NBER WORKING PAPER SERIES

CITIES, HETEROGENEOUS FIRMS, AND TRADE

Jan David Bakker
Alvaro Garcia Marin
Andrei V. Potlogea
Nico Voigtländer
Yang Yang

Working Paper 32377
http://www.nber.org/papers/w32377

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2024, Revised October 2025

We would like to thank Dominick Bartelme, Pierre-Phillipe Combes, Fabian Eckert, Pablo Fajgelbaum, Oleg Itskhoki, Tim Kehoe, Isabelle Mejean, Fernando Parro, Thomas Sampson, Oren Ziv as well as numerous seminar and conference audiences. Bakker thanks his advisors and members of his committee, Kalina Manova, Ferdinand Rauch, Daniel Sturm, Gabriel Ulyssea, and Tony Venables. Francisco Diaz-Valdes, Marcel Sifan Kammerer, and Alexander Valenzuela provided superb research assistance. This paper combines earlier independent work from Bakker's dissertation and from Garcia-Marin, Potlogea, Voigtländer and Yang. We gratefully acknowledge research support from UK Research and Innovation (ESRC New Investigator Grant, Grant Reference ES/W008017/1, project title "Cities, Trade and Productivity"), from the UCLA Ziman Center's Rosalinde, Arthur Gilbert Program in Real Estate, Finance and Urban Economics, and from CONICYT-Chile through projects Fondecyt Regular 1191536 and 1241616. Corresponding author: Nico Voigtländer (nico.v@ucla.edu). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Cities, Heterogeneous Firms, and Trade
Jan David Bakker, Alvaro Garcia Marin, Andrei V. Potlogea, Nico Voigtländer, and Yang Yang
NBER Working Paper No. 32377
April 2024, Revised October 2025
JEL No. F1, R1

ABSTRACT

Does international trade affect the growth of cities, and vice versa? Assembling disaggregate data for four countries, we document a novel stylized fact: Export activity is disproportionately concentrated in larger cities – even more so than overall economic activity. We rationalize this fact by marrying a standard quantitative spatial economics model with a heterogeneous firm model that features selection into the domestic and the export market. Our model delivers novel predictions for the bi-directional interactions between trade and urban dynamics: On the one hand, trade liberalization shifts employment towards larger cities, and on the other hand, liberalizing land use raises exports. We structurally estimate the model using Chinese and French firm-level data. We find that trade policies have quantitatively meaningful impacts on urban outcomes and vice versa, and that the aggregate effects of trade and urban policies differ from standard models that do not account for the interaction between trade and cities. In addition, a distinguishing prediction of our model – which we confirm in the data – is that local trade elasticities vary systematically with city size, so that a country's aggregate trade elasticity depends on the spatial distribution of production within its borders.

Jan David Bakker
Bocconi University
Department of Economics
jan.bakker@unibocconi.it

Alvaro Garcia Marin
Universidad de los Andes, Chile
Department of Economics
alvarof.garciam@gmail.com

Andrei V. Potlogea
The University of Edinburgh
School of Economics
Andrei.Potlogea@ed.ac.uk

Nico Voigtländer
University of California, Los Angeles
Anderson School of Management
and NBER
nico.v@anderson.ucla.edu

Yang Yang
International Monetary Fund
yyang5@imf.org

A data appendix is available at http://www.nber.org/data-appendix/w32377

1 Introduction

Over the last decades, two mega-trends have shaped economies across the globe: rapid urbanization and a surge in international trade.[1] These trends have traditionally been examined by two separate strands of literature – international trade and economic geography.[2] We ask whether the growth of cities and trade are related and what implications such a connection would have for welfare and income inequality.

We examine the relationship between city size and 'export intensity' (city-level exports relative to overall city-level revenues). Using data for China, France, Brazil, and the United States, we establish a novel stylized fact: Larger cities systematically export a higher fraction of their output. The majority of this association can be attributed to variation *within* industries. Thus, differences in industrial composition (such as car manufacturing vs. real estate services) cannot explain why exporting is concentrated in larger cities. Neither can systematic differences in variable trade costs, as we show by several complementary pieces of evidence: First, including proxies for trade costs and market access does not change our results. Second, we use historical soil suitability as an instrument, which is a strong predictor of city size but is not related to transport infrastructure. Cities that became large because they were surrounded by fertile land have a significantly higher export intensity today. Finally, the intensive margin (i.e., the export intensity of exporting firms, through which systematic differences in variable trade costs would work) is relatively unimportant. Instead, we show that the extensive margin is playing a central role: In larger cities, a higher proportion of firms participate in export markets. This implies that heterogeneous firms play an important role in giving rise to the novel stylized fact.

Motivated by our empirical findings, we propose a simple model that marries core mechanisms from international trade and urban economics. The model features multiple locations and – in each location – heterogeneous firms that select into serving the domestic and export market (Melitz, 2003). We first highlight this mechanism in a simple Rosen-Roback framework without domestic trade costs and symmetric countries. In equilibrium, larger cities host more productive firms, as only those can profitably cover the higher wages. Crucially, however, differences in *average* productivity across cities are not sufficient to produce differences in export *intensity*. This results because average productivity affects the domestic and the export entry thresholds proportionately so that the ratio of exports relative to total output of a location remains the same. However, adding a well-known empirical regularity to the model can replicate the disproportionate concentration of exporters in larger cities: We introduce differences in the upper tail of the productivity distribution

[1] The worldwide average urbanization rate grew from 37 to 57 percent between 1970 and 2022. During the same period, exports as a share of GDP almost tripled, from 13 to 31 percent (https://data.worldbank.org/indicator).

[2] Recently, a literature at the intersection of these fields has emerged, but it focuses more broadly on the spatial effects of trade (and its sectoral patterns) within countries, thus speaking only indirectly to urbanization. We discuss this literature in detail below.

across locations, reflecting the empirical pattern documented in the literature on city size and productivity (see Combes, Duranton, Gobillon, Puga, and Roux, 2012) – a productivity pattern that we also document in our data (see Figure 2). In our model, entrants in more productive (and hence larger) cities draw from Pareto distributions with thicker upper tails.[3] This increases the proportion of high-productivity firms (which select into exporting), leading to a disproportionate concentration of exporting in larger cities.[4]

Our simple model delivers another novel prediction: The (international) trade elasticity varies across space and is *smaller* for larger cities. This spatial heterogeneity in trade elasticities is driven by the different shapes of the local productivity distributions.[5] While our key assumptions naturally lead to this prediction, it does not follow from workhorse spatial models without differences in the upper tail of the firm productivity distribution (e.g., Allen and Arkolakis, 2014; Redding, 2016). Thus, we can use this prediction to validate the core assumption of our model vis-à-vis alternative mechanisms. We find strong support, using the identification strategy from Pierce and Schott (2016) for Chinese exporters. Our results show that in large cities, Chinese exports reacted significantly less elastically to trade costs after China was granted permanent normal trade relations (PNTR) with the U.S. in 2001.

In addition to explaining these patterns in the data, our model delivers novel predictions that are relevant for policy-making, highlighting important interactions between trade and economic geography. First, a reduction in international trade costs shifts economic activity towards larger cities. Second, the model leads to a novel mechanism that connects trade and wage inequality. A large literature has argued that trade can raise inequality. Our model makes a similar prediction, but with an important twist: Trade liberalization shifts population towards larger and more productive cities, where nominal wages are higher. Thus, trade raises *nominal* wage inequality across space. To the best of our knowledge, we are the first to identify this spatial effect of trade on (nationwide) wage inequality. However, prices are also higher in larger cities. Hence, using nominal wages as the sole metric may mislead policymakers, as it exaggerates the effect of trade on inequality. Third, the model also allows us to study the reverse relationship – between urbanization and international trade. This leads to our fourth prediction: Deregulating housing supply raises nationwide exports

[3]We are – to the best of our knowledge – the first to explore the implications of the thicker productivity upper tail in larger cities for the spatial distribution of exporting, and more broadly for aggregate outcomes. We provide a microfoundation for the thicker upper tails by introducing learning from local experts into a stochastic firm growth process à la Gabaix (2009), with agglomeration leading to lower peer-meeting costs in larger cities.

[4]This goes beyond what can be explained by differences in foreign market access. In fact, an important strand of the trade literature predicts the opposite pattern: A direct implication of the gravity model is that larger cities (or countries) have *lower* export intensities (Anderson and van Wincoop, 2004) because they sell disproportionately to the local (home) market.

[5]Intuitively, the thicker upper tail in larger cities implies a relatively higher mass of firms above the export cut-off. Thus, the mass of firms that enter the export market after a reduction in variable trade costs is lower relative to the mass of firms that are already exporting.

because it leads to disproportionate growth of larger cities.

To assess the quantitative importance of our mechanism, we extend the simple model into a multi-location quantitative spatial general equilibrium framework with heterogeneous firms. This quantitative model embeds our upper-tail mechanism while incorporating the standard features emphasized in the literature – arbitrary bilateral trade costs, asymmetric countries, spatial variation in amenities, mobility frictions due to idiosyncratic tastes for locations, as well as a richer pattern of productivity differences across space – combining elements from Redding (2016) and Melitz (2003). By incorporating this full set of features, the model enables us to quantify the contribution of our mechanism relative to other forces typically highlighted in quantitative spatial economics and to conduct counterfactual exercises. We structurally estimate this model using Chinese and French firm-level data. The model can account for the bulk of the correlation between export intensity and city size in the data, and it also provides a good fit for the relationship between local trade elasticities and city size. We show that for both features, the thick productivity upper tail in large cities is the crucial driving force, while differences in trade costs or in average productivity across cities play only a limited role. To quantify the policy implications highlighted above, we perform two types of counterfactual experiments: reducing variable international trade costs and increasing the nationwide housing supply elasticity. First, moving from autarky to trade increases the spatial concentration of economic activity (measured as the share of population in the highest relative to the lowest decile of city sizes) by 11.2% (7.9%) for China (France). Second, trade benefits workers in larger relative to smaller cities as it increases real wages in the highest relative to the lowest decile of the city size distribution by 2.5% (2.4%) relative to autarky. Note that the higher gains in larger cities mean that rising house prices there only partially erode the benefits of increasing nominal wages and decreasing tradable goods prices. Third, raising the housing supply elasticity nationwide from the 25% to the 75% percentile increases exports by 4.8 percent for China and 0.9 percent for France.

Lastly, we examine whether our model has quantitatively different predictions for the welfare gains from trade and increased housing supply, when compared to standard models that do not feature interactions between trade and geography. The gains from trade in our model are smaller as compared to i) a standard heterogeneous firm model without geography, and ii) a geography model with homogeneous firms. This results from the fact that exporting is concentrated in larger cities, where wages and house prices are higher, which makes it more costly for firms to expand. This dimension is missing in standard models. Hence, our unified setup has significant implications for the welfare gains from trade and urban policies.

Our paper is related to several strands of the literature. The effect of international trade on the spatial distribution of economic activity has been one of the foundational questions of economic geography. Early – mostly theoretical – contributions by Krugman and Livas Elizondo (1996), Monfort and Nicolini (2000), and Behrens, Gaigne, Ottaviano, and Thisse (2006a,b, 2007, 2009) identify different mechanisms through which trade can increase or decrease the spatial concentration of economic activity. We bring novel empirical evidence and a new mechanism to this literature. Our model marries recent advances in spatial economics (Allen and Arkolakis, 2014; Redding, 2016) with the standard trade model with heterogeneous firms (Melitz, 2003).[6] In order to replicate the observed relationship between city size and export intensity, these standard ingredients need to be combined with a thicker upper tail of the firm productivity distribution in larger cities (as documented by Combes et al., 2012). Thus, our results highlight the practical relevance of this well-known empirical pattern.

Our work also relates to a large literature on trade and inequality (see Helpman, 2018, for a review). Our findings highlight that it is important to distinguish between nominal wages and welfare when examining how trade affects inequality. This results from local prices adjusting differentially across space – a feature that is missing in trade models and that was first systematically analyzed by Moretti (2013).[7] In addition, our result that the spatial distribution of production affects the country-level trade elasticity contributes to a growing literature on the variability of this parameter (Helpman, Melitz, and Rubinstein, 2008; Novy, 2013; Melitz and Redding, 2015; Adão, Arkolakis, and Ganapati, 2020; Carrère, Mrázová, and Neary, 2020; Lind and Ramondo, 2023).

Finally, we contribute to strands of the literature within international trade and economic geography. Regarding the former, we relate to the rich literature on heterogeneous firms in trade (c.f. Bernard and Jensen, 1999; Pavcnik, 2002; Melitz, 2003), as well as to more recent work on the heterogeneous effects of trade across locations within a country (Autor, Dorn, and Hanson, 2013; Dauth, Findeisen, and Suedekum, 2014; Cheng and Potlogea, 2020). In urban economics, our novel stylized fact adds to a growing literature on sorting, selection, and agglomeration across city size (Behrens, Duranton, and Robert-Nicoud, 2014; Eeckhout, Pinheiro, and Schmidheiny, 2014; Davis and Dingel, 2014; Gaubert, 2018; Davis and Dingel, 2019; Fajgelbaum and Gaubert, 2020; Schoefer and Ziv, 2022; Oh, 2023), and our policy counterfactual on housing supply constraints speaks to the literature that quantifies their importance (c.f. Hsieh and Moretti, 2019). Exploring the systematic differences in the effects of trade across cities of different sizes also relates to a growing literature on regional divergence across cities in advanced economies: Giannone (2022) studies the role of skill-biased technical change, Eckert, Ganapati, and Walsh (2022) emphasize the importance in the rise of tradable services, Walsh (2023) underlines the importance of new firm creation, and Chen, Novy, Perroni, and Wong (2023) study the role of trade and structural

[6] The quantitative part of our paper is also related to Coşar and Fajgelbaum (2016) and Fajgelbaum and Redding (2022), who use quantitative spatial equilibrium models to highlight the role of domestic transport costs for the local effects of country-level trade openness.

[7] Note that we abstract from endogenous amenity changes as highlighted by Diamond (2016). If those adjusted endogenously following a trade shock, there would be another margin to consider when studying the effects of trade on inequality.

transformation for the growth of French cities. We show that combining the core elements of both literatures – heterogeneity across firms and across space – leads to novel insights with important policy implications.

Summarizing our contribution, this paper documents a novel complementarity between trade and urbanization: where economic activity is concentrated within a country shapes how that country engages with the global economy – and vice versa. This complementarity has three important implications. First, trade liberalization does not only reallocate activity across sectors, but also across space – shifting population and economic activity toward already large, high-wage cities. In this sense, our findings complement Baldwin's (2006; 2016) account of globalization's "great unbundlings," whereby falling transport and coordination costs enabled production and consumption to separate geographically, fostering industrial concentration. We add that these forces operate disproportionately in larger cities. Second, greater geographic concentration of economic activity deepens globalization, as large cities boost international trade. Third, the spatial concentration of firms makes globalization more resilient to trade-cost shocks: We show that the trade elasticity is systematically *lower* in larger cities. This runs counter to standard models and has direct implications for the benefits of trade. These points highlight why urban and trade policies should not be analyzed in isolation. Our counterfactual exercises demonstrate that the interactions between trade and urban features are quantitatively important, underscoring the policy relevance of our findings.

