(\alpha_i - \frac{1}{2})} < 0 \tag{A.99}$$

which concludes our proof by establishing the negative relationship between the shape parameter of the steady-state firm productivity distribution and city size.

E Quantitative Model

E.1 Equilibrium

This section lists the system of equations that represents the equilibrium of the quantitative model.

1. Expenditure and revenues are equalized for all locations

$$w_i L_i = \sum_{n \in N \cup \{c\}} \pi_{ni} w_n L_n \quad \forall i \in N \cup \{c\} \tag{A.100}$$

Where π_{ni} denotes the expenditure share of goods sourced from i in the overall expenditure of location n. These expenditure shares comprise a combination of domestic and foreign consumption:

$$\pi_{ni} = \frac{p_i^d M_i^e (w_i d_{ni})^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k d_{nk})^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.101}$$

$$\pi_{ci} = \frac{p_i^x M_i^e (w_i d_{ci})^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k d_{ck})^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.102}$$

$$\pi_{cc} = \frac{p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k d_{ck})^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.103}$$

$$\pi_{nc} = \frac{p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k d_{nk})^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.104}$$

where the expenditure share π_{ni} reflects domestic consumption (goods from city i consumed in location n), the share π_{ci} refers to foreign consumption of goods from city i, π_{cc} reflects foreign consumption of foreign goods, and π_{nc} is the share of foreign goods in the expenditure of domestic location n. As before, M_i^e denotes the measure of entrants at each location i. The series of variables of the format $\tilde{\psi}_a^b$ represent the measures of average productivity of firms in location a serving market b, with $a = \{i, c\}$ and $b = \{d, x\}$. Using this notation, $\tilde{\psi}_a^b$ is given by:

$$\tilde{\psi}_a^b = \left[\frac{1}{1 - G_a(\psi_a^{b*})} \int_{\psi_a^{b*}} \psi^{\sigma-1} g_a(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_a}{\alpha_a - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_a^{b*} \tag{A.105}$$

Moreover, the series of terms p_a^b that appear in the expressions for the expenditure shares denote the probability of a firm at location $a = \{i, c\}$ choosing to serve market $b = \{d, x\}$. The terms p_a^b are given by:

$$p_a^b = \left(\frac{A_a}{\psi_a^{b*}} \right)^{\alpha_a} \tag{A.106}$$

2. Zero profit conditions for marginal entrants into the domestic market at each location i (see equation (12))

3. Zero profit conditions for marginal entrant into exporting at each location i (see equation (12))

4. Free entry conditions at all locations

$$\bar{\pi}_i = p_i^d \pi_i^d (\tilde{\psi}_i^d) + p_i^x \pi_i^x (\tilde{\psi}_i^x) = w_i F_e \tag{A.107}$$

5. Price indices at each location are given by

$$P_i = \left(\frac{p_i^d M_i^e}{\pi_{ii}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_i}{\tilde{\psi}_i^d}\right) \tag{A.108}$$

6. Labor markets clear at each location i

$$M_i^e \cdot \bar{l}_i = L_i \iff M_i^e = \frac{L_i}{\sigma\left[F_e + p_i^d F_d + p_i^x F_x\right]} \tag{A.109}$$

where $\bar{l}_i$ denotes average labor employed by a firm entering at location i

7. The share of the population at each location is given by:

$$\frac{L_i}{\bar{L}} = \frac{B_i(w_i/P_i^{\beta} r_i^{\gamma(1-\beta)})^{\epsilon}}{\sum_{k\in N} B_k(w_k/P_k^{\beta} r_k^{\gamma(1-\beta)})^{\epsilon}} \tag{A.110}$$

8. Land markets clear in each city

$$r_i = \frac{1-\beta}{\beta}\gamma\frac{w_i L_i}{N_i} \tag{A.111}$$

E.2 Welfare Derivations

To derive the formulas for welfare changes presented in the main text, we first derive a formula for aggregate welfare in the setting of our quantitative model. We combine the formula for welfare in equation (27) with the population share formula in (A.110) to obtain:

$$\bar{U} = \delta\left[B_n(v_n/P_n^{\beta}(p_n^h)^{1-\beta})^{\epsilon}\frac{\bar{L}}{L_n}\right]^{\frac{1}{\epsilon}} \tag{A.112}$$

Plugging equations (A.108), (17) and (18) into (A.112), and noting that $v_i = \frac{w_i}{\beta}$, yields after rearranging an equilibrium welfare formula of the form:

$$\bar{U} = \frac{\delta B_n^{\frac{1}{\epsilon}}(\tilde{\psi}_n^d)^{\beta} w_n^{(1-\beta)(1-\gamma)} N_n^{\gamma(1-\beta)}\left(\frac{1}{\pi_{nn}}\right)^{\frac{\beta}{\sigma-1}} L_n^{-(\frac{1}{\epsilon}+\gamma(1-\beta)-\frac{\beta}{\sigma-1})}\left[\frac{p_n^d}{\sigma[F_e+p_n^d F_d+p_n^x F_x]}\right]^{\frac{\beta}{\sigma-1}}}{\beta\left(\frac{\sigma}{\sigma-1}\right)^{\beta}\frac{1}{(1-\gamma)^{(1-\gamma)(1-\beta)}}\left(\frac{1-\beta}{\beta}\right)^{\gamma(1-\beta)}\bar{L}^{-\frac{1}{\epsilon}}} \tag{A.113}$$

Using the expression (A.113) we can then write the welfare change between two trading equilibria T and T' (note that given that we study an economic geography setting where locations are interpreted as cities within the same country, the traditional "autarky" benchmark is not relevant):

$$\frac{\bar{U}^T}{\bar{U}^{T'}} = \left(\frac{\pi_{nn}^{T'}}{\pi_{nn}^T}\right)^{\frac{\beta}{\sigma-1}}\left(\frac{L_n^{T'}}{L_n^T}\right)^{\frac{1}{\epsilon}+\gamma(1-\beta)-\frac{\beta}{\sigma-1}}\left(\frac{w_n^T}{w_n^{T'}}\right)^{(1-\beta)(1-\gamma)}\left(\frac{\psi_n^{d*T'}}{\psi_n^{d*T}}\right)^{\beta(\frac{\alpha_n}{\sigma-1}-1)} \tag{A.114}$$

The presence of wages in the formula above for the change in welfare associated with changing trade costs stems from our assumptions that capital income is redistributed locally.[23] To ease comparison with other models in the literature that do not feature an explicit housebuilding sector, we switch off the housebuilding channel by setting $\gamma = 1$ which leads us to the simplified formula for the welfare changes associated to changing trade costs presented in the main text:

$$\frac{\bar{U}^T}{\bar{U}^{T'}} = \left(\frac{\pi_{nn}^{T'}}{\pi_{nn}^T}\right)^{\frac{\beta}{\sigma-1}} \left(\frac{L_n^{T'}}{L_n^T}\right)^{\frac{1}{\epsilon}+(1-\beta)-\frac{\beta}{\sigma-1}} \left(\frac{\psi_n^{d*T'}}{\psi_n^{d*T}}\right)^{\beta(\frac{\alpha_n}{\sigma-1}-1)} \tag{A.115}$$

The expression in (A.115) is similar to that found in the increasing returns specification of Redding (2016). As in Redding (2016), computing the common (across all locations within the country) changes in welfare from changing trade costs involves taking into account not only the own-city trade shares of each location (which affect consumption price indices in each city) but also population redistributions across locations (which affect the price of the immobile factor land in each location). The agglomeration effects due to the endogenous measure of varieties in each location are captured by the exponent on relative populations ($-\beta/(\sigma-1)$). Beyond these terms already present in Redding (2016), our welfare formula includes an additional term that captures the endogenous selection margin. As a city's market access improves, the gains from trade are tempered not only by the congestion force of scarce real estate, but also by the more stringent entry selection, as higher wages associated with improved market access cause a higher fraction of firms that enter at a given location to choose not to serve the domestic (and hence the own-city) market.[24]

Finally, we note that the expression for welfare changes in response to trade cost changes (equations (A.114) and (A.115)) can be written as a more explicit formula involving only own-city trade shares, population and wages. To perform this rewriting, we first write down the domestic (i.e., within China) sales of the marginal producer at location i.

$$r_i^d(\psi_i^{d*}) = \sigma w_i F_d \tag{A.116}$$

We can then write the sales of location i to a particular domestic destination n

$$r_i^n(\psi_i^{d*}) = \frac{R_n P_n^{\sigma-1} d_{ni}^{1-\sigma}}{\sum_k R_k P_k^{\sigma-1} d_{ki}^{1-\sigma}} \sigma w_i F_d \tag{A.117}$$

Thus for an active firm of arbitrary productivity $\psi \geq \psi_i^{d*}$ at location i, sales to location n will be

[23] Intuitively, in our setting changes in trade costs will affect housing demand in each location, thus changing the allocation of where capital income is realised. In turn, as capital is taxed and redistributed locally this will affect the size of subsidies received by the inhabitants of each location