The rest of the paper is organized as follows: Section 2 presents the data and Section 3 our stylized fact and its robustness. In Section 4, we develop a simple model that illustrates the economic forces behind the stylized fact and explores their policy implications. Section 5 extends this to a quantitative model and performs counterfactual policy analyses. Section 6 concludes.

2 Data

Our main empirical analysis uses firm-level data from the 2004 Chinese Economic Census of Manufacturing and from the 2000 French Unified Corporate Statistics System (FICUS). One important advantage of the Chinese and French data is that they provide detailed information on the location of firms. This allows us to study the distribution of firms and exporters across cities. In addition, we use more aggregate information at the city level from the United States (at the MSA level) and Brazil (at the microregion level) for 2012 to confirm the main patterns we derive for China and France. We begin by discussing the Chinese and French data in detail, followed by a description of the U.S. and Brazilian data. For each country, we discuss what constitutes a "city" in our data.

2.1 China

Data for the Chinese Economic Census of Manufacturing are collected by the *National Bureau of Statistics*, covering the universe of firms in China, irrespective of their size. The Chinese data contain detailed information on plant characteristics such as sales, spending on inputs and raw materials, employment, investment, and export value. The reported location of firms reflects the county where their headquarters are based; we address this potential issue in Section 3.2.

Our main analysis defines Chinese cities as metropolitan areas with contiguous lights in night-time satellite images. We use the correspondence constructed by Dingel, Miscio, and Davis (2019) to map counties into metropolitan areas.[8] Since these urban areas are similar to Metropolitan Statistical Areas for other countries, we refer to them as "MSA." For each MSA, we use information on the urban population of the underlying counties, which is provided by the Chinese Population Census of 2010 (i.e., the Chinese Census distinguishes between rural and urban population within each county). We then define 'city size' as the aggregate urban population of the Metropolitan Area.

The Chinese Census of Manufacturing contains information for approximately 1,375,000 firms located in cities to which we can match population information in 2004. We drop firms with zero or missing sales (74,595 observations, corresponding to 5.4% of the sample), non-manufacturing or missing industry codes (134,392, 9.8% of the sample), or export intensity above 100% (6,070 observations, 0.4% of the sample). We also drop processing trade (11,215 observations, defined as firms where processing exports account for over 90% of their sales). In addition, to ensure meaningful variation in export intensity at the city level, we only consider cities and industries with at least 250 firms. Our final sample consists of 915,718 firms in 629 cities (MSAs). For auxiliary data analysis that requires a panel dimension (Section 4.4) we rely on the Annual Survey of Manufacturing (ASM), which is available at a higher frequency than the Census of Manufacturing. We do not use the ASM for our main cross-sectional results because it i) only contains data on one-fourth of the largest firms in the Census of Manufacturing and ii) the lowest geographical identifier is at the more aggregate prefecture level.

2.2 France

Our analysis for France uses firms from the 2000 Unified Corporate Statistics System (FICUS). FICUS is an administrative data set collected by the French National Statistical Institute (*Institut National de la Statistique et des Études Économiques*, INSEE), covering the universe of private sector firms. It reports information on domestic and export revenue, industry classification, headquarters location, employment, capital, value-added, and production.

We define cities in France in terms of employment zones ("Labor Market Areas"). We use

[8] We use a threshold for light intensity of 30. This value is in the middle of the set of thresholds provided by these authors. Importantly, our results do not depend on the particular light intensity threshold. Counties are the third administrative division in China, below provinces and prefectures. For reference, large metropolitan areas (urban population over 1 million) have, on average, about nine counties. In contrast, most small metropolitan areas (population below 100,000) comprise a single county. Numerous papers using information from China define cities in terms of prefecture-level cities (e.g., Brandt, Van Biesebroeck, and Zhang, 2014). We opt for the finer, county-based definition because prefecture boundaries may fragment economically integrated areas into distinct cities or circumscribe places, including rural areas. We note that our main results hold also for prefecture-level cities (see Table A.5).

the definition of employment zones published by INSEE in 2011 that assigns municipalities (*communes*) to commuting zones to create "*geographical area[s] within which most of the labour force lives and works*."[9] City size reflects the overall employment zone population, which we obtain by aggregating municipality-level data from the French Population Census of 1999. As for China, we restrict our analysis to firms with positive information on exports and sales and to cities and industries with at least 250 firms. The final sample contains all 304 employment zones in mainland France.

2.3 United States

In the case of the United States, we define cities in terms of Metropolitan Statistical Areas (MSA). MSAs are defined by the United States Office of Management and Budget as one or more adjacent counties with at least one urban area with a population of at least 50,000 inhabitants and characterized by a high degree of social and economic integration, as measured by commuting flows to work and school. As Dingel et al. (2019) show, MSAs are well-approximated by cities defined in terms of contiguous areas of lights in nighttime satellite images, making the city definition comparable to China. Our analysis considers 348 U.S. MSAs with a population over 100,000 inhabitants in 2012.

To develop our main analysis, we combine data from several sources. Data for exports at the MSA level are provided by the International Trade Administration of the U.S. Department of Commerce and include overall exports.[10] We combine this with establishment-level information of sales aggregated at the MSA level from the 2012 Economic Census.[11] Finally, we use MSA population from the population projections of the U.S. Census Bureau.[12]

2.4 Brazil

Finally, in the case of Brazil, we consider microregions as the main unit of analysis. Microregions are defined by the Brazilian Institute of Geography and Statistics (IBGE) as urban agglomerations of economically integrated contiguous municipalities with similar geographic and productive characteristics.[13] Although microregions do not directly capture commuting flows (in contrast to U.S. MSAs, or French employment zones), they are constructed according to information on integration of local economies, which is closely related to the notion of local labor markets. Our sample includes 316 microregions with more than 100,000 inhabitants in 2012.

We use municipality-level overall exports in 2012 from the COMEX Stat database (compiled

[9] See the INSEE definition at https://www.insee.fr/en/metadonnees/definition/c1361.

[10] We use exports in 2012 from https://www.trade.gov/data-visualization/metropolitan-export-map.

[11] https://data.census.gov/table.

[12] https://www2.census.gov/programs-surveys/popest/datasets/2010-2013/metro/totals/.

[13] A number of researchers have used microregions as their main unit of analysis (see Kovak, 2013; Dix-Carneiro and Kovak, 2015, 2017b, 2019; Costa, Garred, and Pessoa, 2016; Chauvin, Glaeser, Ma, and Tobio, 2017).

by the Brazilian *Ministry of Industry, Foreign Trade and Services*).[14] We complement this data source with municipal-level GDP in 2012 from IBGE.[15] We aggregate both exports and GDP from the municipality level to the level of microregions using the correspondence provided by the IBGE. Finally, we use population projections from the 2010 population Census to measure city (microregion) size.[16]

2.5 Export Intensity and Summary Statistics

We define export intensity as the share of a city's sales that are exported. Correspondingly, we define the export intensity in city i (ρ_i) as follows:

$$\rho_i = \frac{x_i}{r_i}, \tag{1}$$

where x_i and r_i denote city-level exports and revenues, respectively (i.e., aggregated across all firms operating in city i).[17] The distribution of export intensity is positively skewed for all countries in our sample, with a substantially thicker upper tail in Brazil and China than in France and the United States (see Figure A.1 in the appendix). In France, all cities in our sample have exporting firms; in contrast, about 2, 5 and 6 percent of the cities record no export activity in China, the U.S. and Brazil, respectively. While noteworthy, the presence of cities with zero exports does not affect the magnitude of our results because these cities represent a small fraction of output (0.3%, 0.6% and 0.9% of the production in China, the U.S. and Brazil, respectively).

Before turning to our empirical results, we show descriptive statistics for our sample of cities in China, France, the United States, and Brazil. Table 1 shows statistics for the distribution of population and export intensity for the four samples. Average city size varies substantially across the datasets. U.S. cities are larger on average (about 750,000 inhabitants), followed by China (710,000), Brazil (464,000), and France (193,000). These reflect the fact that in our sample, the population in the U.S. and China is more concentrated in larger cities.[18] While for the U.S. and

[14] The COMEX Stat database can be publicly accessed through an interactive interface on the website http://comexstat.mdic.gov.br. We use the version compiled by the Ministry of Economics from the information in the COMEX Stat database at the municipal level (https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas/base-de-dados-bruta).

[15] https://www.ibge.gov.br/en/statistics/economic/national-accounts/19567-gross-domestic-product-of-municipalities.html

[16] https://www.ibge.gov.br/en/statistics/social/education/18391-2010-population-census.html?=&t=microdados.

[17] Note that for China, France, and the U.S. we use total sales for the respective "city-level" areas, while we use "city-level" (Microregion) GDP for Brazil, where firm sales data are not available. Since GDP is a value-added measure, the variable for Brazil is conceptually slightly different compared to the other three countries. However, as Table A.2 shows, results are similar when we use value-added-based export intensity for China and France (where both sales and value added are available). Note also that for China, exports and sales are for manufacturing only, while they comprise all sectors for the other countries.

[18] This is consistent with evidence in Au and Henderson (2006), who show that about half of prefecture-level cities in China are smaller than their optimal size. They argue that this is most likely due to the existence of strong migration restrictions.

China, between 25 and 30 percent of the cities in our sample have populations above 500,000, in Brazil, 17 percent of the cities surpass this threshold, and in France, only 6 percent.

3 Empirical Results: Export Activity and City Size

This section presents our main empirical results. Using data from China, France, the United States, and Brazil, we show that export activity is concentrated in larger cities. We show that this pattern (i) is predominantly driven by within-industry variation, (ii) holds when using different definitions of export activity and city size, and (iii) is not driven by manufacturing alone. We then provide evidence that heterogeneity in firm productivity is an important underlying mechanism.

3.1 Baseline Results: The Stylized Fact in Four Countries

We run the following regression to establish our core stylized fact:

$$\ln(\rho_i) = \alpha + \beta \ln(pop_i) + \gamma X_i + \varepsilon_i, \tag{2}$$

where ρ_i is the export intensity of city i, as defined in (1), pop_i is city population, and X_i is a vector of several proxies for domestic and international trade costs: average distance to other domestic cities, distance to the border, distance to the coast, border city dummies, and coastal dummies.[19] Our coefficient of interest is β, reflecting how export intensity varies with city size. Table 2 presents the results: We obtain statistically highly significant estimates for all four countries, ranging from 0.18 for the U.S. to 0.34 for China and Brazil. Thus, doubling city size in China is associated with raising export intensity by about one-third. With an average export intensity of 8.8%, this corresponds to an increase in the fraction of exported (relative to total) local output by about 3 percentage points. Importantly, the coefficients remain quantitatively similar and statistically highly significant when we include our proxies for market access X_i (reported in even columns in Table 2). This provides a first check suggesting that our results are not confounded by larger cities having better access to international trade. We discuss this potential issue in more detail below. Figure 1 provides binscatter plots for our regressions, illustrating that our results are not driven by outliers.

3.2 Robustness Checks

We implement several tests to check the robustness of our main finding for China and France, where we have the most detailed data. We summarize these results here; the corresponding tables and figures are shown in Appendix A.

Alternative measure of export intensity. Our baseline export intensity measure uses city-level exports relative to sales. While this is the natural normalizing variable, we present results with an

[19] Border city dummies take on value 1 if a city's boundaries (defined in Section 2) overlap with a border segment.

alternative denominator – city population. Using this variable, we repeat our main analysis with *per capita* exports as a proxy for export intensity. Table A.1 shows that our stylized fact holds with this alternative specification: we obtain statistically highly significant coefficients for all countries, with magnitudes similar to our baseline results. In Table A.2, we show that the results are also very similar when we define export intensity relative to value added (for China and France, where these data are available). This suggests that our core stylized fact is not driven by differences in the 'upstreamness' across cities, i.e., their production of intermediate vs. final products.

Population density. Our baseline specification uses city population as the main explanatory variable. An alternative measure of agglomeration that is widely used in the literature is population density (see for instance Combes et al., 2012). This measure can deviate substantially from city size, especially when cities vary widely in the size of their geographic areas.[20] Panel A of Table A.3 shows that our results hold when we use population *density* as the explanatory variable. Finally, we also obtain similar results when using per-capita exports as dependent variable in combination with population density (see Panel B of Table A.3).

Export intensity in manufacturing vs. services. Our baseline analysis for China only considers production and export information for the manufacturing industry. The French data allow us to estimate regression (2) separately for manufacturing, services, and the primary sector. As shown in Table A.4, the gradient of export intensity with city size is remarkably similar for firms in manufacturing and in services. We also provide a consistency check, showing that export activity of the primary sector – which depends on land abundance and natural resources – is indeed if anything *less* concentrated in larger cities.

China-specific robustness checks. We perform a number of robustness checks for our results for China in the appendix. Table A.5 shows that our results are not affected by i) controlling for Special Economic Zones (SEZ) and Coastal Development Areas (CDA), ii) by using only direct exports when computing export intensity, iii) when using prefecture-level Chinese cities as the main unit of analysis (e.g., Au and Henderson, 2006). In Table A.6, we show that our results for China hold when we i) exclude state- and collective-owned enterprises, ii) exclude firms with foreign ownership, iii) use a particularly restrictive subsample that includes only privately owned domestic firms.

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 export-intensive companies with production based in small cities locate their headquarters in larger ones.[21] We address this concern

[20] See Henderson, Nigmatulina, and Kriticos (2021), who discuss the power of different measures to estimate urban agglomeration effects in the context of six Sub-Saharan African countries.

[21] This problem is not present in the U.S. data, where exports are aggregated to the city level from the establishment level. For Brazil, the data is aggregated from the tax domicile of the company, which can be at the establishment or the firm level, depending on how the firm files taxes.

in Appendix A.4. For China, we restrict the sample to firms that only have one industrial unit (Table A.7, Panel A). For France, we i) restrict the analysis to firms that are active in a single employment zone (Panel B), and ii) we assign (for firms with establishments in multiple employment zones) domestic and export revenues to establishments proportionally to their respective wage bill, which we observe at the establishment level (Panel C). Throughout all these specifications, we obtain results that are remarkably similar to our baseline.