[24] In our setting, this stronger selection effect associated with improved market access is reflected in higher domestic entry thresholds

given by:

$$r_i^n(\psi) = \left(\frac{\psi}{\psi_i^{d*}}\right)^{\sigma-1} \frac{R_n P_n^{\sigma-1} d_{ni}^{1-\sigma}}{\sum_k R_k P_k^{\sigma-1} d_{ki}^{1-\sigma}} \sigma w_i F_d \tag{A.118}$$

We can then write the total sales of location i to location n (which we denote X_{ni}) as

$$X_{ni} = \int_{\psi_i^{d*}}^{\infty} r_i^n(\psi) M_i^e g_i(\psi) d\psi \tag{A.119}$$

Which gives us after some manipulations

$$X_{ni} = \left(\frac{A_i}{\psi_i^{d*}}\right)^{\alpha_i} \frac{R_n P_n^{\sigma-1} d_{ni}^{1-\sigma}}{\sum_k R_k P_k^{\sigma-1} d_{ki}^{1-\sigma}} \sigma w_i F_d M_i^e \frac{\alpha_i}{\alpha_i - \sigma + 1} \tag{A.120}$$

Noting that we can write

$$M_i^e = \frac{\sigma - 1}{\sigma \alpha_i} \frac{L_i}{F_e} \tag{A.121}$$

and that $X_i = w_i L_i$ we can write for any location i (by substituting (A.121) into (A.120)):

$$\pi_{ii} = \left(\frac{A_i}{\psi_i^{d*}}\right)^{\alpha_i} \frac{F_d}{F_e} \frac{R_i P_i^{\sigma-1}}{\sum_k R_k P_k^{\sigma-1} d_{ki}^{1-\sigma}} \frac{\sigma - 1}{\alpha_i - \sigma + 1} \tag{A.122}$$

which can in turn be rewritten as

$$(\psi_i^{d*})^{\alpha_i} = A_i^{\alpha_i} \frac{F_d}{F_e} \underbrace{\frac{R_i P_i^{\sigma-1}}{\sum_k R_k P_k^{\sigma-1} d_{ki}^{1-\sigma}}}_{\pi_{iid}} \frac{\sigma - 1}{\alpha_i - \sigma + 1} \frac{1}{\pi_{ii}} = A_i^{\alpha_i} \frac{F_d}{F_e} \frac{\sigma - 1}{\alpha_i - \sigma + 1} \frac{\pi_{iid}}{\pi_{ii}} \tag{A.123}$$

where the key step is noting that the expression captured by the underbrace constitutes the "domestic" own trade share (π_{iid}) of location i (i.e. the share of own city revenues relative to all revenues made from the domestic market). Equation (A.123) thus gives us an expression for the domestic productivity thresholds at each location as a function of the domestic and overall own expenditure shares. Substituting equation (A.123) into equation (A.114) yields an expression of the welfare gains from trade:

$$\begin{aligned}
\frac{\bar{U}^T}{\bar{U}^{T'}} &= \left(\frac{\pi_{ii}^{T'}}{\pi_{ii}^T}\right)^{\frac{\beta}{\alpha_i}} \left(\frac{\pi_{iid}^{T'}}{\pi_{iid}^T}\right)^{\beta\left(\frac{1}{\sigma-1} - \frac{1}{\alpha_i}\right)} \left(\frac{L_i^{T'}}{L_i^T}\right)^{\frac{1}{\epsilon} + \gamma(1-\beta) - \frac{\beta}{\sigma-1}} \left(\frac{w_n^T}{w_n^{T'}}\right)^{(1-\beta)(1-\gamma)} \\
&= \left(\frac{\pi_{ii}^{T'}/\pi_{iid}^{T'}}{\pi_{ii}^T/\pi_{iid}^T}\right)^{\frac{\beta}{\alpha_i}} \left(\frac{\pi_{iid}^{T'}}{\pi_{iid}^T}\right)^{\frac{\beta}{\sigma-1}} \left(\frac{L_i^{T'}}{L_i^T}\right)^{\frac{1}{\epsilon} + \gamma(1-\beta) - \frac{\beta}{\sigma-1}} \left(\frac{w_n^T}{w_n^{T'}}\right)^{(1-\beta)(1-\gamma)}
\end{aligned} \tag{A.124}$$

This expression reveals a variation of the standard formulation of welfare changes associated with changes in market access in the tradition of Arkolakis et al. (2012, ACR for short), who show that in a large class of trade models, the own-trade share and the trade elasticity are sufficient to calculate the gains from trade. Our 'trade plus economic geography' setting with many locations features each city i's own-trade share *relative* to domestic expenditure (π_{iid}) and *relative* to to overall expenditure (π_{ii}), as well as population reallocation (L_i) and wages (w_i).[25] As before, to compare our formula for the welfare gains from trade to other benchmark models, we set $\gamma = 1$ to abstract from the house-building channel. This yields:

$$\frac{\bar{U}^T}{\bar{U}^{T'}} = \left(\frac{\pi_{ii}^{T'}/\pi_{iid}^{T'}}{\pi_{ii}^{T}/\pi_{iid}^{T}}\right)^{\frac{\beta}{\alpha_i}} \left(\frac{\pi_{iid}^{T'}}{\pi_{iid}^{T}}\right)^{\frac{\beta}{\sigma-1}} \left(\frac{L_i^{T'}}{L_i^{T}}\right)^{\frac{1}{\epsilon}+(1-\beta)-\frac{\beta}{\sigma-1}} \tag{A.125}$$

This welfare formula nests similar formulas in Redding (2016), and Melitz (2003) with Pareto. Our welfare formula collapses to the standard ACR formula for the Melitz/Chaney case if the population of the entire country was located in one domestic location and there was no mobility. It collapses to the formula in Redding in the absence of firm heterogeneity.

E.3 Calibration Procedure

This section provides a detailed description of the calibration procedure of the quantitative model outlined in section 5.2.

Our calibration proceeds in two steps. In the first stage, we calibrate the parameters and exogenous variables that can be directly linked to the existing empirical literature or the data using the structure of the model ($\sigma, \beta, \gamma, \epsilon, \alpha_n, F_d, F_x, F_e, d_{ni},$ and d_{nc}). Table A.17 summarizes these parameters and the corresponding data moments. We set the elasticity of substitution $\sigma = 3$ (following estimates by Eaton et al., 2011, for France)[26] and set the housing expenditure share $(1 - \beta)$ in both countries to 25%.[27] The land cost share in housing production (γ) matches the median housing supply elasticity across Chinese cities estimated in Wang et al. (2012), and across U.S. metropolitan areas in Saiz (2010) for the case of France.[28] The dispersion of the idiosyncratic

[25]Our formula features two different own-trade shares – one relative to expenditure on domestic goods (π_{iid}) and one relative to overall expenditure (π_{ii}) – because the intra- and the international trade elasticity differ in our model. While the international trade elasticity is heterogeneous across locations and equal to α_i, the domestic trade elasticity is given by $\sigma - 1$. This is because in our setting there are no fixed domestic trade costs and hence no extensive margin adjustments to domestic trade – i.e., all active firms serve all domestic locations. This is an artifact of our specification of fixed market entry costs (i.e., assuming joint fixed cost for entry to all domestic markets). We make this assumption both to keep closer to the spirit of Melitz (2003) and with our quantitative exercise in mind, as city-specific fixed entry costs would be difficult to pin down empirically.

[26]As Chaney (2008) shows, a larger value for σ makes the intensive margin more responsive and the extensive margin less responsive to trade costs.

[27]The value we adopt for the housing expenditure share is in line with Davis and Ortalo-Magne (2011), who report a stable housing expenditure share over time and across MSA in the United States of around 0.25. In the case of China, this value closely matches official statistics produced by the National Bureau of Statistics of China. See http://www.stats.gov.cn/english/PressRelease/202201/t20220118_1826649.html.

[28]The housing supply in the model is given by $(1 - \gamma)/\gamma$, which is monotonically decreasing in the value of the land cost share in housing production, γ.

amenity draws, ϵ, takes a value of 3.18, following the estimates in Bryan and Morten (2019) using Indonesian data.[29] To calculate the matrix of domestic bilateral trade costs between domestic locations, d_{ni}, or between any domestic location and the nearest port or border, d_{pi}, we use the bilateral straight-line distances to the power of 0.33, as in Redding and Sturm (2008).[30] The value for the variable trade cost of taking a good from the border to the rest of the world, τ, is set in the second stage of the calibration procedure using the structure of the model. Finally, relying on the structure of the model, the fixed production and export costs (F_d,F_x) are given by the ratio of average domestic and export revenues to wages (see detail below in equation (A.126)). The fixed cost of entry (F_e) only scales the Pareto scale parameter and does not affect any endogenous variables in the model. We, therefore, choose a value that ensures selection effects are active in the model (i.e., the corresponding scale parameter is low enough for a non-empty subset of firms (those with low draws) not to find it profitable to produce.