3.3 Soil-Suitability-Driven City Population: 2SLS Results

An important concern with our empirical results is that unobserved endogenous factors may drive the relationship between export intensity and city size. In particular, larger cities may have formed in locations with better access to international markets, or they may have invested disproportionately in export-enabling infrastructure. While we used control variables for market access in Table 2, our proxies may fail to fully capture this alternative mechanism. To further address this issue, we implement a two-stage least squares (2SLS) strategy for the two main countries in our analysis – China and France. Our approach is to instrument for city size with a historical variable that is predetermined with respect to modern infrastructure: We use the caloric suitability index for agricultural crops after 1500 (CSI) from Galor and Özak (2016).[22] Our instrument reflects the fact that historically, agriculturally more productive areas could sustain higher urban population. We provide evidence for this mechanism in Table A.8 in Appendix A.5, showing that CSI has strong predictive power for historical urban population in China (using prefecture-level urban population around 1580) and France (in 1876 – the first year with sufficiently detailed data that we can match to today's employment zones). Importantly, this effect is still meaningful today: As shown in our first-stage results in Table 3, historical caloric suitability has a strong positive effect on modern urban population in both countries. Note that we allow for a differential effect of CSI in coastal areas – where fishing likely also played an important role – by including an interaction between CSI and a dummy for coastal prefectures/employment zones. In terms of magnitude, a one-standard-deviation higher historical caloric suitability predicts a 40% higher urban population today for China, and 25% for France. This effect is not significantly different in coastal areas in China, while it is muted in coastal zones in France.

Columns 3 and 4 (for China) and 7 and 8 (for France) in Table 3 show our 2SLS results, with export intensity as the dependent variable. The coefficient on city size (predicted by CSI in the first stage) is highly significant and economically important: We find that doubling suitability-driven

[22] The CSI is available for approx. 10x10 km cells, which we match to Chinese prefectures and French employment zones. Specifically, for each grid cell, we use the caloric yield for the crop that maximizes potential yield, given the set of crops that are suitable for cultivation after the Columbian Exchange. This includes the potato, which had important effects on city growth after 1500 (Nunn and Qian, 2011). We then compute, for each prefecture (China) and employment zone (France) the average attainable regional yield based on the underlying grid cells. For China, we run this analysis at the prefecture level because historical population for today's MSAs is not available. See Appendix A.5 for details. For a straightforward interpretation of coefficients, we standardize the CSI variable.

city size raises export intensity by about 60% in China and by about 40% in France in the specifications that include controls for coastal access (as a proxy for international trade costs).[23] Note that our 2SLS results are larger than their OLS counterparts (see Table 2). One explanation is that by exploiting only the variation in today's population that is driven by soil suitability, we exclude factors that historically fostered city growth via market access. Since historical trade was largely domestic, the fact that our OLS results include this channel would tend to bias them downward (i.e., higher domestic sales in cities with historically high domestic market access leading to lower export intensity). Table A.9 in the appendix presents the corresponding reduced-form estimates, documenting a strong and statistically significant relationship between CSI and export intensity.

The exclusion restriction for our 2SLS analysis is that historical soil suitability does not affect today's city-level exports via channels other than population size. We focus on a potential violation that is most relevant for our context: high soil suitability possibly also leading to better infrastructure (e.g., to transport the locally produced crops to other parts of the country). We address this possibility in Appendix A.5. We use data for France, where detailed information on Roman Roads is available, allowing us to compute the density of this important historical transportation network in each employment zone. We show in Table A.10 that CSI is not correlated with the density of Roman Roads, and it is also uncorrelated with modern highways, railways, and canals. Thus, CSI predicts city population independent of infrastructure.[24] While one might think of other possible violations of the exclusion restriction, our results render the most immediate concern – endogenous infrastructure – unlikely.[25] Overall, our 2SLS results support a direct connection between city size and export intensity that is not confounded by modern transport infrastructure.

3.4 Mechanisms

In what follows, we shed light on the mechanism that drives our novel stylized fact. We build on prominent features of existing trade models to motivate two decompositions that distinguish our new-new-trade-theory mechanism from alternative explanations.

Within- and between-industries variation. Theories of comparative advantage could explain our stylized fact through differences across industries, while new-new-trade-theory-based mechanisms

[23]For China, we use prefecture-level travel time to the nearest port from Egger, Loumeau, and Loumeau (2023). In contrast to geographic distances, travel time also captures variation in transport infrastructure across locations. Using this variable as coastal access control in column 3 is thus a conservative choice, allowing potentially endogenous infrastructure investments to explain export intensity in addition to the effect that operates through soil-suitability-driven urban population. For France, we perform additional analyses in Appendix A.5, using historical and contemporaneous infrastructure variables.

[24]In an additional consistency check, we show in Table A.11 that indeed, Roman Roads strongly predict today's highway network in France and that they also predict contemporaneous population. On the other hand, adding our instrument CSI to these regressions with Roman Roads, we confirm that CSI does not predict modern highways, while it remains a strong predictor of today's urban population.

[25]Note that possible effects of CSI on *urban productivity* (e.g., productive agriculture stimulating proto-industrialization and thus city growth) would be coherent with our mechanism, as productivity lies at its core.

would generate variation in export intensity *within* industries, across firms. To distinguish between these mechanisms, we decompose export intensity into its underlying within- and between-industry components (see Appendix A.6 for details). The between-industry component is the counterfactual measure that would result if city-level export intensity only varied due to differences in industry composition across cities.[26] The within-industry component is the residual variation that is not accounted for by differences in the sectoral composition of cities.

Table 4 shows the results of the decomposition for our main datasets, China and France. Note that by construction, the within- and between-industry coefficients add up to the overall coefficient on export intensity from Table 2. Columns 3 and 6 in Table 4 report the share of the overall variation that is accounted for by the within-industry component. For China, the relationship between export intensity and city size is exclusively driven by differences within industries, while this component accounts for 52% in France. The fact that differences across industries play a larger role in France is likely a result of the French data comprising all sectors, including many with minimal exporting, such as construction (recall that the Chinese data only include manufacturing). In fact, if we restrict the sample to manufacturing in France, the within-component accounts for 67%. Thus, within-industry differences account for the majority of our stylized fact in both countries.[27] In sum, for both countries, the majority of the variation is driven by systematic differences in exporting behavior across firms *within* the same industry.

Intensive margin and differences in transport costs. Above we implemented two checks suggesting that our core stylized fact is not driven by larger cities having systematically better international market access: We controlled for trade costs in Table 2, and we instrumented for city size using soil suitability in Table 3. Next, we provide a third, complimentary test, checking whether a standard gravity model could explain our stylized fact through differences in variable trade costs. In order to test for this possible mechanism, we decompose export intensity into i) the intensive margin x_i/r_i^x, reflecting export revenue over the total revenue of exporters; and ii) the (value) extensive margin of export intensity r_i^x/r_i, reflecting the total revenue of exporters over the total revenue of all firms:[28]

$$\ln\left(\frac{x_i}{r_i}\right) = \ln\left(\frac{x_i}{r_i^x}\right) + \ln\left(\frac{r_i^x}{r_i}\right) \tag{3}$$

[26]We use detailed, 4-digit industries in our decomposition. The decomposition cannot be performed for Brazil and the United States, as for these countries we only have access to aggregate city-level exports (i.e., not by sector).

[27]Another way to gauge the importance of the industry dimension is to run our main regression (2) at the *city-industry* level with and without industry fixed effects. Table A.12 in Appendix A.6 shows that adding industry fixed effects does not change our core result for China, while it reduces the city-size – export intensity coefficient by about one-third for France. We prefer our main specification at the city level as it avoids issues with the presence of zeros at the city-industry level and allows for the entire analysis to be conducted at the same level of aggregation.

[28]Note that our intensive margin of export intensity differs from the commonly used "average exports per exporter" definition of the intensive margin of exporting. The latter's relationship with variable trade costs is highly model-dependent (see Fernandes, Klenow, Meleshchuk, Pierola, and Rodríguez-Clare, 2023), while our measure increases as variable trade costs fall for a large number of different model assumptions.

Under a wide range of model assumptions (e.g., all gravity models), differences in variable transport costs would affect export intensity through the intensive margin.[29] In Table 5, we regress these components on city size and geographical controls. Since this is an exact decomposition, the coefficients add up to our baseline result (reported again in columns 1 and 4). The intensive margin is negatively correlated with city size in China (column 2) and uncorrelated with city size in France (column 5).[30] For both countries, we find an economically large and statistically significant correlation between city size and the importance of exporters in local economic activity (r_i^x/r_i). Most importantly, however, the negative or insignificant role of the intensive margin implies that our stylized fact is not driven by systematically lower variable trade costs in larger cities.

In sum, the results of the two decompositions suggest that our stylized fact is driven by differences across firms *within* industries that are unrelated to differences in variable trade costs. This leaves systematic differences in productivity across small and large cities (as first documented by Combes et al., 2012) as the natural candidate-mechanism behind our stylized fact.

4 Simple Model

In this section, we present a simple model to highlight the key mechanism that can explain our novel stylized fact. The model also serves as the foundation for our quantitative analysis in Section 5. It incorporates firm heterogeneity (Melitz, 2003) driven by differences in the upper tail of the local productivity distribution into an open economy Rosen-Roback framework. In the simple model, locations only differ in their exogenous firm productivity distribution, which we assume to be Pareto with location-specific shape parameters. Domestic and export market entry of firms, together with labor mobility across locations, give rise to the equilibrium distribution of economic activity, where productive locations i) host larger cities and ii) are more export-intensive.

4.1 Setup

We consider a world economy with two symmetric countries.[31] Each country is endowed with an exogenous population L of identical workers and an exogenous number of locations $i = 1, ..., I$. In each location i, a city with endogenous population L_i emerges. Workers are assumed to be perfectly mobile across locations within countries, but immobile internationally. In order to focus on the role of differences in the productivity distribution across locations, our simple model abstracts

[29] Note that for competing models without firm heterogeneity to match our main stylized facts, it is not sufficient for larger cities to have systematically better foreign market access. What is required instead is that larger cities have systematically better foreign *relative to domestic* market access than smaller cities. This, in turn, would lead to a higher intensive margin of exporting in larger cities.

[30] The negative coefficient for China can be rationalized as follows: Larger cities have larger domestic market access due to the larger local home market. This will lead c.p. to *lower* exports relative to domestic sales (i.e., lower export intensity). Thus, to account for our stylized fact, any differences in international variable trade cost would have to outweigh and overcompensate the home market effect for larger cities. Our results suggest that this is far from being the case in China, while some of the negative home market effect is absorbed in France.

[31] We therefore suppress country subscripts in the description of the model.

from within-country trade costs and differences in amenities.

Preferences. Workers are homogeneous; they decide where to locate and consume a bundle of goods and housing, while earning the local wage w_i in city i. Worker preferences are given by:

$$U_i = c_i^{\beta} h_i^{1-\beta} \tag{4}$$

where h_i denotes (individual) housing consumption in city i, and c_i is a CES composite of the consumption of tradable varieties:

$$c_i = \left[\int c_i(x)^{\frac{\sigma-1}{\sigma}} dx \right]^{\frac{\sigma}{\sigma-1}} \tag{5}$$

In each city, housing is produced by atomistic housing developers j, using land and capital according to the technology:

$$h_i(j) = k_i(j)^{1-\gamma} n_i(j)^{\gamma} \tag{6}$$

where $k_i(j)$ and $n_i(j)$ denote capital and land used by developer j at location i. Both land and housing markets are assumed to be perfectly competitive at the local level. Capital is provided in perfectly elastic supply at the exogenous rental rate $r_K = 1$, while each location i has the same land endowment N. Land market clearing implies $\int_j n_i(j) = N$. Total housing supply at the city level is then given by

$$H_i = K_i^{1-\gamma} N^{\gamma} \tag{7}$$

where $K_i = \int_j k_i(j)$. Land and capital income are assumed to be fully taxed by local governments and redistributed lump-sum to local residents as in Helpman et al. (1995).

Production. In each country and city, there is an unlimited supply of potential entrants. Each firm produces a differentiated tradable variety using local labor. Once they have chosen a city to enter and paid the entry cost F_e in units of local labor, firms draw their productivity ψ from the location-specific productivity distribution $g_i(\psi)$. In each location, productivity is distributed Pareto with the corresponding cumulative distribution function:[32]

$$G_i(\psi) = 1 - \left(\frac{A_i}{\psi} \right)^{\alpha_i}, \quad \text{for } \psi \geq A_i \tag{8}$$

where α_i is the location-specific productivity shape parameter, and A_i is the scale parameter. In our simple model, we use a common scale parameter $A_i \equiv 1$ for all locations, which allows us to

[32] The Pareto distribution combines analytical tractability with a good approximation of the right tail of the firm productivity distribution. This is the quantitatively relevant part of the distribution, especially for exporting – see Axtell (2001) for evidence from the US, and Eaton, Kortum, and Kramarz (2011) for France.

focus on the role of the shape parameter.[33] Thus, the probability that a productivity draw exceeds a value $x \geq 1$ is given by $\Pr(\psi > x) = x^{-\alpha_i}$. A firm with productivity ψ that uses local labor input l produces the following quantity of output:

$$q = \psi l \tag{9}$$

Depending on their productivity, firms choose to exit immediately, serve the domestic market paying fixed cost F_d (in units of local labor), or serve both the domestic and the export market (paying an additional fixed cost F_x to export). All fixed costs as well as the variable costs of exporting (τ) are identical across locations. We adopt the standard simplifying assumption of zero domestic transport costs.[34]

Firms engage in monopolistic competition. Given CES demand, this implies that prices are set at a constant mark-up over marginal cost. Because firms can serve all domestic locations without incurring variable trade costs, and because imported varieties also reach each location at the same international trade cost, the price index of the tradable good is the same across locations. We choose the CES composite tradable good as the numeraire ($P = 1$).

For a firm with productivity draw ψ, operational profits from serving the domestic and foreign market, respectively, are given by:

$$\pi_i^d(\psi) = \frac{r_i^d(\psi)}{\sigma} - w_i F_d = \frac{R}{\sigma}\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} - w_i F_d \tag{10}$$

$$\pi_i^x(\psi) = \frac{r_i^x(\psi)}{\sigma} - w_i F_x = \tau^{1-\sigma}\frac{R}{\sigma}\left(\frac{\psi\rho}{w_i}\right)^{\sigma-1} - w_i F_x \tag{11}$$

where $r_i^d(\psi)$ and $r_i^x(\psi)$ denote the revenue from serving the domestic and foreign market, R is the aggregate revenue of the tradable sector in each country, w_i is the wage in location i, and $\rho \equiv \frac{\sigma-1}{\sigma}$.

As in Melitz (2003), the industry equilibrium (in our model, for each location) can be characterized using two zero-profit cutoff conditions (12) that yield expressions for the minimum productivity threshold for successful entry into the domestic market (ψ_i^{d*}) and for successful entry into the foreign market (ψ_i^{x*}), as well as a free entry condition (13). The latter states that the expected operational profits of market entry $\bar{\pi}_i$ are equal to the fixed cost of market entry (in terms of local

[33] Note that a smaller α_i implies not only a thicker upper tail but also higher average productivity in location i. In our quantitative model in Section 5, we allow A_i to vary across locations. This allows us to counter-balance the effect of α_i on average productivity.