Table A.17: Calibrated Parameters and Target Moments

Parameters	**Calibration Strategy**	**Value**	
Calibrated Parameters		CHN	FRA
β	Housing expenditure share (Davis and Ortalo-Magne, 2011)	0.25	0.25
σ	Elasticity of Substitution (Eaton et al., 2011)	3.0	3.0
ϵ	Dispersion of idiosyncratic amenity draws (Bryan and Morten, 2019)	3.18	3.18
γ	Median housing supply elasticity (Wang et al., 2012; Saiz, 2010)	5.4	1.75
F_d	Average domestic sales/ Average wages	19.7	0.97
F_x	Average export sales/ Average wages	30.7	1.81
d_{ni}	$\max\{1, (dist_{ni}/100)^{\phi}\}$; $\phi = 0.33$ (Redding and Sturm, 2008)	–	–
d_{pi}	$\max\{1, (dist_{pi}/100)^{\phi}\}$; $\phi = 0.33$ (Redding and Sturm, 2008)	–	–
τ	Country-level sales-weighted export intensity	2.35	2.31

Notes: The table summarizes the target moments used for determining the parameters calibrated externally to the model.

A central object for our quantitative analysis is the value for the Pareto shape parameter in each location, α_i. To estimate these parameters, we follow Head et al. (2014) estimating QQ regressions for each location using firm-level domestic revenue data, accounting for industry fixed effects. We then multiply the resulting coefficient by $(\sigma - 1)$ following di Giovanni et al. (2014) to recover the underlying productivity distribution Pareto shape parameter. In a nutshell, the QQ method

[29] We are not aware of any similar estimates produced for China or France.

[30] In particular, we compute d_{ni} as $\max\{1, (dist_{ni}/100)^{\phi}\}$, with $\phi = 0.33$, and set $d_{ni} = 1$ for location pairs separated by fewer than 100 kilometers, as for these short-distance cases, the marginal effect of distance on trade is arguably negligible.

estimates the domestic sales Pareto distribution parameters, minimizing the distance between the empirical and theoretical quantiles. In our baseline estimation of α_i, we use the 10 percent largest firms in each city and bootstrap the estimates to minimize small-sample bias.[31] For the rest of the world, we pool firms from all cities in China or France to estimate the corresponding RoW Pareto shape parameter.

Table A.18 shows the distribution of the estimated Pareto shape parameters for China and France. We find significant dispersion of the estimates across cities in both countries, with a mean of 3.17 for China (2.97 for France) and a standard deviation of 0.67 (0.36 for France). Moreover, as Figure 2 shows, the estimated Pareto shape parameters show a strong negative correlation with city size in both countries, with estimated coefficients at -0.18 for China and -0.21 for France (both coefficients statistically significant at the 1 percent level). This negative correlation lends support to our model: Large cities feature productivity distributions with thicker upper tails (i.e., smaller shape parameters), which increases the proportion of high-productivity firms that select into exporting.

Table A.18: Estimated City-Level Pareto Shape Parameters: Descriptive Statistics

				Percentiles							
	Obs.	Mean	St.Dev.	1st	5th	10th	25th	50th	75th	90th	95th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
China	615	3.17	0.67	2.18	2.36	2.47	2.70	3.05	3.44	4.09	4.56
France	304	2.97	0.36	2.22	2.47	2.58	2.75	2.91	3.15	3.42	3.56

Notes: The table shows the distribution of the estimated Pareto shape parameters for cities in China and France.

To calibrate fixed production and export costs, F_d and F_x, we rely on the convenient properties of the Pareto distribution, which implies that average revenues in each market can be written as a scaled version of revenues for the threshold productivity firm, that is, $\overline{r_{d,i}} = \alpha_i/(\alpha_i - \sigma + 1)r^*_{d,i}$ and $\overline{r_{x,i}} = \alpha_i/(\alpha_i - \sigma + 1)r^*_{x,i}$. Combining these expressions with the zero profit cut-off conditions (12) yields the following expressions for the fixed production and export cost for firms located at location i:

$$F_{d,i} = \frac{(\alpha_i - \sigma + 1)}{\alpha_i \sigma} \frac{\overline{r_{d,i}}}{w_i}, \qquad F_{x,i} = \frac{(\alpha_i - \sigma + 1)}{\alpha_i \sigma} \frac{\overline{r_{x,i}}}{w_i} \tag{A.126}$$

Equation (A.126) states that the value of the fixed production (export) cost $F_{d,i}$ ($F_{x,i}$) can be computed using values for the elasticity of substitution, the Pareto shape parameter, plus data on av-

[31] In approximately 22% of the cities in China and 4% of French cities the estimated shape parameters are below $(\sigma - 1)$. For our baseline quantitative exercise, we re-scale the estimated shape parameters such that the city with the minimum value has a value for the shape parameter just above $(\sigma - 1)$. Censoring the sample at $(\sigma - 1)$ delivers similar results but a lower correlation with the actual shape parameters (that is, before censoring or re-scaling).

erage domestic (export) sales and wages in each location. To isolate the estimated model from differences in fixed costs across locations, we calibrate national-level values for F_d and F_x, computing (A.126) with the RoW's Pareto shape parameter, as well as average nationwide values for domestic sales, export sales, and wages.[32]

The second stage recovers the fundamental parameter (A_n, B_n) and the value for the variable trade cost between ports and the rest of the world (τ) using the values for the parameters calibrated in the first stage and data on population, wages, and country-level export intensity. The algorithm for obtaining the values for these variables consists of the following steps:

Step 1. Guess a value for τ, and solve for $\{A_i, P_i, M_i^e, \psi_i^{d*}, \psi_i^{x*}\}$ using equations (12), (A.100), (A.101), (A.108) and (A.109).

Step 2. Repeat step 1 for a grid of values for τ. Compute national-level sales-weighted export intensity according to equation (A.127). Choose the value of τ that minimizes the distance with the national-level export intensity computed from the data.

$$\text{Export Intensity} = \frac{\sum_i (\psi_i^{x*})^{\sigma-1-\alpha_i} R_c \left(\frac{P_c}{d_{ip}\tau}\right)^{\sigma-1}}{\sum_i \left[(\psi_i^{x*})^{\sigma-1-\alpha_i} R_c \left(\frac{P_c}{d_{ip}\tau}\right)^{\sigma-1} + (\psi_i^{d*})^{\sigma-1-\alpha_i} \sum_n R_n \left(\frac{P_n}{d_{ni}}\right)^{\sigma-1}\right]} \tag{A.127}$$

Step 3. Using the optimal value of τ derived in the previous step, invert the model as in step 1 to find the values of $\{A_i\}$ consistent with the actual distribution of population and wage across locations in each country.

Step 4. Using information on population and wages, compute location income, v_i, and land rents, r_i, in each location using equations (16) and (18).

Step 5. Using the equilibrium values for each location's aggregate price levels, P_i and location income, v_i, and land rents, r_i solve the system of equations (A.110) to recover the fundamental parameters $\{B_i\}$ for each location.

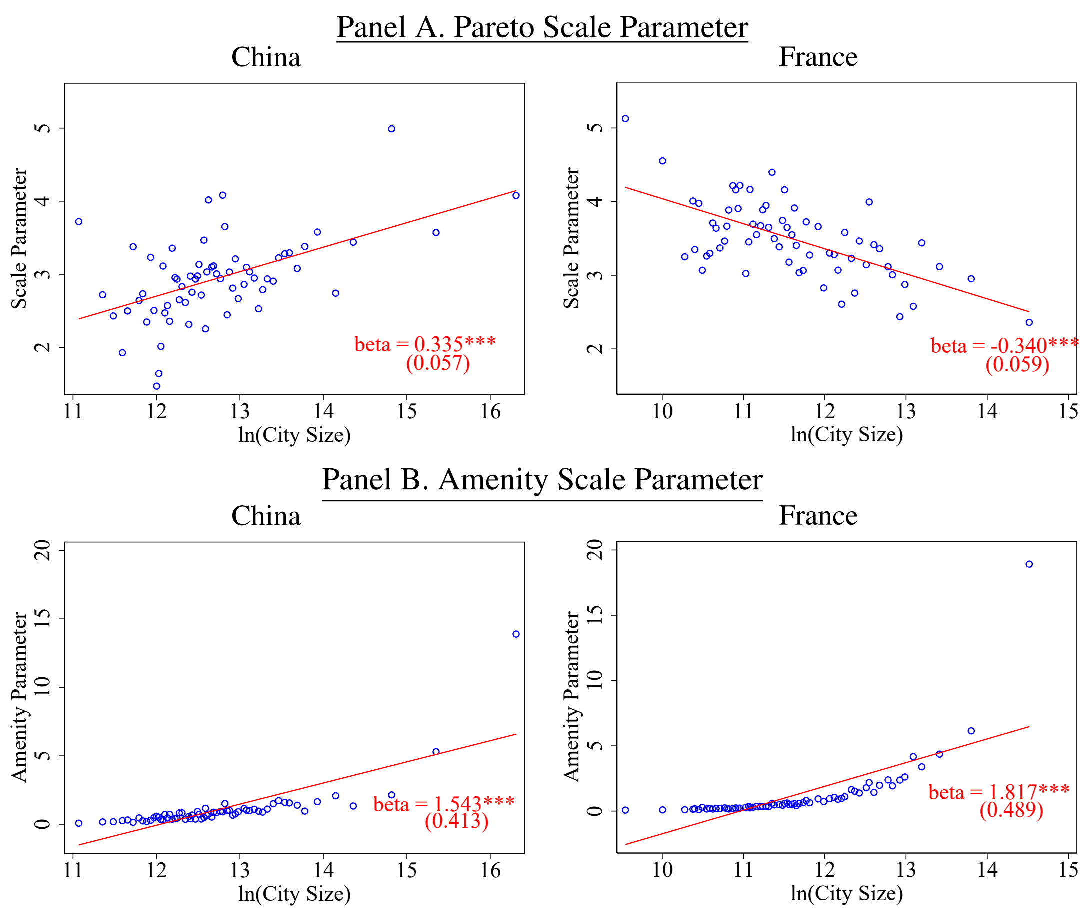
E.4 Estimated Parameters and City Size

Productivity and amenity scale parameters. Panel A in Figure A.2 shows, for each country, the estimated values of the Pareto productivity *scale* parameter, plotted in a binscatter against city size

[32] Alternatively, we can allow for city-specific fixed costs of exporting, using the ratio of export revenues to wages for each city along city-specific shape parameters. This yields very similar results (available upon request).

(x-axis).[33] The Pareto scale parameter features a positive relationship with city size in China but a negative relationship with city size in France. This suggests that, in larger cities, the productivity distribution is shifted to the right in China and to the left in France. Panel B in Figure A.2 plots the Frechet amenity scale parameter. In both countries, this parameter shows a positive relationship with city size. Moreover, this relationship seems to be driven to a certain extent by the estimated parameter in the largest cities, with values that easily surpass the average across cities several times.

Figure A.2: Inverted Model Fundamentals by City Size



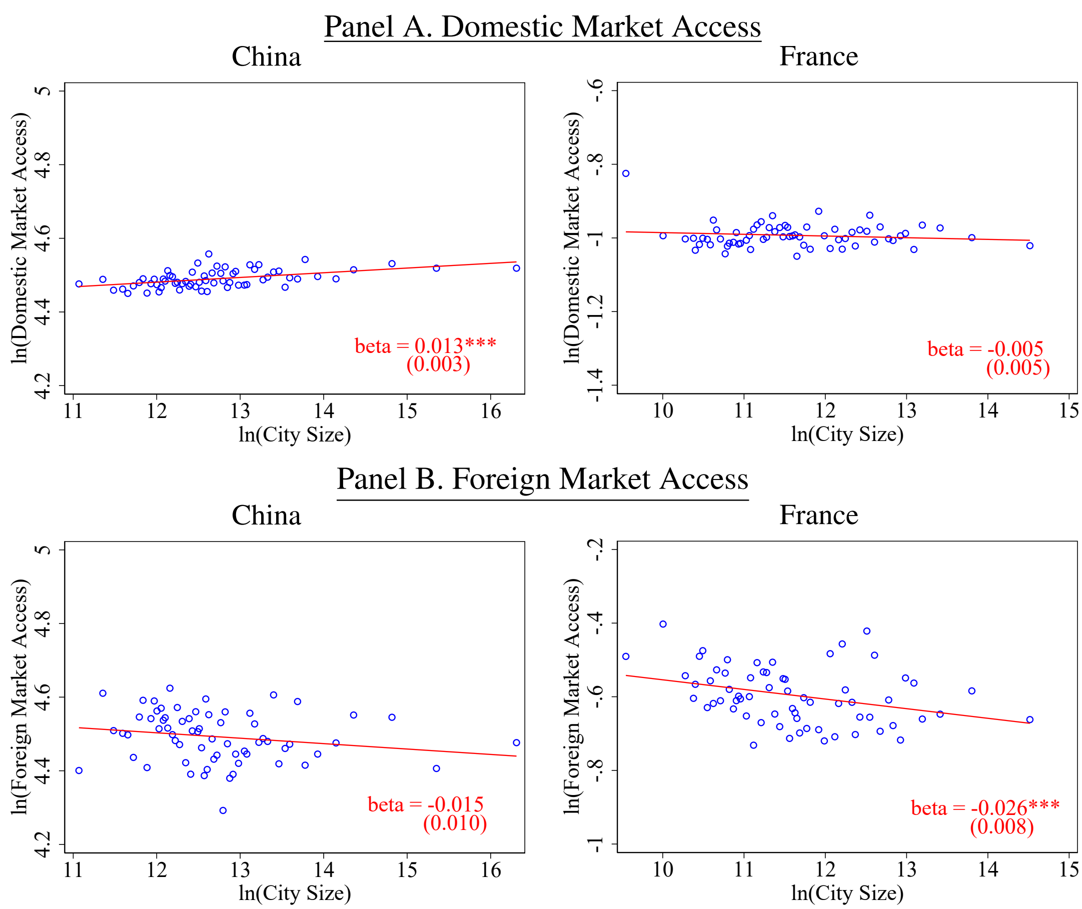
Notes: The figure plots the correlation of the two fundamentals of our quantitative model with city size for China and France: the Pareto scale parameter (A_i) and the city-specific scale parameter of the amenity distribution (B_i), and city size. Note that the distribution of the Pareto *shape* parameter (α_i) is shown in Figure 2 in the paper. Each dot (bin) represents 10 underlying cities for China, and 5 cities for France. See Section 5.2 for a description of the calibration procedure and Appendix E.3 for additional details.

Domestic and foreign market access. Figure A.3 plots the estimated domestic (Panel A) and for-

[33] Note that we plot the same relationship for the Pareto *shape* parameter in Figure 2 in the paper.

eign market access (Panel B) in the baseline model, plotted in a binscatter against city size (x-axis). For each location i, domestic market access is defined as $\sum_n d_{ni}^{1-\sigma} w_n L_n P_n^{\sigma-1}$; higher values of this variable are interpreted as better domestic market access – a lower average trade cost of accessing other domestic markets. Correspondingly, foreign market access for location i is defined as $d_{ci}^{1-\sigma} w_c L_c P_c^{\sigma-1}$, with higher values denoting, again, a lower cost of shipping goods, but to the rest of the world. The figure shows that larger cities in China tend to have better foreign market access but a similar domestic market access as smaller cities. Larger cities in France, in contrast, tend to have better domestic *and* foreign market access.

Figure A.3: Market Access and City Size for China and France



Notes: This figure displays the correlation between city size and domestic and foreign market access for China and France at the model equilibrium implied by the data for our baseline model. Domestic market access$_i$ = $\sum_n d_{ni}^{1-\sigma} w_n L_n P_n^{\sigma-1}$; Foreign market access$_i$ = $d_{xi}^{1-\sigma} w_F L_F P_F^{\sigma-1}$. The red lines are the bivariate regression lines with slope coefficient beta. Each dot (bin) represents 10 underlying cities for China, and 5 cities for France.

E.5 Model Validation: Local Trade Elasticity and City Size

In this appendix section, we compare the predicted relationship between the local international trade elasticity and city size across different versions of the model presented in Section 5.5. For each model, we first compute the international trade elasticities by location at the data-implied equilibrium. We then regress this value on the natural logarithm of city size.[34] Table A.19 reports the estimated coefficients for China (Panel A) and France (Panel B). The first row shows results for the baseline model (also shown in Figure 4 in the paper), with city-specific Pareto shape and scale parameters. Recall from our discussion in Section that a positive regression coefficient β indicates that larger cities have a less negative (i.e., smaller) trade elasticity (see footnote 47 in the paper). The second row shows the estimated coefficients for a model where firms are homogeneous within cities, but there can be variation in productivity levels across cities. In contrast to the baseline, this model produces a small *negative* relationship between the local trade elasticity and city size. This negative relationship is driven by domestic trade costs – the remaining force when we shut down firm heterogeneity within locations.[35] The third row in Table A.19 corresponds to the baseline model but shuts down within-country trade costs. This model produces a slightly larger relationship between the trade elasticity and city size than our baseline, in line with our argument above that variation in domestic trade costs in isolation delivers a *negative* relationship between city size and local trade elasticities and can therefore not account for the pattern in the data. Finally, the last row in Table A.19 shows results for a model featuring heterogeneity in market access across cities, within-city firm-level heterogeneity, but no differences in the thickness of the upper tail. In both countries, this model again delivers a slightly *negative* relationship between the trade elasticity and city size, driven by the same domestic-trade-cost mechanics as model (2). Thus, our core assumption – the thicker productivity upper tail in larger cities – is a crucial distinguishing feature of our model that replicates the observed pattern of local trade elasticities in the data.

[34]See Appendix F for a detailed description of the equilibrium conditions in each model.