[34] We impose the standard parametric restrictions $\tau^{\sigma-1}F_x > F_d$, which ensures that firms serving the foreign market represent a strict subset of firms operating domestically, and $\alpha_i > \sigma - 1$, which ensures that average firm size is finite.

labor).

$$\pi_i^d(\psi_i^{d*}) = 0 \qquad \text{and} \qquad \pi_i^x(\psi_i^{x*}) = 0 \tag{12}$$

$$\bar{\pi}_i = \int_{\psi_i^{d*}}^{\infty} \pi_i^d(\psi) g_i(\psi) d\psi + \int_{\psi_i^{x*}}^{\infty} \pi_i^x(\psi) g_i(\psi) d\psi = w_i F_e \tag{13}$$

We use (13) to solve for the measure of entrants (M_i^e) in each city:[35]

$$M_i^e = \frac{L_i}{\sigma \left[F_e + \psi_i^{d*-\alpha_i} F_d + \psi_i^{x*-\alpha_i} F_x\right]} \tag{14}$$

Thick productivity upper tail. A key ingredient of our model is the thicker upper tail (i.e., smaller shape parameter α_i) of the productivity distribution $g_i(\psi)$ in larger cities. This assumption reflects the empirical pattern first documented by Combes et al. (2012). We show in Section 5.2 that the same pattern holds in our data. For simplicity, we take the differences in the upper tail across cities as given in this simple version of the model. In Appendix C, we provide a model extension that additionally features agglomeration economies in the upper tail of the local firm productivity distribution and a microfoundation based on introducing learning from local experts into a stochastic firm growth process à la Gabaix (2009).[36] We adopt the core of this mechanism in an extension of our quantitative model that allows for agglomeration economies, such that the local shape parameter α_i depends on city size L_i (see Section 5.5).

Residential choice in spatial equilibrium. In spatial equilibrium, utility is equal across all locations within each country and given by:

$$U_i = \beta^\beta (1-\beta)^{1-\beta} \frac{v_i}{(p_i^h)^{1-\beta}} = \bar{U} \qquad \forall i \tag{15}$$

where v_i denotes total income per worker, and p_i^h is the price of housing at location i. Since we assume that capital and land income are fully taxed by local governments and redistributed lump

[35] M_i^e represents the measure of firms that pay the fixed cost of entry F_e at each location. This measure includes both active firms (that serve at least one market) and inactive firms (i.e., firms with 0 total revenues) in each city.

[36] Learning has been widely recognized as a fundamental force of agglomeration benefits (e.g., Glaeser, Kallal, Scheinkman, and Shleifer, 1992; Duranton and Puga, 2004; Rosenthal and Strange, 2004). While this previous literature has focused on average productivity in a location, we highlight the effect of learning on the *shape* of the productivity distribution. In our microfoundation, firms choose how much time to invest in order to meet and learn from local experts with technology or business expertise. The marginal cost of those meetings decreases with city size, leading to higher learning in larger cities. We build on the work by Gabaix (2009) and model a stochastic firm growth process whose drift component depends on local learning effects. Using the Fokker-Planck equation, we show that this process yields a stable Pareto distribution that features a thicker upper tail in larger cities. We then take this stable Pareto distribution as an input for our static spatial equilibrium model. All proofs from this section also hold in the model with agglomeration economies – as long as these are not too strong – and are available from the authors.

sum to workers, total income in location i is given by labor income plus expenditure on housing:

$$v_i L_i = w_i L_i + (1-\beta) v_i L_i = \frac{w_i L_i}{\beta} \tag{16}$$

The price of housing follows from (7) and is given by:

$$p_i^h = \left(\frac{r_K}{1-\gamma}\right)^{1-\gamma} \left(\frac{r_i}{\gamma}\right)^{\gamma} = \left(\frac{1}{1-\gamma}\right)^{1-\gamma} \left(\frac{r_i}{\gamma}\right)^{\gamma} \tag{17}$$

where r_i denotes the (endogenous) land rental rate at location i. Finally, land market clearing yields the equilibrium land rent:

$$r_i = \frac{1-\beta}{\beta} \gamma \frac{w_i L_i}{N} \tag{18}$$

General equilibrium. The general equilibrium of the model is represented by, for each location, a measure of workers L_i, a set of entry thresholds for the domestic and the export markets, ψ_i^{d*} and ψ_i^{x*}, respectively, a set of wages w_i, as well as the level of utility in the economy $\bar{U}$ that satisfy the following system of equations:

1. The industry equilibrium consisting of the zero profit cut-off conditions and the free entry condition (equations 12 and 13)

2. Trade balance at each location (wage income in each location equals expenditure on goods produced in that location)

$$w_i L_i = M_i^e \left(\int_{\psi_i^{d*}} r_i^d(\psi) g_i(\psi) d\psi + \int_{\psi_i^{x*}} r_i^x(\psi) g_i(\psi) d\psi \right) \qquad \forall i \tag{19}$$

3. Spatial equilibrium for workers among all domestic locations

$$U_i = \bar{U} \qquad \forall i \tag{20}$$

4. National labour markets clear

$$\sum_i L_i = L \tag{21}$$

4.2 Matching the Stylized Fact

We now show that our simple model can match the novel stylized fact that we documented in Section 3. We begin with an auxiliary result that will facilitate our exposition.

Lemma 1. *Cities whose exogenous productivity distributions feature a thicker upper tail (i.e., smaller shape parameters α_i) will have higher average firm productivity as well as higher wages, house prices, and population in equilibrium.*

Proof: See Appendix D.1.

Cities with a thicker upper tail feature more firms with high productivity draws. This ex-ante difference is reinforced by the endogenous selection into entry, since cities with more productive firms feature stronger labor demand for a given wage and hence stronger selection effects. We will therefore refer to cities with thicker productivity upper tails simply as "more productive" cities. The presence of highly productive firms implies higher local labor demand, so that these locations feature higher equilibrium population, wages, and house prices.[37] We now proceed to stating our first theoretical result, connecting city size to export intensity

Proposition 1. *In equilibrium:*

(i) Larger cities are more export intensive.

(ii) The higher export intensity 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.2.

This proposition provides our central prediction: More productive – and therefore larger – cities also have a higher export intensity. This is in contrast to standard models that only feature differences in the *level* of productivity across locations (e.g., Redding, 2016): As in Melitz (2003), the productivity thresholds for serving the domestic and the export market are proportional to each other in our model, and the ratio is common to all locations:

$$\frac{\psi_i^{x*}}{\psi_i^{d*}} = \tau \left(\frac{F_x}{F_d}\right)^{\frac{1}{\sigma-1}} \qquad \forall i \tag{22}$$

Thus, standard models that build on differences in *average* productivity across locations do not feature spatial heterogeneity in export intensity. In our setting, the thicker upper tail of the productivity distribution in more productive cities means that for a given ratio in (22), a higher share of firms are above the export threshold. This, in turn, leads to more productive locations having a higher export intensity. Note that this is driven exclusively by the export share (i.e., the fraction of firms that export), with the intensive margin of exporting (i.e., the export intensity of exporters) being constant across space.[38]

[37] While our main focus (and the distinguishing feature of our model) is the relationship between city size and export intensity, we also provide empirical evidence for a strong positive correlation between city population, wages, and house prices in Appendix A.7.

[38] This feature of the model is a direct result of the fact that under CES preferences (as in Melitz, 2003), conditional on exporting, firm-level export intensity is constant with respect to firm productivity.

Taken together, our simple model can account for our novel stylized fact and for the result from the decomposition in Section 3.4, showing that it is driven by the extensive margin of exporting.

4.3 Comparative Statics

Our model provides novel insights into the joint determination of international trade and economic geography outcomes. In this section, we highlight these interactions by studying the impact of trade and spatial policies on several variables of interest.

Implications of international trade liberalization for internal geography. In our simplified setting, trade liberalization leads to a spatial reallocation of employment:

Proposition 2. *A reduction in the variable international trade costs τ leads to a shift in population towards larger cities, as well as an increase in aggregate productivity and welfare.*[39]

Proof: See Appendix D.3.

Trade liberalization leads to two types of reallocation of economic activity: First, within each location, economic activity reallocates from less to more productive firms. This raises the location-specific average productivity and puts upward pressure on wages. Second, this effect is stronger in larger cities due to their higher initial export intensity. Thus, there is a reallocation of workers to larger cities. Both of these channels increase aggregate productivity and welfare.

Trade costs and inequality. Since we focus on equilibrium outcomes with mobile, homogeneous workers, there are no distributional effects of trade in terms of utility. However, many studies on the distributional effects of trade examine wage inequality. We thus highlight how our mechanism naturally gives rise to both regional nominal wage inequality (i.e., wage inequality across cities) and aggregate wage inequality among workers (e.g., the nation-wide Gini coefficient). These results highlight the importance of accounting for geography and spatial equilibrium when studying the effects of trade on inequality. Formally:

Proposition 3. *A reduction in the variable international trade costs τ leads to*

(i) an increase in nominal wage inequality across cities, with nominal wages in larger cities increasing relative to smaller cities. However, trade liberalization does not lead to welfare inequality across cities.

(ii) a change in nominal wage inequality at the country level. The direction of this effect is ambiguous and depends on the interplay of the wage changes from (i) and population reallocations. However, trade liberalization does not lead to a change in aggregate welfare inequality.

[39]This proposition refers to the long run, with perfectly mobile labor. If labor is immobile, a reduction in τ will instead lead to differences in welfare across space, with workers in small cities achieving a lower level of utility than workers in large cities.

Proof: See Appendix D.4.

We already established that trade liberalization disproportionately increases the demand for labor in larger and more productive cities. This not only leads to disproportionate population growth in more productive cities (Proposition 2) but also pushes up their wages, as they need to attract workers while compensating for rising house prices. Additionally, less productive cities lose population, and housing becomes cheaper. Thus, wages in these places decline relative to larger cities. Consequently, trade liberalization increases the pre-existing nominal wage gap between small and large cities. However, due to perfect worker mobility, this rise in nominal wage inequality does not imply inequality in *utility* in spatial equilibrium. This is similar in spirit to Moretti (2013) and Diamond (2016), highlighting the importance of getting the inequality measure right.

Aggregate nominal wage inequality in the model depends on differences in nominal wages across cities and on the distribution of population across cities. Trade liberalization increases nominal wage differences across cities as outlined above (Proposition 3, i) and also changes the distribution of population across space (Proposition 2). Since the effect of the latter on aggregate wage inequality is ambiguous, we cannot sign this effect and will explore it more thoroughly below in the quantitative model.

Spatial policy and trade. The proposition below characterizes the implications of an increase in housing supply elasticities for international trade:

Proposition 4. *Planning policies that increase the housing supply elasticity lead to an increase in country-level export intensity and aggregate productivity.*

Proof: See Appendix D.5.

Intuitively, relaxing housing supply restrictions reduces the cost of housing in all locations, thus making all cities more attractive. However, this effect is stronger for larger and more productive cities, as these locations were more constrained by the scarcity of land and housing. As a result of this mechanism, workers move towards the most productive cities. Thus, the relative mass of firms shifts to the more productive locations, where firms are also more likely to be exporters. This process causes an increase in aggregate productivity and in aggregate export intensity.

Local and aggregate trade elasticity. Our simple model makes two novel predictions regarding a key object of interest in the international trade literature: the trade elasticity. First, the model predicts that city-level international trade elasticities (i.e., the elasticity of city-level exports with respect to the variable foreign trade cost τ) are heterogeneous across cities and systematically *lower* in larger cities. Second, the aggregate (or country-level) trade elasticity depends on the endogenous distribution of exporters across space. To derive these predictions formally, we make the simplifying assumption that each city is small relative to the aggregate economy:

Proposition 5. *Assume each city is small relative to the national economy and that for each city, foreign demand is small relative to domestic demand (i.e., τ and/or F_x are large). Then:*

(i) each city's (international) trade elasticity is given by the shape parameter of its firm productivity distribution.

(ii) the city-level (international) trade elasticity is decreasing with city size. These differences are driven entirely by the variation in the number of exporters.

(iii) the country-level trade elasticity is endogenous and given by an export revenue-weighted average of city-level trade elasticities.

Proof: See Appendix D.6.

Under these assumptions, the response of exports to changes in trade costs in location i – the city-level trade elasticity – is equal to α_i, i.e., the shape parameter of the local productivity distribution. This result extends the country-level analysis from Chaney (2008) to our setting. Since in equilibrium larger cities have a lower α_i (i.e., a thicker upper tail, see Lemma 1), the trade elasticity is *decreasing* with city size.[40] Intuitively, the thicker upper tail in larger cities implies a relatively higher mass of firms above the export cut-off. Thus, the mass of firms that enter the export market following a reduction in variable trade costs is lower relative to the mass of firms that are already exporting. Decomposing the growth in exports as τ falls, it is entirely driven by changes in the number of exporters (i.e., the extensive margin), while the average exports per exporter (the intensive margin) remains unchanged. This is because in each city, the increased exports of existing exporters are exactly offset by the relatively lower exports of the new entrants.

The country-level trade elasticity can be decomposed as the export value-weighted average of city-level trade elasticities. Since these city-level elasticities are heterogeneous across cities and the distribution of population and economic activity across cities depends on international trade costs, the country-level trade elasticity is not constant and depends on the overall level of international trade costs.

4.4 Model Validation: City Size and Trade Elasticity

Proposition 5 provides a novel prediction that arises naturally from our key assumptions. In contrast, the trade elasticity does not vary by city size in workhorse spatial models without differences in the upper tail of the firm productivity distribution across locations (e.g., Allen and Arkolakis, 2014; Redding, 2016). Thus, we can use Proposition 5 to validate our model.

To empirically test whether the trade elasticity is lower in larger cities, we need exogenous variation in the trade costs of exporters. We exploit the granting of permanent normal trade relations

[40] In line with the large literature on this subject, we focus on the elasticity with respect to a change in *variable* trade costs (τ_x). However, as we show in Appendix D.6, our model predicts the same qualitative result for the elasticity with respect to the *fixed* cost of exporting (f_x).

(PNTR) for Chinese exports to the US in 2001. As pointed out by Pierce and Schott (2016), the conferral of PNTR removed uncertainty regarding future increases in US duties on imports from China and thereby promoted Chinese exports. Building on the identification strategy in Pierce and Schott (2016), we run the following regression:

$$\begin{aligned}\ln(\text{exports}_{ijt}) =& \beta_1 \left[PostPNTR_t \times NTR\,Gap_j\right] + \beta_2 \left[PostPNTR_t \times NTR\,Gap_j \times \ln(pop_i)\right] + \\ & \delta \mathbf{X}_{ijt} + \gamma_j + \gamma_t + [\gamma_i] + \varepsilon_{ijt} \end{aligned} \tag{23}$$

where the dependent variable $\ln(exports_{ijt})$ denotes the exports of industry j in prefecture i in year t. As we mentioned in Section 2, we use the Annual Survey of Manufacturing (ASM) for this analysis because the Census is not available at a sufficiently high frequency. Since the ASM is at the prefecture (and not at the MSA) level, pop_i is now the urban population in prefecture i in the year 2000. $PostPNTR_t$ is a dummy that takes on value one for $t \geq 2001$, and $NTRGap_j$ is the difference (for industry j) between the US most favored nation (MFN) tariffs and non-MFN ("column 2") tariff rates. The former became permanent under PNTR. Thus, $NTRGap_j$ measures the counterfactual increase in trade costs that would have occurred if China's normal trade relations (NTR) status had not been extended. We calculate this NTR gap at the 3-digit Chinese industry level using data from Pierce and Schott (2016).[41] Importantly for our identification, $NTRGap_j$ exhibits substantial variation across industries, with mean and standard deviation of 30% and 14%, respectively. We standardize $NTRGap_j$ and $\ln(pop_i)$ for a straightforward interpretation of coefficients.

We run regression (23) for the years 1998, 2000, 2003, 2006, and 2007. The start year is determined by data availability. We stop in 2007 in order to avoid the negative export demand shock induced by the Great Recession. The vector $\mathbf{X}_{ijt}$ includes city-level controls (city population, distance to the coast, distance to the border, and capital-labor ratio, all in logs), as well as each of these variables' interactions with $PostPNTR_t$, $NTRGap_j$, and with $PostPNTR_t \times NTR\,Gap_j$ (this is why $\mathbf{X}_{ijt}$ has subscript $_{ijt}$). Finally, γ_j, γ_t, and γ_i are industry, year, and city fixed effects, respectively. We report regressions with and without city fixed effects. Note that these absorb the city-level controls but not their interactions with $PostPNTR_t$ and $NTRGap_j$. We weigh each city-industry cell by its share in overall exports of the industry, and we cluster standard errors at the industry as well as at the city level.

Holding everything else equal, we expect a larger response of exports post-PNTR in sectors with higher $NTRGap_j$ – that is, we expect a positive coefficient β_1. However, our main coefficient of interest in testing Proposition 5 is β_2, which measures how the response of city-industry-level

[41] We match the original HS6 products to 3-digit Chinese industries using the crosswalk from Dean and Lovely (2010). Overall, this yields NTR gap data for 112 industries. Non-MFN tariffs were on average 34% across the 112 Chinese industries in 1999, while MFN tariffs were merely 4%.

exports to PNTR varies with city size (pop_i). We expect a negative coefficient ($\beta_2 < 0$), corresponding to a declining trade elasticity with city size. To provide an even more stringent test of our model we also separately look at the extensive (number of exporters) and intensive margin response (average exports of exporters), where the model predicts $\beta_1 > 0, \beta_2 < 0$ for the extensive margin and $\beta_1 = \beta_2 = 0$ for the intensive margin (see Appendix D.6).

Table 6 presents our results. We begin with a parsimonious specification in column 1 that controls only for the population terms as well as industry and year fixed effects. We then introduce city-level controls $\mathbf{X}_{ijt}$ (col 2), add the interactions of city controls with $PostPNTR_t$ and $NTRGap_j$ (cols 3), and include city fixed effects. We find strong support for Proposition 5 throughout, with $\beta_1 > 0$ and $\beta_2 < 0$, both statistically highly significant. According to the estimates for β_1 in Table 6, a one-standard-deviation (std) increase in $NTRGap_j$ led to a 16-18 percent increase in exports post-PNTR (for a city with average population size). This serves as a validation of our approach, as the reduction in uncertainty about U.S. import tariffs indeed led to higher exports from China. Most importantly, this effect is weaker in larger cities, as shown by the statistically significant, negative coefficients on the interaction (β_2): For small cities (in the bottom decile of population), the net effect of a one-std higher $NTRGap_j$ is a 40% increase in exports post-PNTR, while it is only 8% for large cities (in the top decile). Thus, the reduction in trade elasticities in larger cities is not only statistically but also economically significant.

Columns 5 and 6 decompose our results into the intensive and the extensive margin. In line with Proposition 5, we find β_1 and β_2 close to zero and statistically insignificant for the intensive margin, while for the extensive margin, the log(number of exporters) increases significantly ($\beta_1 > 0$), but less so in larger cities ($\beta_2 < 0$). These results validate our core mechanism of cross-city differences in the tail of firm productivity distributions – the distinguishing feature of our model that gives rise to Proposition 5.

A possible concern with the PNTR shock is that it led to large inflows of foreign (especially U.S.) firms investing in China for export purposes (Pierce and Schott, 2016). We address this in Appendix A.9, showing that our results hold when we reproduce Table 6 excluding foreign firms.

Finally, we can also use the PNTR shock to provide empirical evidence for Proposition 2's prediction that population will move towards larger cities. In Appendix A.8, we show that exposure to the PNTR shock led to significantly larger in-migration and employment growth in larger relative to smaller cities in China. We also explore Proposition 3's prediction on nominal wage inequality across cities. However, we find that city-level wages react much less to the PNTR shock (which, as we show in Appendix A.8, is likely driven by compositional effects due to elastic unskilled migration). Thus, we delegate the analysis of Proposition 3 to our quantitative model in Section 5, where we can isolate this mechanism from other confounding forces.

5 Quantitative Analysis

In this section, we embed the mechanism we highlighted in our simple model in an open-economy quantitative spatial model, which we then fit to Chinese and French microdata. We evaluate its key predictions quantitatively, studying the relationship between city size and i) export intensity and ii) the local trade elasticity. We also study the effects of counterfactual trade and spatial policies on welfare, productivity, wage inequality, spatial concentration and the aggregate trade elasticity. Finally, we compare our results to more standard models studied in spatial economics and international trade.

5.1 Quantitative Model

In order to make our framework amenable to quantitative analysis, we adapt the simple model presented in the previous section in four ways: First, we model the export market as one potentially asymmetric country (subscript c) without internal geography, while the home country consists of N locations. Second, we allow for a full matrix of bilateral domestic iceberg trade costs across the N locations: d_{ni} units of a commodity need to be shipped from origin i for one unit of the commodity to arrive at destination n.[42] We assume that exporting happens through specific locations that are located at the border or feature large ports in the data. Specifically, the cost of exporting from location i is equal to $d_{ci} = d_{pi} \times \tau$, where d_{pi} is the distance to the closest point of exit (ports or border crossings) and τ is the cost of exporting. Overall, a city's export intensity will be the larger (c.p.) the lower its export cost is *relative to* its domestic trade costs. Third, we allow for a richer specification of productivity differences across locations. While we maintain the assumption that firm productivity is drawn from location-specific Pareto distributions given by (8), we allow these distributions to differ across locations not only in their shape parameters (α_i), but also in their scale parameters (A_i). A higher *scale* parameter A_i increases average city productivity while having no effect on export intensity. As we explained in our simple model, export intensity is solely driven by the *shape* parameter α_i, which *ceteris paribus* not only renders a thicker upper tail but also higher average productivity. The quantitative model effectively allows A_i to counteract this 'side effect' of α_i on average productivity. Fourth, we introduce amenities in the model that vary across locations, and we allow workers to have heterogeneous preferences for different locations, which renders local labor supply imperfectly elastic and allows for differences across locations in real wage as well as ex-post utility.

[42] Introducing domestic trade costs with $d_{ni} > 1$ and $d_{ii} = 1$ generates a standard home market effect, where larger cities (c.p.) have higher domestic market access. This is because their local market and therefore their overall domestic market is larger. This creates a force that leads to a *negative* correlation between city size and export intensity, thus working against our core stylized fact.

Preferences. Worker preferences are given by:

$$U_i(\omega) = b_i(\omega)c_i(\omega)^{\beta} h_i(\omega)^{1-\beta}, \tag{24}$$

with c_i and h_i as defined above, and $b_i(\omega)$ denoting location-worker-specific amenity shocks that are drawn independently from the Fréchet distribution $F_i(b) = \exp\left(-B_i b^{-\epsilon}\right)$. The scale parameters B_i determine average amenity levels for each location i, and the shape parameter ϵ controls the dispersion of amenities across workers for each location. With this specification of preferences, the indirect utility function at each location is given by:

$$U_i(\omega) = \frac{b_i(\omega)v_i}{P_i^{\beta}\left(p_i^h\right)^{1-\beta}}, \tag{25}$$

where P_i denotes the price index, which is now location-specific due to within-country trade costs. As in our simple model, v_i denotes total income per worker, and p_i^h is the price of housing at location i.

Given the extreme-value assumption on the amenity draws, utility is distributed Fréchet. Within the home country, the population share of each location is given by

$$\frac{L_i}{L} = \frac{B_i(v_i/P_i^{\beta}(p_i^h)^{1-\beta})^{\epsilon}}{\sum_{k\in N} B_k(v_k/P_k^{\beta}(p_k^h)^{1-\beta})^{\epsilon}} \tag{26}$$

The expected utility at each location is given by

$$\bar{U} = \delta\left[\sum_{k\in N} B_k(v_k/P_k^{\beta}(p_k^h)^{1-\beta})^{\epsilon}\right]^{\frac{1}{\epsilon}} \tag{27}$$

where $\delta \equiv \Gamma\left((\epsilon-1)/\epsilon\right)$, with $\Gamma(\cdot)$ denoting the Gamma function.

General equilibrium of the quantitative model. The general equilibrium is represented by a measure of workers for each location L_i together with the following sets for each location: domestic market entry thresholds for firms ψ_i^{d*}, entry-into-exporting thresholds ψ_i^{x*}, firm entrants M_i^e, wages w_i, price indices P_i, and land rents r_i, as well as their corresponding variables for the rest of the world $w_c, \psi_c^{d*}, \psi_c^{x*}, M_c^e, P_c, r_c$. The equilibrium is determined by the following conditions in each location: trade balance, zero profit condition for serving the local and the export market, firm free entry condition, price index equation, land market clearing, location choice probabilities, and labor market clearing (see Appendix E.1 for the system of equations).

Welfare gains from trade in the quantitative model. We derive a tractable expression for welfare changes induced by variations in trade costs, in the tradition of the Arkolakis, Costinot, and

Rodríguez-Clare (2012) "ACR" formula. This expression establishes a direct link between the welfare gains from trade in our framework and those obtained in benchmark models such as Melitz (2003) and Redding (2016). The derivation and a detailed discussion of this welfare result are provided in Appendix E.2.

5.2 Calibration

Pareto shape parameter. 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, Mayer, and Thoenig (2014), estimating QQ regressions for each location using firm-level domestic revenue data and accounting for industry fixed effects. We then multiply the resulting coefficient by $(\sigma-1)$ following di Giovanni, Levchenko, and Mejean (2014) to recover the underlying productivity distribution Pareto shape parameter. In a nutshell, the QQ method estimates the domestic sales Pareto distribution parameters, minimizing the distance between the empirical and theoretical quantiles.

Table A.18 shows the summary statistics 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). Importantly, as Figure 2 illustrates, the estimated Pareto shape parameters show a strong negative, statistically highly significant correlation with city size in both countries, with estimated coefficients at -0.18 for China and -0.21 for France. In other words, large cities feature productivity distributions with thicker upper tails (i.e., smaller shape parameters). Note that these estimates stem directly from the data, without imposing any structure from our model. Thus, the strong negative correlation in Figure 2 directly supports our central assumption.

Other parameters. The rest of the calibration extends Redding (2016) for the case of heterogeneous firms with fixed production and export costs. We relegate the technical details to Appendix E.3. In short, the procedure inverts the model using the values of externally calibrated parameters together with data on population (L_i), wages (w_i), and country-level exports to recover the locational fundamentals: amenities (B_i), the Pareto *scale* parameter (A_i), and the value for the variable trade cost between ports and the rest of the world, τ.[43]

5.3 Fitting the Stylized Fact: Export Intensity and City Size

In this section, we evaluate the fit of the quantitative model to our key stylized fact: the higher export intensity of larger cities. Note that the relationship between export intensity and city size was not a targeted moment in our calibration. Thus, matching this empirical moment is a stringent test for our model and its underlying mechanisms. Figure 3 plots binned scatter diagrams for city-level export intensity and city size for China and France, respectively. Each dot in the graph

[43] Table A.17 summarizes the values of the main parameters and the corresponding data moments. We display the correlation of the resulting fundamentals (A_i, B_i) with city size in Figure A.2: As we would expect, the level of amenities and the productivity *scale* parameter correlate positively with city size.

corresponds to approximately 10 cities in China and 5 cities in France.[44] The model produces a good fit: It can account for more than two-thirds of the correlations observed in the data.

To further evaluate the performance of our model, we turn off specific features one by one and examine how that affects its ability to match the data. This allows us to shed light on the role played by the various mechanisms in matching the empirical relationship between export intensity and city size. Table 7 presents the correlation between export intensity and city size in the data (row 1 in each panel) and in the different versions of our model (rows 2-5). We also report correlations for the model-based decompositions into intensive and extensive margin (in columns 2 and 3), similar to the empirical counterpart in Table 5. We report all results including the geographic controls used in the empirical analysis (but the results are very similar when excluding these controls).

We begin with our baseline model (row 2). For both China and France, most of the correlation between export intensity and city size is driven by the importance of exporters (extensive margin) rather than the exporting intensity of exporters (intensive margin). This broadly reflects the patterns in the data: For France, the model matches both margins very well; for China, there is a negative correlation between city size and the intensive margin in the data, while the model predicts a modest positive relationship. Next, we compare the fit of our baseline model to the fit of two polar opposite benchmarks. The first benchmark ("Homogeneous firms model") features *homogeneous* firms within cities, but we allow for variation in productivity levels across cities.[45] This benchmark is similar to the increasing returns model in Redding (2016), allowing us to check whether differences in foreign- relative to domestic-market access can produce the observed relationship between export intensity and city size. The results (row 3 of Table 7) show that without firm heterogeneity, the model can only account for a small fraction of our stylized fact: at most 6% (0.016/0.258) in China and less than 4% in France (0.007/0.185). In addition, the small correlation between export intensity and city size is entirely due to the intensive margin, which in turn is driven by differences in the calibrated international (relative to domestic) market access across cities). This was to be expected because in the absence of within-city firm heterogeneity, all firms engage in export markets in the same way, and there is no selection-into-exporting (extensive-margin) mechanism. These results suggest that differential market access can, at most, play a very modest role in accounting for the observed relationship between export intensity and city size.

Our second benchmark ("Baseline without internal trade costs") is a model in which the role of

[44] To make model and data directly comparable in the figure, we i) normalize the average export intensity across cities to zero both in the data and in the model (our model matches the aggregate sales-weighted export intensity, leading to slight differences in the average level as compared to the data), and ii) we control for the same set of geographic characteristics included in Table 2 when plotting the data.

[45] We use the same calibration strategy as in the baseline model. Since this version has no firm heterogeneity, we do not need to fit the shape parameter but set it equal to infinity, which implies a degenerate productivity distribution for all cities. We keep the structure of the internal variable trade costs between domestic locations the same as in our baseline model.

market access is shut down by assuming costless within-country trade, while city-level productivity distributions are kept the same as in our baseline model. The results in row 4 of Table 7 show that the remaining core mechanism – heterogeneity in firm-level productivity distributions across cities – can account for about 50 percent of the relationship between export intensity and city size in China, and for more than 95 percent in France.[46] Finally, we explore the role of cross-city heterogeneity in the thickness of the upper tail of the firm productivity distributions – the key component in our theory. We introduce a third benchmark ("Baseline with common shape parameter"): a model featuring heterogeneity in market access across cities, within-city firm-level heterogeneity, but *no* differences in the thickness of the upper tail (we allow *scale* parameters to differ across cities but impose a common *shape* parameter). As shown in row 5 of Table 7, this setup can explain only a small fraction of the correlation between city size and export intensity. In addition, this is almost entirely driven by the intensive margin of trade, thus contradicting the pattern in the data.