[35]Domestic trade costs introduce differential changes in local demand across locations when an economy opens up to trade. This, in turn, generates general equilibrium adjustments that deviate from the assumptions under which the trade elasticity is equal to $(\sigma - 1)$ in the homogenous firm model, or the Pareto shape parameter in a heterogenous firm model (Chaney, 2008). In particular, as international trade costs fall, two countervailing forces coexist that are driven by the higher initial foreign market access of larger cities (see Figure A.3 for this pattern in the context of our baseline model). On the one hand, higher market access leads to higher import intensity in larger cities, which in turn means that the local price index decreases more. This increases population size c.p. and thereby exports. On the other hand, better market access also means a higher initial export intensity of larger cities (even if we showed above that this effect is quantitatively small). Higher initial export intensity implies that for a given proportional increase in exports, population and, therefore, wages need to rise more, which c.p. means that (in proportional terms) exports increase less in larger cities. Given elastic labor supply to each location, these wage changes also induce firm entry as the additional profits are absorbed by both higher wages and additional entry. Evidence in Table 8 shows relatively modest wage relative to population effects. This means that the population channel dominates, raising exports disproportionately in larger cities. Thus, the net effect leads to a slightly larger (more negative) trade elasticity in larger cities in the homogeneous firm model.

Table A.19: Model Comparison: Predicted Local Trade Elasticity and City Size

	(1) Beta	(2) Std. Error.	(3) R^2
Panel A: China			
(1) Baseline: City-specific Pareto shape and scale parameters	0.137***	(0.021)	0.05
(2) Homogeneous firms model: City-specific average productivity	-0.018***	(0.006)	0.06
(3) Baseline without internal trade costs	0.151***	(0.020)	0.06
(4) Baseline with common shape parameter	-0.018***	(0.006)	0.01
Panel B: France			
(1) Baseline: City-specific Pareto shape and scale parameters	0.164***	(0.022)	0.24
(2) Homogeneous firms model: City-specific average productivity	-0.008	(0.005)	0.28
(3) Baseline without internal trade costs	0.172***	(0.021)	0.28
(4) Baseline with common shape parameter	-0.008	(0.005)	0.01

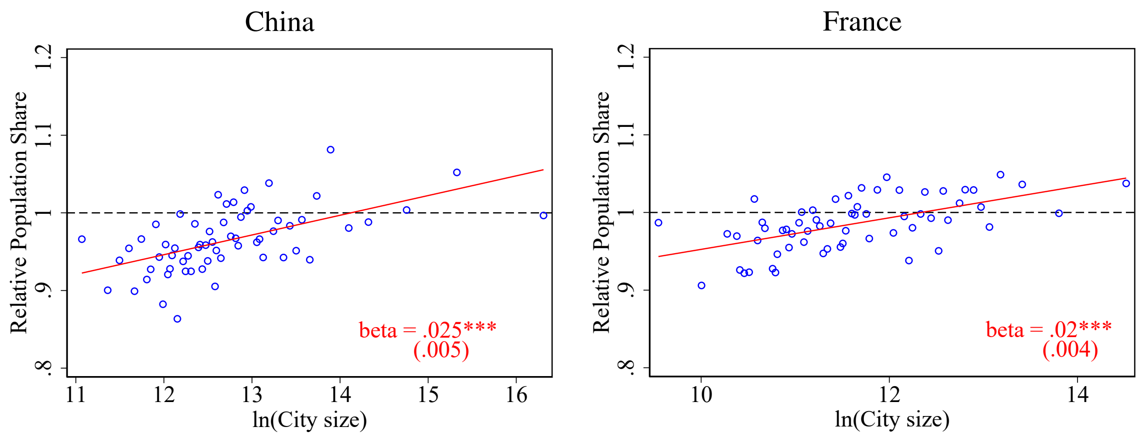
Notes: The table shows the estimated fitted values of a regression of the local international trade elasticity implied by the different versions of the model presented in Section 5.5 and discussed in more detail in Appendix F against city size (in natural logarithms). Column 1 shows the estimated coefficient, while columns 2 and 3 show the robust standard error (in parenthesis) and the R-squared of the regression, respectively. Row 1 presents results for our baseline model that allows for city-varying productivity dispersion, with city-specific Pareto shape Parameters. Row 2 presents results for a version of the model with homogeneous firms. Rows 3 and 4 present results for the baseline model but shutting down domestic trade costs and restricting the Pareto shape parameter to be the same for all cities, respectively.

E.6 Counterfactual City Population in Trade vs. Autarky

This appendix compares city population under trade vs. autarky for a variety of models. Figure A.4 shows, for each model, the population share for each location in the actual trade relative to the autarky equilibrium (on the y-axis), plotted in a binscatter against city size (x-axis). Values above one on the y-axis indicate a larger population share in the location under the actual trade equilibrium. The red-solid line in each panel represents the fitted value between the population share and city size. Panel A visualizes the differences in population reallocation from autarky to trade along the city size distribution for the baseline model. Following a reduction in trade costs, smaller cities shrink while larger cities grow. In quantitative terms, raising city size by 1 log-point increases the population share in the trade equilibrium relative to autarky by 2.5 percentage points in China and 2.0 percentage points in France. Importantly, all competing models deliver smaller spatial reallocation. Panel B visualizes the differences in population reallocation from autarky to trade under a model with homogeneous firms, while Panels C and D show the population reallocation in the baseline model, shutting down domestic trade costs and restricting the Pareto shape parameter to be the same in all cities, respectively. Interestingly, restricting the upper tail of the productivity distributions to be the same across cities (Panel D) leads to similar effects as a model with homogeneous firms (Panel B). Shutting down internal trade costs (Panel C) reduces the spatial reallocation relative to the baseline (Panel A), although significantly more in China than in France.

Figure A.4: Comparing Models: City Population in Trade vs. Autarky

Panel A. Baseline: City-Specific Pareto Shape and Scale Parameters



Panel B. Homogeneous Firms Model: City-Specific Average Productivity

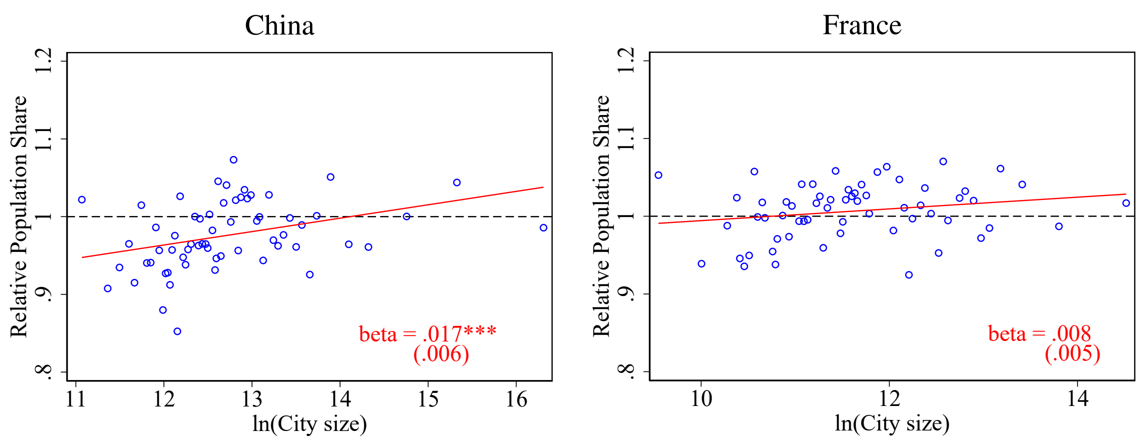
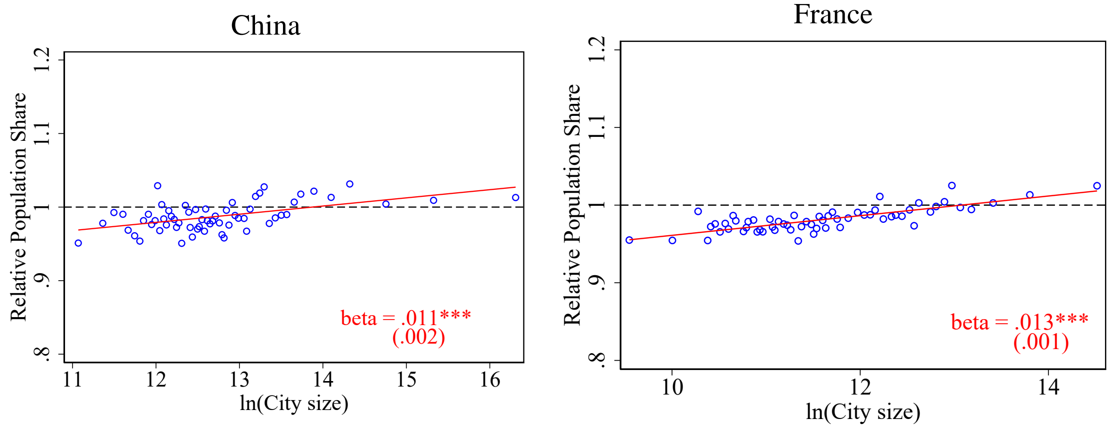
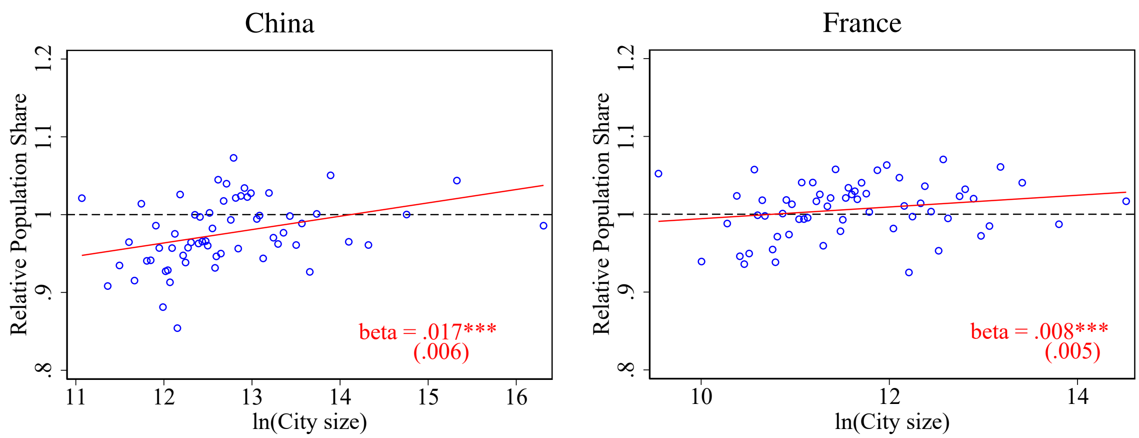