Taken together, our results imply that differences in the thickness of the productivity upper tail are crucial to match our key stylized fact. This novel mechanism, in isolation, allows us to account for most of the reduced-form relationship between export intensity and city size, while differences in foreign- relative to domestic-market access play only a minor role.

5.4 Validating the Model: City Size and Local Trade Elasticity

Our simple model in Section 4 predicts a lower trade elasticity in larger cities. To test whether this prediction holds under the more general assumptions of our quantitative model, and to assess its quantitative importance, we calculate the trade elasticity for each city from a marginal decrease in trade costs at the observed equilibrium. Figure 4 plots the local trade elasticity from our full quantitative model against city size. For China, the estimated gradient is 0.137 (0.164 for France). Thus, our quantitative model confirms that the trade elasticity varies with city size in an economically meaningful way. Note that these estimates provide another untargeted moment. It is thus reassuring that they are similar to our PNTR regression results from Section 4.4, where we found a gradient of about 0.10 for China.[47]

Finally, in Appendix E.5 we show that the relationship between city size and trade elasticities is a distinguishing feature of our model, as compared to the other models discussed in Section 5.3: The city-size-elasticity relationship results from our baseline quantitative model (even in a simpler version without internal trade costs). Crucially, the relationship becomes minuscule (and

[46]Note that in this model, all effects operate via the extensive margin, as symmetric market access produces a null intensive margin: conditional on exporting, firms in all locations export the same fraction of their production.

[47]The city-size-gradient coefficients (i.e., the triple interactions) in Table 6 are negative, reflecting a smaller reaction of exports to *falling* trade costs ($NTRGap$) in larger cities (i.e., a smaller trade elasticity in absolute terms). In the model, the size- gradient-coefficients are positive, reflecting a less negative trade elasticity (i.e, also smaller in absolute terms).

reverses its sign) when we use a common Pareto shape parameter across cities, or when using a homogeneous-firms model.

5.5 Counterfactual Analysis

This section analyzes the effects of commonly discussed trade and spatial policies in the context of our model. We illustrate how the aggregate productivity and welfare implications of these policies differ relative to the predictions of existing workhorse models.

Trade Liberalization

We first study the effects of trade openness by comparing the actual trade equilibrium to an autarky counterfactual. This counterfactual is identical to the actual equilibrium along all key parameters, but international trade costs are assumed to be prohibitive. To contrast our model's counterfactual predictions, we compare the quantitative effects of trade in this model to four benchmarks commonly considered in the literature, as well as to a version of our model with endogenous shape parameters. We separately calibrate each model to the data (see Appendix F for details). Table 8 shows the results for: (1) Our baseline model; (2) A model with the same rich internal geography as in our baseline but with homogeneous firms;[48] (3) A model equal to the baseline model except that it does not feature any domestic trade costs and, therefore, no differences in the cost of exporting across locations; (4) A heterogeneous firm model with geography that only features differences in the level of productivity but not in the tail of the productivity distribution across cities (i.e., we restrict the shape parameter of the Pareto distribution to be constant across locations while the scale parameter is location-specific); (5) A model featuring heterogeneous firms but with no internal geography (i.e., the entire country is modeled as a single location), which therefore closely resembles traditional trade models in the spirit of Melitz (2003) and Chaney (2008); and (6) A model with endogenous agglomeration in the tail of the firm productivity distribution, where the Pareto shape parameter changes as cities grow or shrink following counterfactual policies (see Appendix C).

We first discuss the welfare gains from trade (column 1 in Table 8). Our baseline model predicts gains equal to 4.7% and 4.5% for China and France, respectively. In all alternative models, the welfare gains from trade are meaningfully larger than in our model, especially in the case of China. Gains from trade arise as a reduction in trade costs increases the value of exports and imports in the economy. Since in our model (as in the data), exporters are spatially concentrated in larger cities, exporting firms face relatively high local wages due to high housing costs, and those increase over-proportionally in large cities when exports expand following a reduction in trade costs. In other words, our model implies that the spatial concentration of exporting in large cities limits

[48] Conceptually, this benchmark is closely related to the increasing-returns-to-scale model presented in Redding (2016). We choose the increasing returns model as it is closest to our model.

the ability to expand exporting, which dampens the gains from trade.[49] When we endogenize the Pareto shape parameter in row 6, gains from trade are slightly larger, but they remain below those derived from models without internal geography. These results highlight that our mechanism has important implications for aggregate statistics.

Next, we quantitatively evaluate the prediction from our simple model that, by reallocating workers to larger cities, opening up to trade increases the nominal wage gap as well as the real income gap between larger and smaller cities. We report the effect of trade on the ratio between the 90th and the 10th percentile of city-level population, nominal wages, and real wages in columns 2-4 in Table 8.[50] In our baseline model, trade increases the population ratio by 11.2% in China and by 7.9% in France, while the nominal wage ratio increases by 2.0% and 2.5%. The real income ratio increases by 2.5% in China and 2.4% in France. Note that the difference between nominal and real wage changes is not just driven by house prices (which rise in larger relative to smaller cities) but also due to relative changes in the price index of the tradable good, whose change is not monotonically tied to city size. These results highlight the important effect that international trade has on regional inequality.[51] Following a reduction in trade cost, smaller cities shrink while larger cities grow, and workers who remain in smaller cities experience a reduction in their real income while workers in larger cities experience real wage gains. Except for the model with endogenous shape parameter (row 6), all competing models deliver smaller spatial reallocation and changes in nominal wages / real income due to trade liberalization. Interestingly, restricting the upper tail of the productivity distributions to be the same across cities (row 4) leads to similar effects as a model with homogeneous firms (row 2).[52] Shutting down internal trade costs (row 3) also reduces spatial reallocation, although significantly more in China than in France. In contrast, allowing for an endogenous response of the Pareto shape parameter to city size leads to significantly larger spatial reallocation and to higher wage- and real-income-gaps between larger and smaller cities.

In our model, changing trade costs affects the aggregate trade elasticity. Figure 5 illustrates this prediction, plotting the country-level trade elasticity and the spatial concentration of popula-

[49]This important role for land scarcity in cities for the gains from trade echoes the recent work by Ducruet, Juhasz, Nagy, and Steinwender (2020), who emphasize the high land consumption (and local disamenities) of container ports, which offsets their market access gains in port communities. Similar to our work, Ducruet et al. (2020) also highlight how a reduction in international trade costs can have heterogeneous effects across domestic locations. In their case, this heterogeneity is driven by differences in the cost of port development, while in our setting, differences in productivity tails play a central role.

[50]Figure A.4 visualizes the differences in population reallocation from autarky to trade along the city size distribution.

[51]The effect of trade on wage inequality is diluted in the data by compositional changes of high- and low-skilled workers across space (in particular, by low-skilled workers moving to large cities). We thus study this effect through the lens of our quantitative model. We discuss these issues in more detail in Appendix A.8.

[52]Intuitively, the firm heterogeneity in our model by itself does not lead to a heterogeneous transmission of an aggregate trade shock. However, it amplifies the effects induced by differences in variable trade costs through the selection margin.

tion for different levels of trade costs (τ_c). As trade costs decline (moving to the left on the x-axis), economic activity and exporting become more concentrated in larger cities. As we have shown both theoretically (in Proposition 5) and empirically (in our PNTR regression in Table 6), the trade elasticity is smaller in larger cities. Thus, when overall trade costs decline, the aggregate trade elasticity falls in absolute terms. This result complements recent findings highlighting how the trade elasticity changes with trade costs (Adão et al., 2020), providing a novel, spatial mechanism. Our results, in particular, imply that an open economy with relatively low housing supply constraints will be more resilient to short-term tariff hikes than a relatively closed one. This is because in the former, trade has led to a concentration of exports in large cities, thereby lowering its elasticity with respect to trade shocks.

Spatial Policies

Our second counterfactual exercise studies the effects of reducing land-use restrictions on aggregate productivity and export intensity at the country level. We implement this policy as a reduction in the parameter γ, which measures the intensity of land use in the housing production function, in the home country. Reducing γ tilts the housing supply curves such that the price of housing increases less steeply with population. As a result, the cost of living and, therefore, wages increase less steeply with city size, as workers require lower wages to be willing to move to larger cities. This, in turn, increases labor demand from firms in larger cities relative to smaller cities, leading to a reallocation of population to the former. Overall, this affects aggregate export intensity in a number of ways: First, since firms in larger cities are more productive and are more likely to export, this pushes up country-level export intensity (this is the only mechanism that is also active in our simple model). Second, this reallocation increases total income in the domestic country and thereby raises domestic relative to foreign demand, reducing aggregate export intensity. Third, this positive effect on aggregate export intensity gets attenuated as the additional entry into the export market reduces the price index in the foreign country and thereby the exports of all firms. Finally, the spatial reallocation of factors also affects domestic sales across locations in an ex-ante ambiguous way. Thus, while total exports unambiguously increase, the overall effect of planning deregulation on export *intensity* is ambiguous.

To quantitatively explore these effects, we increase the housing supply elasticity in the model from the 25th to 75th percentile of previous estimates. For China, we use the range of elasticities estimated by Wang, Chan, and Xu (2012). For France, we use estimates based on US data from Saiz (2010). Table 9 displays the results. Column 1 shows that the concentration of population in large cities increases substantially. Next, columns 2-4 show that for China, aggregate exports grow more than sales, so that export intensity increases as a result of deregulated urban planning. For France, in contrast, sales grow more than exports, so that export intensity declines. Finally, aggregate productivity (column 5) increases particularly strongly (by 6%) in China as a result of the more elastic housing supply. This relatively large number (as compared to 1.6% in France) is likely not only a result of housing supply constraints, but also of China's restrictive Hukou system, which limits migration to productive cities. These aggregate effects are towards the lower end of findings in the existing literature, which is largely focused on the US.[53]

6 Conclusion

Trade policy has received renewed interest in recent years, as globalization has been blamed for widening disparities in many developed countries (Autor et al., 2013; Ezcurra and Pose, 2013; Dix-Carneiro and Kovak, 2017a; Potlogea, 2018). In response to this interest, a nascent literature has begun to analyze the interplay between trade and economic geography within countries. We contribute to this literature in three ways. First, using information from four major trading nations – China, France, the United States, and Brazil – we have documented a novel and highly robust stylized fact: Exporting is more unevenly distributed than overall economic activity, and in particular, it is disproportionately concentrated in larger cities. Importantly, this stylized fact is not driven by larger cities benefiting from better foreign market access. Second, we show that a relatively simple framework – an extension of the standard quantitative spatial equilibrium framework that includes firm heterogeneity and a mechanism of selection into exporting in the spirit of Melitz (2003) – can explain this stylized fact. Third, we structurally estimate the model and use it to undertake counterfactual policy analyses.

Our model is designed to assess the effects of both trade policies and domestic spatial policies, giving rise to novel interactions between these two levers. The corresponding welfare implications are richer and differ from those in more parsimonious models that are nested in our framework: a standard trade model that ignores within-country geography, and an economic geography model without rich firm heterogeneity. Looking forward, our theoretical framework opens the door to exploring the rich interactions between internal and international trade policies (e.g., infrastructure investments vs. trade agreements). Our novel mechanism also has implications for the design of optimal spatial policies and transfers, in the spirit of Fajgelbaum and Gaubert (2020), constituting a promising avenue for future research.

[53] In a model with endogenous housing supply restrictions, Parkhomenko (2023) finds that lowering regulations in a group of 10 US 'superstar' cities would increase productivity by 3.6%. By contrast, Hsieh and Moretti (2019) find that reducing regulations in New York, San Francisco, and San Jose alone would raise output by 3.7%-8.9% while Herkenhoff, Ohanian, and Prescott (2018), in a state-level analysis, find that reducing regulations everywhere in the US to half the level of Texas would increase aggregate productivity by 12.4% to 19.4%.

References

Adão, R., C. Arkolakis, and S. Ganapati (2020). Aggregate Implications of Firm Heterogeneity: A Nonparametric Analysis of Monopolistic Competition in Trade Models. Technical report, National Bureau of Economic Research.

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

Anderson, J. E. and E. van Wincoop (2004). Trade Costs. *Journal of Economic Literature 42*(3), 691–751.

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.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review 103*(6), 2121–2168.

Axtell, R. L. (2001). Zipf distribution of u.s. firm sizes. *Science 293*(5536), 1818–1820.

Baldwin, R. (2006). Globalisation: The Great Unbundlings. In *Globalisation Challenges for Europe*, pp. 5–47. Helsinki: Prime Minister's Office, Economic Council of Finland.

Baldwin, R. (2016). *The Great Convergence: Information Technology and the New Globalization*. Cambridge, MA: Harvard University Press.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy 122*(3), 507–553.

Behrens, K., C. Gaigne, G. I. Ottaviano, and J.-F. Thisse (2006a). How density economies in international transportation link the internal geography of trading partners. *Journal of Urban Economics 60*(2), 248–263.

Behrens, K., C. Gaigne, G. I. Ottaviano, and J.-F. Thisse (2006b). Is remoteness a locational disadvantage? *Journal of Economic Geography 6*(3), 347–368.

Behrens, K., C. Gaigne, G. I. Ottaviano, and J.-F. Thisse (2007). Countries, regions and trade: On the welfare impacts of economic integration. *European Economic Review 51*(5), 1277–1301.

Behrens, K., C. Gaigne, G. I. Ottaviano, and J.-F. Thisse (2009). Industry location and welfare when transport costs are endogenous. *Journal of Urban Economics 65*(2), 195–208.

Bernard, A. B. and J. B. Jensen (1999). Exceptional Exporter Performance: Cause, Effect, or Both? *Journal of International Economics 47*(1), 1–25.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2014). Challenges of working with the Chinese NBS firm-level data. *China Economic Review 30*(C), 339–352.

Carrère, C., M. Mrázová, and J. P. Neary (2020). Gravity without apology: the science of elasticities, distance and trade. *The Economic Journal 130*(628), 880–910.

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

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics 98*(C), 17–49.

Chen, N., D. Novy, C. Perroni, and H. C. Wong (2023). Urban-biased structural change.

Cheng, W. and A. Potlogea (2020). Trade Liberalization and Economic Development: Evidence from China's WTO Accession.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica 80*(6), 2543–2594.

Coşar, A. K. and P. D. Fajgelbaum (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics 8*(1), 24–56.

Costa, F., J. Garred, and J. P. Pessoa (2016). Winners and losers from a commodities-for-manufactures trade boom. *Journal of International Economics 102*(C), 50–69.

Dauth, W., S. Findeisen, and J. Suedekum (2014). The rise of the east and far-east: German labor markets and trade integration. *Journal of the European Economic Association 12*(6), 1643–1675.

Davis, D. R. and J. I. Dingel (2014). The Comparative Advantage of Cities. NBER Working Papers 20602, National Bureau of Economic Research, Inc.

Davis, D. R. and J. I. Dingel (2019). A Spatial Knowledge Economy. *American Economic Review 109*(1), 153–170.

Dean, J. M. and M. E. Lovely (2010). Trade growth, production fragmentation, and china's environment. In *China's growing role in world trade*, pp. 429–469. University of Chicago Press.

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

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *The American Economic Review 106*(3), 479–524.

Dingel, J. I., A. Miscio, and D. R. Davis (2019). Cities, Lights, and Skills in Developing Economies. NBER Working Papers 25678, National Bureau of Economic Research, Inc.

Dix-Carneiro, R. and B. Kovak (2017a). Trade Liberalization and Regional Dynamics. *American Economic Review 107*, 2908–2946.

Dix-Carneiro, R. and B. K. Kovak (2015). Trade Liberalization and the Skill Premium: A Local Labor Markets Approach. *American Economic Review 105*(5), 551–557.

Dix-Carneiro, R. and B. K. Kovak (2017b). Trade Liberalization and Regional Dynamics. *American Economic Review 107*(10), 2908–2946.

Dix-Carneiro, R. and B. K. Kovak (2019). Margins of labor market adjustment to trade. *Journal of International Economics 117*(C), 125–142.

Ducruet, C., R. Juhasz, D. K. Nagy, and C. Steinwender (2020). All aboard: The effects of port development.

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.

Eckert, F., S. Ganapati, and C. Walsh (2022). Urban-biased growth: a macroeconomic analysis. Technical report, National Bureau of Economic Research.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial Sorting. *Journal of Political Economy 122*(3), 554–620.

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.

Ezcurra, R. and A. R. Pose (2013). Does Economic Globalization affect Regional Inequality? A Cross-country Analysis. *World Development 52*, 92–103.

Fajgelbaum, P. and S. J. Redding (2022). Trade, structural transformation, and development: Evidence from argentina 1869–1914. *Journal of Political Economy 130*(5), 1249–1318.

Fajgelbaum, P. D. and C. Gaubert (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

Fernandes, A. M., P. J. Klenow, S. Meleshchuk, M. D. Pierola, and A. Rodríguez-Clare (2023). The intensive margin in trade: How big and how important? *American Economic Journal: Macroeconomics 15*(3), 320–354.

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.

Giannone, E. (2022). Skilled-biased technical change and regional convergence. *CREI. Unpublished manuscript*.

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.

Helpman, E. (2018). *Globalization and Inequality*. Harvard University Press.

Helpman, E. et al. (1995). *The size of regions*. Foerder Institute for Economic Research.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics 123*(2), 441–487.

Henderson, J. V., D. Nigmatulina, and S. Kriticos (2021). Measuring urban economic density. *Journal of Urban Economics 125*(C).

Herkenhoff, K., L. E. Ohanian, and E. C. Prescott (2018). Tarnishing the Golden and Empire States: Land-Use Regulations and the U.S. Economic Slowdown. *Journal of Monetary Economics 93*, 89–109.

Hsieh, C.-T. and E. Moretti (2019). Housing Constraints and Spatial Misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Kovak, B. K. (2013). Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization? *American Economic Review 103*(5), 1960–1976.

Krugman, P. and R. Livas Elizondo (1996). Trade policy and the Third World metropolis. *Journal of Development Economics 49*(1), 137–150.

Lind, N. and N. Ramondo (2023). Trade with correlation. *American Economic Review 113*(2), 317–353.

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

Melitz, M. J. and S. J. Redding (2015). New Trade Models, New Welfare Implications. *American Economic Review 105*(3), 1105–46.

Monfort, P. and R. Nicolini (2000). Regional Convergence and International Integration. *Journal of Urban Economics 48*(2), 286–306.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics 5*(1), 65–103.

Novy, D. (2013). International Trade without CES: Estimating Translog Gravity. *Journal of International Economics 89*(2), 271–282.

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.

Oh, R. (2023). Spatial sorting of workers and firms. *Yale University, Job Market Paper*.

Parkhomenko, A. (2023). Local causes and aggregate implications of land use regulation. *Unpublished manuscript*.

Pavcnik, N. (2002). Trade Liberalization, Exit and Productivity Improvements: Evidence from Chilean Plants. *Review of Economic Studies 69*, 245–76.

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

Potlogea, A. (2018). Globalization and the Skilled City. *Journal of Urban Economics 107*, 1–30.

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

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 Economics 125*(3), 1253–1296.

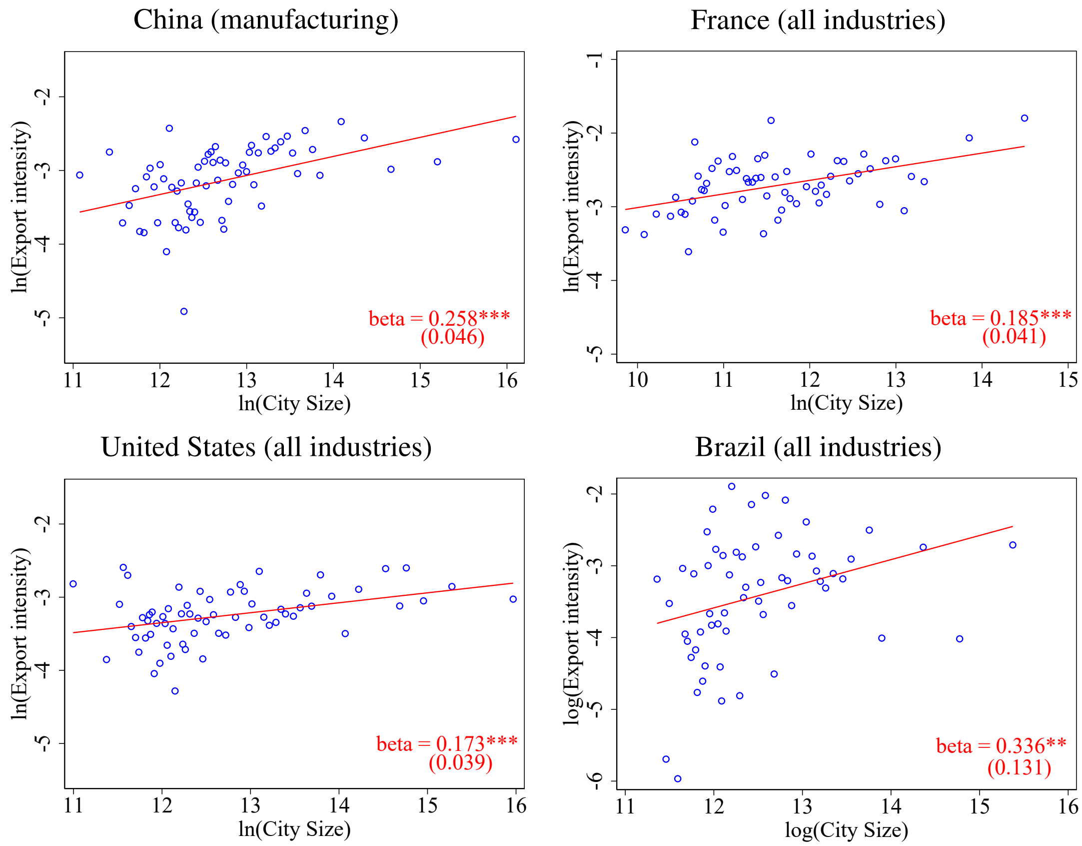
Schoefer, B. and O. Ziv (2022). Productivity, place, and plants. *Review of Economics and Statistics*, 1–46.

Walsh, C. (2023). The startup multiplier.

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.

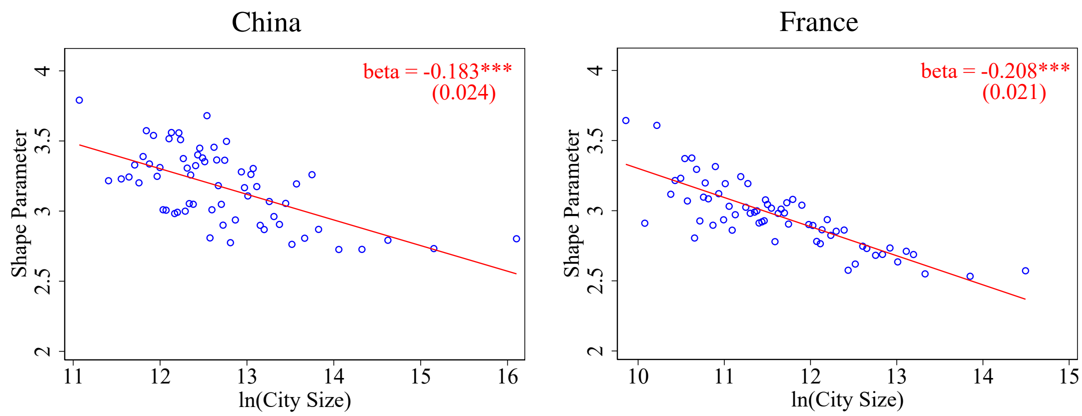
FIGURES

Figure 1: Export Intensity and City Size in China, France, the United States, and Brazil



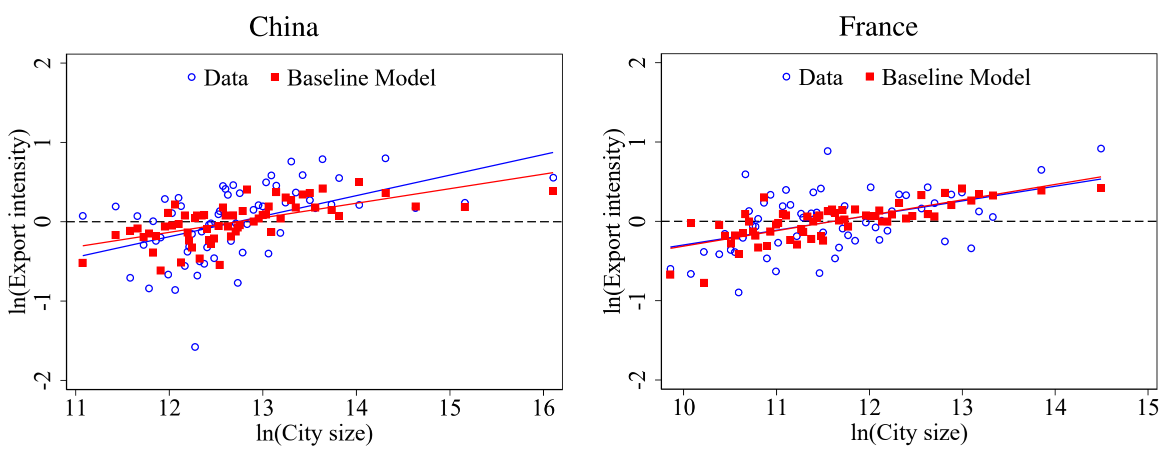
Notes: The figure shows a binned scatter plot between city-level log export intensity (exports relative to revenues) and log city size. Each dot (bin) represents 10 underlying cities for China, and 5 cities for the other three countries. Cities are defined in terms Microregions for Brazil, Metropolitan Areas for China and the United States; and Employment Zones for France. The analysis considers cities with positive exports and at least 250 manufacturing firms for China and France. For Brazil and the United States, the analysis considers cities with a population above 100,000 inhabitants. All figures include the following controls: ln average distance to other domestic cities, ln distance to the border, ln distance to the coast, border dummies, and coastal dummies.

Figure 2: Estimated Pareto Shape Parameter and City Size



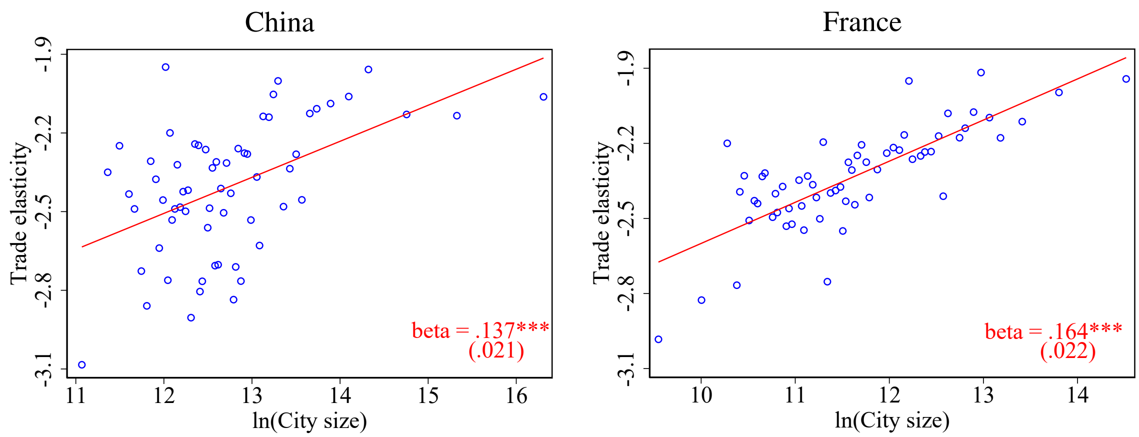
Notes: The figure provides evidence for a central assumption in our model: The thicker upper tail (i.e., lower Pareto shape parameter α_i) in larger cities. We show a binned scatter plot between the estimated α_i and city size for China and France. Each dot (bin) represents 10 underlying cities for China and 5 cities for France. See Section 5.2 for a description of the calibration procedure and Appendix E.3 for additional details.

Figure 3: Model vs. Data: (Predicted) Export Intensity and City Size



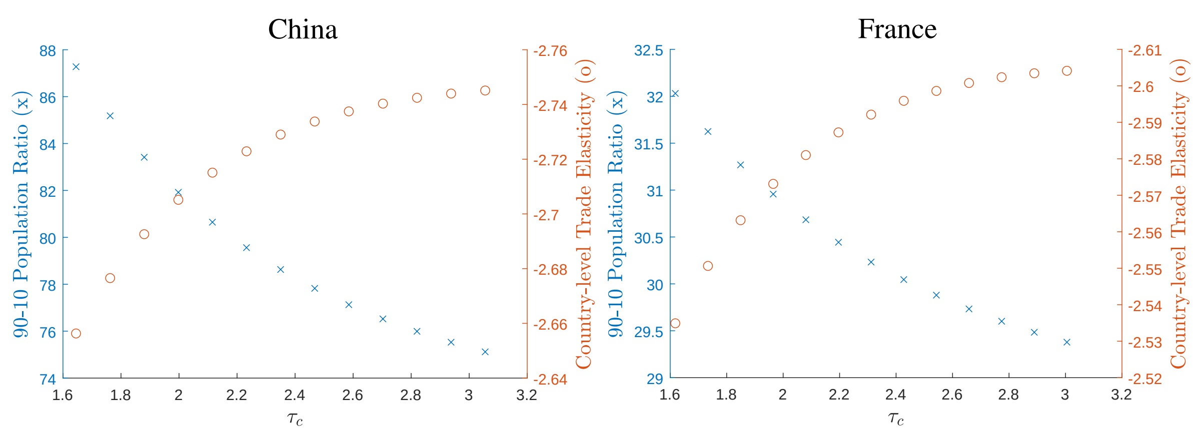
Notes: The figure plots a binned scatter plot between export intensity and city size for China and France for the empirical correlation in the data (blue-hollow circles) and the output of our quantitative model (red-solid squares). Each dot (bin) represents 10 underlying cities for China, and 5 cities for France. Both data and model output include linear regression lines. The elasticities of export intensity with respect to city size for China and France, respectively, are 0.183 and 0.194 in the model, as compared to 0.258 and 0.185 in the data.