Figure A.4: Comparing Models: City Population in Trade vs. Autarky (cont.)

Panel C. Baseline without Internal Trade Costs



Panel D. City-Specific Scale, Common Shape Parameter Across Cities



Notes: These figures plot the relative change in population from autarky to the data-implied equilibrium (with trade) against population size for all cities in China and France in various models. Each dot (bin) represents 10 underlying cities for China, and 5 cities for France. On the y-axis, values above one indicate a larger population share in the location under the trade equilibrium.The baseline model and its calibration are described in Sections 5.1, 5.2 and in Appendix E.3, while Appendix F provides details on the other models.

F Details on Comparison Models

F.1 Heterogeneous Firm Model with Constant Shape Parameter

Model equilibrium. The model with constant shape parameters is a simplified version of our baseline model that omits heterogeneity in the upper tail of the city-level firm productivity distributions as a potential mechanism for accounting for city-level differences in export intensity, while allowing differences in (foreign) market access across cities to affect city level export intensities via both the extensive and the intensive margins. The system of equations describing the equilibrium of this simplified model can be described by setting $\alpha_i = \alpha$ for all cities i in the system of equations describing the equilibrium of our baseline model. This results in the following system of equations:

$$w_i L_i = \sum_{n \in N \cup \{c\}} \pi_{ni} w_n L_n \tag{A.128}$$

$$\pi_{ni} = \frac{p_i^d M_i^e (w_i d_{ni})^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k d_{nk})^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.129}$$

$$\pi_{ci} = \frac{p_i^x M_i^e (w_i d_{ci})^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k d_{ck})^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.130}$$

$$\pi_{cc} = \frac{p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k d_{ck})^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.131}$$

$$\pi_{nc} = \frac{p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k d_{nk})^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{nc})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.132}$$

$$\tilde{\psi}_i^d = \left[\frac{1}{1 - G_i(\psi_i^{d*})} \int_{\psi_i^{d*}} \psi^{\sigma-1} g_i(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha}{\alpha - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_i^{d*} \tag{A.133}$$

$$\tilde{\psi}_i^x = \left[\frac{1}{1 - G_i(\psi_i^{x*})} \int_{\psi_i^{x*}} \psi^{\sigma-1} g_i(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha}{\alpha - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_i^{x*} \tag{A.134}$$

$$\tilde{\psi}_c^d = \left[\frac{1}{1 - G_i(\psi_c^{d*})} \int_{\psi_c^{d*}} \psi^{\sigma-1} g_c(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha}{\alpha - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_c^{d*} \tag{A.135}$$

$$\tilde{\psi}_c^x = \left[\frac{1}{1 - G_c(\psi_c^{x*})} \int_{\psi_c^{x*}} \psi^{\sigma-1} g_c(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha}{\alpha - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_c^{x*} \tag{A.136}$$

$$p_i^d = \left(\frac{A_i}{\psi_i^{d*}}\right)^{\alpha} \tag{A.137}$$

$$p_i^x = \left(\frac{A_i}{\psi_i^{x*}}\right)^{\alpha} \tag{A.138}$$

$$p_c^d = \left(\frac{A_c}{\psi_c^{d*}}\right)^{\alpha} \tag{A.139}$$

$$p_c^x = \left(\frac{A_c}{\psi_c^{x*}}\right)^{\alpha} \tag{A.140}$$

$$P_i = \left(\frac{p_i^d M_i^e}{\pi_{ii}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_i}{\tilde{\psi}_i^d}\right) \tag{A.141}$$

$$P_c = \left(\frac{p_c^d M_c^e}{\pi_{cc}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_c}{\tilde{\psi}_c^d}\right) \tag{A.142}$$

$$M_i^e = \frac{L_i}{\sigma\left[F_e + p_i^d F + p_i^x F_x\right]} \tag{A.143}$$

$$M_c^e = \frac{L_c}{\sigma\left[F_e + p_c^d F + p_c^x F_x\right]} \tag{A.144}$$

$$\frac{\psi_i^{x*}}{\psi_i^{d*}} = \left(\frac{F_x}{F_d}\frac{\sum_{k\in N} w_k L_k P_k^{\sigma-1} d_{ki}^{1-\sigma}}{w_c L_c P_c^{\sigma-1} d_{ci}^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.145}$$

$$\frac{\psi_c^{x*}}{\psi_c^{d*}} = \left(\frac{F_x}{F_d}\frac{w_c L_c P_c^{\sigma-1}}{\sum_{k\in N} w_k L_k P_k^{\sigma-1} d_{kc}^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.146}$$

$$F_e = \left(p_i^d F_d + p_i^x F_x\right)\left(\frac{\sigma-1}{\alpha-\sigma+1}\right) \tag{A.147}$$

$$F_e = \left(p_c^d F_d + p_c^x F_x\right)\left(\frac{\sigma-1}{\alpha-\sigma+1}\right) \tag{A.148}$$

$$L_i = \frac{B_i^{\frac{1}{1+\gamma\epsilon(1-\beta)}}\left[w_i^{(1-\beta)(1-\gamma)}\left(\frac{\pi_{ii}}{p_i^d M_i^e}\right)^{\frac{\beta}{1-\sigma}}(\tilde{\psi}_i^d)^{\beta} N_i^{\gamma(1-\beta)}\right]^{\frac{\epsilon}{1+\gamma\epsilon(1-\beta)}}}{\left\{\sum_{k\in N} B_k\left[w_k^{(1-\beta)(1-\gamma)}\left(\frac{\pi_{kk}}{p_k^d M_k^e}\right)^{\frac{\beta}{1-\sigma}}(\tilde{\psi}_k^d)^{\beta}\left(\frac{N_k}{L_k}\right)^{\gamma(1-\beta)}\right]^{\epsilon}\right\}^{\frac{1}{1+\gamma\epsilon(1-\beta)}}}\bar{L}^{\frac{1}{1+\gamma\epsilon(1-\beta)}} \tag{A.149}$$

Where equation (A.128) is a market clearing equation (i.e. expenditure equal income for each location), equations (A.129) to (A.132) define the expenditure shares, equations (A.133) to (A.136) pin down the average productivities of domestic and exporting firms at each location, equations (A.137) to (A.140) pin down the probabilities with which domestic and foreign markets are served by firm entrants at each location, equations (A.141) and (A.142) pin down the price indices at all domestic and foreign locations, equations (A.143) and (A.144) are the labor market clearing conditions at each location, equations (A.145) and (A.146) pin down the ratios of the productivity thresholds to serve foreign vs domestic markets at each locations, which are themselves pinned down by the relevant zero profit conditions for marginal entrants, equations (A.147) and (A.148) are the free entry conditions at each location, while equation (A.149) is the national labor market clearing condition.

Calibration strategy. The calibration strategy for this model is identical to the one described in Appendix E.3 but uses the same shape parameter for all cities. This parameter is estimated with QQ regressions using the 10 percent largest firms across all cities in China.

F.2 Homogeneous Firm Model

Model equilibrium. Assuming away within city-firm heterogeneity, our model collapses to the increasing returns model in Redding (2016). In this model city level export intensity is pinned down by each city's foreign market access relative to domestic market access, and all cross-city variation in export intensity is driven by the intensive margin. The equilibrium in this simplified model is characterized by the following system of equations:

$$w_i L_i = \sum_{n \in N \cup \{c\}} \pi_{ni} w_n L_n \tag{A.150}$$

$$\pi_{ni} = \frac{L_i (\frac{w_i d_{ni}}{A_i})^{1-\sigma}}{\sum_{k \in N} L_k (\frac{w_k d_{nk}}{A_k})^{1-\sigma} + L_c (\frac{w_c d_{nc}}{A_c})^{1-\sigma}} \tag{A.151}$$

$$P_i = \left(\frac{L_i}{\sigma F \pi_{ii}} \right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma - 1} \frac{w_i}{A_i} \right) \tag{A.152}$$

$$\frac{L_i}{\bar{L}} = \frac{B_n (v_i / P_i^{\beta} (p_i^h)^{1-\beta})^{\epsilon}}{\sum_{k \in N} B_k (v_k / P_k^{\beta} (p_k^h)^{1-\beta})^{\epsilon}} \tag{A.153}$$

$$v_i = \frac{w_i}{\beta} \tag{A.154}$$

$$r_i = \gamma \frac{1-\beta}{\beta} \frac{w_i L_i}{H_i} \tag{A.155}$$

$$r_c = \gamma \frac{1-\beta}{\beta} \frac{w_c L_c}{H_c} \tag{A.156}$$

$$M_i = \frac{L_i}{\sigma F_d} \tag{A.157}$$

$$M_c = \frac{L_c}{\sigma F_d} \tag{A.158}$$

$$\bar{U} = \delta \left[\sum_{k \in N} B_k (v_k / P_k^{\beta} (p_k^h)^{1-\beta})^{\epsilon} \right]^{\frac{1}{\epsilon}} \tag{A.159}$$

where p_i^h is given by equation (18). Equation (A.150) is a market-clearing condition, (A.151) pins down the expenditure share of each source i in destination n, equation (A.152) pins down the price index at each location, equation (A.153) is national labour market clearing, while equations

(A.154) and (A.155) represent local land market clearing conditions. Finally, equations (A.156) and (A.157) represent local labor market clearing conditions.

Calibration strategy. The calibration strategy follows the one of the baseline model described in appendix E.3 with the following distinctions: Since there is no firm heterogeneity, we do not need to calibrate α_i, which also implies that there is no selection such that we set F_d and F_x to 0.

F.3 Baseline Model without Internal Variable Trade Cost

Model equilibrium. Assuming away heterogeneity in domestic and foreign market access results in a version of our quantitative model that is very similar to the simple model outlined in section 4. This allows us to study how important the key mechanisms outlined in our theory is in driving both the overall configuration of economic geography and overall trade patterns, in the absence of the forces typically present in standard QSE models. The system of equations characterizing the equilibrium in this simplified setting is derived by setting $d_{ni} = 1$ for all n and i and $d_{nc} = d_{cn} = \tau$ for all n, as well as noting that under these circumstances price indices are equalized across space in the home country and thus to composite consumption good can be chosen as the numeraire. This delivers the following set of equilibrium conditions:

$$w_i L_i = \sum_{n \in N \cup \{c\}} \pi_{ni} w_n L_n \tag{A.160}$$

$$\pi_{ni} = \frac{p_i^d M_i^e (w_i)^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k)^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c \tau)^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.161}$$

$$\pi_{ci} = \frac{p_i^x M_i^e (w_i \tau)^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k \tau)^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.162}$$

$$\pi_{cc} = \frac{p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}}{\sum_{k \in N} p_k^x M_k^e (w_k \tau)^{1-\sigma} (\tilde{\psi}_k^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.163}$$

$$\pi_{nc} = \frac{p_c^x M_c^e (w_c \tau)^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}}{\sum_{k \in N} p_k^d M_k^e (w_k)^{1-\sigma} (\tilde{\psi}_k^d)^{\sigma-1} + p_c^x M_c^e (w_c \tau)^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.164}$$

$$\tilde{\psi}_i^d = \left[\frac{1}{1 - G_i(\psi_i^{d*})} \int_{\psi_i^{d*}} \psi^{\sigma-1} g_i(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_i^{d*} \tag{A.165}$$

$$\tilde{\psi}_i^x = \left[\frac{1}{1 - G_i(\psi_i^{x*})} \int_{\psi_i^{x*}} \psi^{\sigma-1} g_i(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_i^{x*} \tag{A.166}$$

$$\tilde{\psi}_c^d = \left[\frac{1}{1 - G_i(\psi_c^{d*})} \int_{\psi_c^{d*}} \psi^{\sigma-1} g_c(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_c}{\alpha_c - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_c^{d*} \tag{A.167}$$

$$\tilde{\psi}_c^x = \left[\frac{1}{1 - G_c(\psi_c^{x*})} \int_{\psi_c^{x*}} \psi^{\sigma-1} g_c(\psi) d\psi \right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_c}{\alpha_c - (\sigma - 1)} \right)^{\frac{1}{\sigma-1}} \psi_c^{x*} \tag{A.168}$$

$$p_i^d = \left(\frac{A_i}{\psi_i^{d*}}\right)^{\alpha_i} \tag{A.169}$$

$$p_i^x = \left(\frac{A_i}{\psi_i^{x*}}\right)^{\alpha_i} \tag{A.170}$$

$$p_c^d = \left(\frac{A_c}{\psi_c^{d*}}\right)^{\alpha_c} \tag{A.171}$$

$$p_c^x = \left(\frac{A_c}{\psi_c^{x*}}\right)^{\alpha_c} \tag{A.172}$$

$$P_c = \left(\frac{p_c^d M_c^e}{\pi_{cc}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_c}{\tilde{\psi}_c^d}\right) \tag{A.173}$$

$$M_i^e = \frac{L_i}{\sigma\left[F_e + p_i^d F + p_i^x F_x\right]} \tag{A.174}$$

$$M_c^e = \frac{L_c}{\sigma\left[F_e + p_c^d F + p_c^x F_x\right]} \tag{A.175}$$

$$\frac{\psi_i^{x*}}{\psi_i^{d*}} = \left(\frac{F_x}{F_d}\frac{\sum_{k\in N} w_k L_k}{w_c L_c P_c^{\sigma-1}\tau^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.176}$$

$$\frac{\psi_c^{x*}}{\psi_c^{d*}} = \left(\frac{F_x}{F_d}\frac{w_c L_c P_c^{\sigma-1}}{\sum_{k\in N} w_k L_k \tau^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.177}$$

$$F_e = (p_i^d F_d + p_i^x F_x)\left(\frac{\sigma-1}{\alpha_i-\sigma+1}\right) \tag{A.178}$$

$$F_e = (p_c^d F_d + p_c^x F_x)\left(\frac{\sigma-1}{\alpha_c-\sigma+1}\right) \tag{A.179}$$

$$\frac{L_i}{\bar{L}} = \frac{B_i(v_i/(p_i^h)^{1-\beta})^{\epsilon}}{\sum_{k\in N} B_k(v_k/(p_k^h)^{1-\beta})^{\epsilon}} \tag{A.180}$$

$$v_i = \frac{w_i}{\beta} \tag{A.181}$$

Where equation (A.160) is a market clearing equation (i.e. expenditure equal income for each location), equations (A.161) to (A.164) define the expenditure shares of each source location i to each destination n, equations (A.165) to (A.168) pin down the average productivities of domestic and exporting firms at each location, equations (A.169) to (A.172) pin down the probabilities with which domestic and foreign markets are served by firm entrants at each location, equation (A.142) pins down the price indicx of the foreign country (all domestic price indices are equal to 1), equations (A.174) and (A.175) are the labor market clearing conditions at each location, equations (A.176) and (A.177) pin down the ratios of the productivity thresholds to serve foreign vs domestic markets at each locations, which are themselves pinned down by the relevant zero profit conditions for marginal entrants, equations (A.178) and (A.179) are the free entry conditions at each location, while equation (A.180) is the national labor market clearing condition.

Calibration strategy. The baseline model without geography departs from the one with geography by defining the internal iceberg cost from location n to i (d_{ni}) and the distance from n to the nearest point of exit (d_{np}) both equal to 1. Conditional on these changes to calibrated parameters, calibration proceeds exactly as in the baseline model (see appendix E.3).

F.4 Melitz Model

Model equilibrium. As we are also interested in the effects of explicitly modeling within-country economic geography on international trade counterfactuals, we also study a Melitz (2003) version of our model, where we model both the domestic and foreign country as having only one region and hence only one country-level productivity distribution that firms are drawn from. The relevant equations are outlined below:

$$w_i L_i = \pi_{ii} w_i L_i + \pi_{ci} w_c L_c \tag{A.182}$$

$$w_c L_c = \pi_{cc} w_c L_c + \pi_{ic} w_i L_i \tag{A.183}$$

$$\pi_{ii} = \frac{p_i^d M_i^e (w_i)^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1}}{p_i^d M_i^e (w_i)^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{ic})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.184}$$

$$\pi_{ci} = \frac{p_i^x M_i^e (w_i d_{ci})^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1}}{p_i^x M_i^e (w_i d_{ci})^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.185}$$

$$\pi_{cc} = \frac{p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}}{p_i^x M_i^e (w_i d_{ci})^{1-\sigma} (\tilde{\psi}_i^x)^{\sigma-1} + p_c^d M_c^e (w_c)^{1-\sigma} (\tilde{\psi}_c^d)^{\sigma-1}} \tag{A.186}$$

$$\pi_{ic} = \frac{p_c^x M_c^e (w_c d_{ic})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}}{p_i^d M_i^e (w_i)^{1-\sigma} (\tilde{\psi}_i^d)^{\sigma-1} + p_c^x M_c^e (w_c d_{ic})^{1-\sigma} (\tilde{\psi}_c^x)^{\sigma-1}} \tag{A.187}$$

$$\tilde{\psi}_i^d = \left[\frac{1}{1 - G_i(\psi_i^{d*})} \int_{\psi_i^{d*}} \psi^{\sigma-1} g_i(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - (\sigma-1)}\right)^{\frac{1}{\sigma-1}} \psi_i^{d*} \tag{A.188}$$

$$\tilde{\psi}_i^x = \left[\frac{1}{1 - G_i(\psi_i^{x*})} \int_{\psi_i^{x*}} \psi^{\sigma-1} g_i(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_i}{\alpha_i - (\sigma-1)}\right)^{\frac{1}{\sigma-1}} \psi_i^{x*} \tag{A.189}$$

$$\tilde{\psi}_c^d = \left[\frac{1}{1 - G_c(\psi_c^{d*})} \int_{\psi_c^{d*}} \psi^{\sigma-1} g_c(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_c}{\alpha_c - (\sigma-1)}\right)^{\frac{1}{\sigma-1}} \psi_c^{d*} \tag{A.190}$$

$$\tilde{\psi}_c^x = \left[\frac{1}{1 - G_c(\psi_c^{x*})} \int_{\psi_c^{x*}} \psi^{\sigma-1} g_c(\psi) d\psi\right]^{\frac{1}{\sigma-1}} = \left(\frac{\alpha_c}{\alpha_c - (\sigma-1)}\right)^{\frac{1}{\sigma-1}} \psi_c^{x*} \tag{A.191}$$

$$p_i^d = \left(\frac{A_i}{\psi_i^{d*}}\right)^{\alpha_i} \tag{A.192}$$

$$p_i^x = \left(\frac{A_i}{\psi_i^{x*}}\right)^{\alpha_i} \tag{A.193}$$

$$p_c^d = \left(\frac{A_c}{\psi_c^{d*}}\right)^{\alpha_c} \tag{A.194}$$

$$p_c^x = \left(\frac{A_c}{\psi_c^{x*}}\right)^{\alpha_c} \tag{A.195}$$

$$P_i = \left(\frac{p_i^d M_i^e}{\pi_{ii}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_i}{\tilde{\psi}_i^d}\right) \tag{A.196}$$

$$P_c = \left(\frac{p_c^d M_c^e}{\pi_{cc}}\right)^{\frac{1}{1-\sigma}} \left(\frac{\sigma}{\sigma-1}\frac{w_c}{\tilde{\psi}_c^d}\right) \tag{A.197}$$

$$M_i^e = \frac{L_i}{\sigma\left[F_e + p_i^d F + p_i^x F_x\right]} \tag{A.198}$$

$$M_c^e = \frac{L_c}{\sigma\left[F_e + p_c^d F + p_c^x F_x\right]} \tag{A.199}$$

$$\frac{\psi_i^{x*}}{\psi_i^{d*}} = \left(\frac{F_x}{F_d}\frac{w_i L_i P_i^{\sigma-1}}{w_c L_c P_c^{\sigma-1} d_{ci}^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.200}$$

$$\frac{\psi_c^{x*}}{\psi_c^{d*}} = \left(\frac{F_x}{F_d}\frac{w_c L_c P_c^{\sigma-1}}{w_i L_i P_i^{\sigma-1} d_{ic}^{1-\sigma}}\right)^{\frac{1}{\sigma-1}} \tag{A.201}$$

$$F_e = (p_i^d F_d + p_i^x F_x)\left(\frac{\sigma-1}{\alpha_i - \sigma + 1}\right) \tag{A.202}$$

$$F_e = (p_c^d F_d + p_c^x F_x)\left(\frac{\sigma-1}{\alpha_c - \sigma + 1}\right) \tag{A.203}$$

$$L_i = \overline{L} \tag{A.204}$$

$$L_c = \overline{L}_c \tag{A.205}$$

Where equations (A.182) and (A.183) are market clearing equations (i.e. expenditure equal income for each country), equations (A.184) to (A.187) define the expenditure shares of each source-destination pair, equations (A.188) to (A.191) pin down the average productivities of domestic and exporting firms in each country, equations (A.192) to (A.195) pin down the probabilities with which domestic and foreign markets are served by firm entrants in each country, equations (A.196) and (A.197) pin down the price indices of each country, equations (A.198) and (A.199) are the labor market clearing conditions of each country, equations (A.200) and (A.201) pin down the ratios of the productivity thresholds to serve foreign vs domestic markets for each country, which are themselves pinned down by the relevant zero profit conditions for marginal entrants, equations (A.202) and (A.203) are the free entry conditions for each country, while equations (A.204) and (A.205) are the national labor market clearing conditions for each country.

Calibration strategy. We model the domestic country as a single location with a land endowment equal to the sum of the land endowments of cities in our baseline model. We perform two further modifications. First, we set the Pareto shape parameter of the single domestic location to the value estimated in model F.1. Second, we set the amenity level in the single domestic location to $B_i = 1$. Conditional on these changes to calibrated parameters, calibration proceeds exactly as in the baseline model (see appendix E.3), where the main change is that at step 3 we only employ data on average (country level) wages to pin down the country level productivity (location) parameter A_i.

References

Allen, T. and C. Arkolakis (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics 129*(3), 1085–1140.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Au, C.-C. and J. V. Henderson (2006). Are Chinese Cities Too Small? *Review of Economic Studies 73*(3), 549–576.

Bai, Y. and R. Jia (2021). The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, AD1000-2000. Technical report. forthcoming.

Bellemare, M. F. and C. J. Wichman (2020, February). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61.

Bleakley, H. and J. Lin (2012). Portage and path dependence. *The quarterly journal of economics 127*(2), 587–644.

Bryan, G. and M. Morten (2019). The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy 127*(5), 2229–2268.

Chaney, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review 98*(4), 1707–1721.

Chengrong, D., L. Lidan, and C. Mengyao (2018). Domestic Migrants in China (2018): Trends, Challenges and Recommendations. Report.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, Volume 5, pp. 247–348. Elsevier.

Combes, P.-P. and H. G. Overman (2004). The spatial distribution of economic activities in the European Union. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, Chapter 64, pp. 2845–2909. Elsevier.

Davis, M. A. and F. Ortalo-Magne (2011, April). Household Expenditures, Wages, Rents. *Review of Economic Dynamics 14*(2), 248–261.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica 82*(4), 1303–1340.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier.

Eaton, J., S. Kortum, and F. Kramarz (2011). An Anatomy of International Trade: Evidence From French Firms. *Econometrica 79*(5), 1453–1498.

Egger, P. H., G. Loumeau, and N. Loumeau (2023). China's dazzling transport-infrastructure growth: Measurement and effects. *Journal of International Economics 142*(103734), 1–22.

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ. 1*(1), 255–294.

Galor, O. and O. Özak (2016, October). The Agricultural Origins of Time Preference. *American Economic Review 106*(10), 3064–3103.

Gaubert, C. (2018). Firm Sorting and Agglomeration. *American Economic Review 108*(11), 3117–3153.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of political economy 100*(6), 1126–1152.

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and trade without pareto. *American Economic Review 104*(5), 310–316.

Imbert, C., J. Monras, M. Seror, and Y. Zylberberg (2025). Floating population: migration with (out) family and the spatial distribution of economic activity.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2022). Migrants and firms: Evidence from china. *American Economic Review 112*(6), 1885–1914.

Kline, P. and E. Moretti (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics 129*(1), 275–331.

McCormick, M., G. Huang, G. Zambotti, and J. Lavash (2013). Roman Road Network (version 2008).

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Nunn, N. and N. Qian (2011). The Potato's Contribution to Population and Urbanization: Evidence from an Historical Experiment. *Quarterly Journal of Economics 126*(2), 593–650.

Pierce, J. R. and P. K. Schott (2016). The surprisingly swift decline of us manufacturing employment. *American Economic Review 106*(7), 1632–62.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Redding, S. J. and D. M. Sturm (2008, December). The costs of remoteness: Evidence from german division and reunification. *American Economic Review 98*(5), 1766–97.

Roca, J. D. L. and D. Puga (2017). Learning by working in big cities. *The Review of Economic Studies 84*(1), 106–142.

Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2119–2171. Elsevier.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *The Quarterly Journal of Eco-*

nomics 125(3), 1253–1296.

Wang, S., S. H. Chan, and B. Xu (2012). The estimation and determinants of the price elasticity of housing supply: Evidence from china. *Journal of Real Estate Research 34*(3), 311–344.