Figure 4: Model Results: Trade Elasticity vs. City Size for China and France



Notes: This figure plots the local international trade elasticity implied by our quantitative model (at the data-implied equilibrium) against the population size of each location for China and France. We calculate the trade elasticity for each city from a marginal decrease in trade costs at the observed equilibrium. Each dot (bin) represents 10 underlying cities for China, and 5 cities for France.

Figure 5: Effect of Trade Costs on Aggregate Trade Elasticity and Population Concentration



Notes: This figure illustrates the effect of international trade costs (τ_c – on the x-axis) on the aggregate (i.e., country-level) trade elasticity (right axis) and on the spatial concentration of economic activity (left axis – the population ratio of cities in the top decile relative to the bottom population decile).

TABLES

Table 1: City Size and Export Intensity: Descriptive Statistics

		Population size		Export intensity	
	(1)	(2)	(3)	(4)	(5)
	Obs.	Mean	Std. dev.	Mean	Std. dev.
China (Metropolitan areas)	629	709.3	2,732.4	8.59	9.75
France (Employment zones)	304	192.5	374.5	8.40	6.13
United States (Metropolitan areas)	348	754.9	1,667.6	5.25	6.30
Brazil (Microregions)	316	463.7	1,129.1	9.03	12.67

Notes: The table shows statistics for the distribution of city sizes and export intensities in China, France, the U.S., and Brazil. Population size is in ’000s and export intensity in percent. Cities are defined in terms of Metropolitan Areas for China and the United States, Employment Zones for France, and Microregions for Brazil. The analysis considers cities with positive exports and at least 250 firms for China and France. For Brazil and the United States, the analysis considers cities with a population above 100,000 inhabitants.

Table 2: Export Intensity and City Size in Brazil, China, France, and the United States

Dependent Variable: City-Level ln(Export Intensity)

	— China —		—— France ——		— United States —		—— Brazil ——	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln City Size	0.338***	0.258***	0.215***	0.185***	0.177***	0.173***	0.337***	0.336**
	(0.052)	(0.046)	(0.043)	(0.041)	(0.037)	(0.039)	(0.121)	(0.131)
Geography Controls		✓		✓		✓		✓
Mean Export Intensity:	.088	.088	.084	.084	.056	.056	.096	.096
R^2	.043	.263	.083	.184	.049	.173	.019	.185
Observations	615	615	304	304	329	329	297	297

Notes: The table analyzes the relationship between city size and export intensity, both in logs. See the note to Table 1 for the definition of cities in the four countries. City-level export intensity is defined as manufacturing exports over manufacturing sales for China; overall exports over sales for France and the United States, and overall exports over (city-level) GDP for the case of Brazil. 'Geography Controls' include: ln(average distance to other domestic cities), ln(distance to the border), ln(distance to the coast), border dummies, and coastal dummies. Robust standard errors in parentheses. Key: ** significant at 1%; ** 5%; * 10%.

Table 3: Export Intensity and City Size: 2SLS Results

Dependent variable: As indicated in table header

	China				France			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	First Stage		2SLS		First Stage		2SLS	
Dep. Variable:	ln(Modern Urban Pop.)		ln(Export Intensity)		ln(Modern Urban Pop.)		ln(Export Intensity)	
ln City Size			0.736***	0.582**			0.446**	0.425**
			(0.248)	(0.275)			(0.206)	(0.209)
CSI	0.442***	0.391***			0.252***	0.246***		
	(0.051)	(0.056)			(0.049)	(0.047)		
CSI × Coastal	0.182	0.201			-0.200*	-0.205*		
	(0.309)	(0.310)			(0.121)	(0.113)		
First Stage F-Statistic			39.4	25.2			13.6	13.5
Geog. Indicators	✓	✓	✓	✓	✓	✓	✓	✓
Coastal Access Control		✓		✓		✓		✓
R^2	0.323	0.335	0.198	0.241	0.053	0.062	0.038	0.067
Observations	329	329	329	329	304	304	304	304

Notes: This table examines the effect of city size, instrumented with the historical caloric suitability index (CSI), on export intensity in China (columns 1-4) and France (columns 5-8). CSI is measured as the maximum attainable crop yield after the Colombian Exchange from Galor and Özak (2016). The analysis is run at the prefecture level for China and at the employment zone level for France. Columns 1, 2, 5, and 6 report the first stage regressions, and columns 3, 4, 7, and 8, the corresponding 2SLS results. We report the (cluster-robust) Kleibergen-Paap rK Wald F-statistic. The corresponding Stock-Yogo value for 10% maximal IV bias is 16.4. The corresponding reduced-form results are reported in Appendix Table A.9. '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). Key: *** significant at 1%; ** 5%; * 10%.

Table 4: Decomposition into Within- and Between-Sector Export Intensity

Dep. Var.: Within- and Between Components of City-Level Export Intensity

	China			France		
	(1)	(2)	(3)	(4)	(5)	(6)
Specification:	Within	Between	% Within	Within	Between	% Within
ln(City Size)	.292***	-.033**	>100%	.097***	.088***	52.4%
	(.043)	(.015)		(.028)	(.023)	
Geography Controls	✓	✓		✓	✓	
R^2	.213	.261		.171	.148	
Observations	615	615		304	304	

Notes: This table decomposes city-level export intensity into a within- and a between-industry component. Note that the coefficients on the two components add up to the overall coefficient in Table 2, col 2 (for China) and col 4 (for France). 'Geography Controls' are listed in the note to Table 2. Robust standard errors in parentheses. Key: *** significant at 1%; ** 5%; * 10%.

Table 5: Checking for the Role of Variable Transport Costs: Two Margins of Export Intensity

Dependent Variable as listed in table header

	China			France		
	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Export intensity)	ln(Intensive margin)	ln(Extensive margin)	ln(Export intensity)	ln(Intensive margin)	ln(Extensive margin)
ln(City Size)	0.258***	-.186***	0.444***	.185***	.018	.167***
	(.046)	(.033)	(.045)	(.041)	(.033)	(.026)
Geog. Controls	✓	✓	✓	✓	✓	✓
R^2	0.263	0.123	0.229	0.184	0.052	0.289
Observations	615	615	615	304	304	304

Notes: The table decomposes overall export intensity (x_i/r_i in cols 1,4 – with x_i and r_i denoting export revenues and total revenues, respectively, in city i) into the importance of exporting among exporters (the intensive margin x_i/r_i^x in cols 2,5 – with r_i^x denoting total (incl. domestic) revenues of exporters) and the (value) extensive margin (r_i^x/r_i in cols 3,6). 'Geog. Controls' are listed in the note to Table 2. Robust standard errors are in parentheses. Key: ** significant at 1%; ** 5%; * 10%.

Table 6: Empirical Test of Proposition 5: Heterogeneous Local Trade Elasticity

Dependent Variable:	ln(exports$_{ijt}$)				$\ln\begin{pmatrix}\text{Intensive}\\ \text{margin}\end{pmatrix}$	$\ln\begin{pmatrix}\text{Extensive}\\ \text{margin}\end{pmatrix}$
	(1)	(2)	(3)	(4)	(5)	(6)
$\beta_1 : [Post\ PNTR_t \times NTR\ Gap_j]$	0.159***	0.154***	0.180**	0.171***	0.032	0.139***
	(0.051)	(0.052)	(0.070)	(0.060)	(0.050)	(0.043)
$\beta_2 : [Post\ PNTR_t \times NTR\ Gap_j] \times ln(pop_i)$	-0.095***	-0.093***	-0.097**	-0.092**	-0.026	-0.066**
	(0.031)	(0.032)	(0.040)	(0.037)	(0.028)	(0.030)
Industry FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
$ln(pop_i)$	✓	✓	✓	(✓)	(✓)	(✓)
$ln(pop_i) \times Post\ NTR_t$	✓	✓	✓	✓	✓	✓
$ln(pop_i) \times NTR\ Gap_j$	✓	✓	✓	✓	✓	✓
City controls		✓	✓	(✓)	(✓)	(✓)
City controls $\times Post\ PNTR_t$			✓	✓	✓	✓
City controls $\times NTR\ Gap_j$			✓	✓	✓	✓
City controls $\times [Post\ PNTR_t \times NTR\ Gap_j]$			✓	✓	✓	✓
City FE				✓	✓	✓
Observations	38,269	38,269	38,269	38,266	38,266	38,266
Pseudo R^2	0.59	0.61	0.62	0.67	0.53	0.71

Notes: The table presents evidence for Proposition 5, which predicts that the local trade elasticity is smaller in larger cities and is driven by the extensive margin (log number of exporters). This is captured by the negative coefficient β_2 on the triple interaction. The dependent variable in cols 1-5 is the log value of exports in prefecture c in year t in industry i. The dependent variable in column 6 is log average exports per exporter), while the dependent variable in column 7 corresponds to the log number of exporters. $PostPNTR_t$ is a dummy that takes on value one for $t \geq 2001$, and $NTRGap_j$ is the counterfactual increase in export tariffs for sector j from China to the US that would have occurred if China's normal trade relations (NTR) status had lapsed, rather than becoming permanent (PNTR) in 2001. Thus, a higher $NTRGap_j$ reflects a larger reduction in tariffs, relative to the counterfactual. The regression uses the standardized $NTRGap_j$, with mean zero and std one. 'City Controls' include distance to the coast (in logs), coastal dummy, distance to the border (in logs), border dummy, the average distance to other domestic cities (in logs), and capital-to-labor ratio (in logs). Standard errors clustered at the city and at the industry level in parentheses. Key: *** significant at 1%; ** 5%; * 10%.

Table 7: Export Intensity and City Size: Various Models vs. Data

Variable for which correlation with ln(city size) is reported:	(1) $\ln\begin{pmatrix}\text{Export}\\\text{intensity}\end{pmatrix}$	(2) $\ln\begin{pmatrix}\text{Intensive}\\\text{margin}\end{pmatrix}$	(3) $\ln\begin{pmatrix}\text{Extensive}\\\text{margin}\end{pmatrix}$
Panel A: China			
(1) Data	0.258***	-0.186***	0.444***
	(0.046)	(0.033)	(0.045)
(2) Baseline Model†	0.183***	0.015**	0.169***
	(0.026)	(0.008)	(0.022)
(3) Homogeneous firms model‡	0.016**	0.016**	0.000
	(0.009)	(0.009)	(0.000)
(4) Baseline without internal trade costs	0.137***	0.000	0.137***
	(0.018)	(0.000)	(0.018)
(5) Baseline with common shape parameter	0.017**	0.016**	0.002**
	(0.010)	(0.009)	(0.001)
Panel B: France			
(1) Data	0.185***	0.018	0.167***
	(0.041)	(0.033)	(0.026)
(2) Baseline Model†	0.194***	0.006	0.188***
	(0.024)	(0.008)	(0.021)
(3) Homogeneous firms model‡	0.007	0.007	0.000
	(0.009)	(0.009)	(0.000)
(4) Baseline without internal trade costs	0.177***	0.000	0.177***
	(0.019)	(0.000)	(0.019)
(5) Baseline with common shape parameter	0.008	0.007	0.001
	(0.009)	(0.008)	(0.001)

Notes: The table reports the coefficients of regressing the variables shown in the table header on ln(city size). All specifications include the 'Geography Controls' listed in the note to Table 2. In row 1, results stem from the actual data. In rows 2-5, the underlying regression input comes from our quantitative model. The intensive margin (col 2) reflects the export intensity of exporters, and the extensive margin (col 3) is the total sales ratio of exporting firms to all firms. Cities are defined in terms of Metropolitan Areas and Employment Zones for China and France, respectively. The analysis considers cities with positive exports and at least 250 firms. Robust standard errors are in parentheses. Key: *** significant at 1%; ** 5%; * 10%.

† City-specific Pareto shape and scale parameters.

‡ City-specific average productivity: firms are *homogeneous* within cities, while allowing for variation in productivity levels *across* cities.

Table 8: Model Counterfactual: Trade Liberalization

	(1) Aggregate Welfare	(2) Population Share 90-10 Ratio	(3) Nominal Wage 90-10 Ratio	(4) Real Income 90-10 Ratio
Panel A: China				
(1) Baseline model[†]	4.7%	11.2%	2.0%	2.5%
(2) Homogeneous firms model[‡]	5.9%	10.2%	0.8%	1.6%
(3) Baseline without internal trade costs	4.8%	2.6%	1.5%	1.3%
(4) Baseline with common shape parameter	5.4%	10.2%	0.8%	1.6%
(5) Melitz – No internal geography	5.2%	—	—	—
(6) Baseline with endogenous shape parameter[§]	5.1%	13.8%	3.3%	3.8%
Panel B: France				
(1) Baseline model[†]	4.5%	7.9%	2.5%	2.4%
(2) Homogeneous firms model[‡]	5.1%	2.7%	0.7%	1.1%
(3) Baseline without internal trade costs	4.6%	5.3%	1.9%	1.3%
(4) Baseline with common shape parameter	4.8%	2.7%	0.7%	1.1%
(5) Melitz – No internal geography	5.0%	—	—	—
(6) Baseline with endogenous shape parameter[§]	4.6%	9.5%	3.2%	2.9%

Notes: The table shows the estimated effects of moving from autarky to trade in terms of aggregate welfare (col 1), spatial differences in population and wages (col 2-3), and overall wage inequality (col 4) for different versions of the model presented in Section 5.5 and discussed in more detail in Appendix F. 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. Row 5 presents results for a model that does not have any domestic trade costs and imposes that the Pareto *shape* parameter is equal across cities, while allowing for differences in the *scale* parameter of the local productivity distributions.

[†] City-specific Pareto shape and scale parameters.

[‡] City-specific average productivity: firms are *homogeneous* within cities, while allowing for variation in productivity levels *across* cities.

[§] Endogenous shape parameter: City-specific productivity shape parameter is allowed to change with city size in the counterfactual economy. See Appendix C.3 for detail and calibration.

Table 9: Model Counterfactual: Less Restrictive Land Use Regulation in the Baseline Model

		(1) Population Share 90-10 Ratio	(2) Aggregate Export	(3) Aggregate Sales	(4) Export Intensity	(5) Aggregate Productivity
(1)	China	73.9%	4.8%	4.3%	0.06% p.p.	6.0%
(2)	France	46.0%	0.9%	1.2%	-0.05 p.p.	1.6%

Notes: The table shows the estimated effects of increasing the housing supply elasticity from the 25th to the 75th percentile of previous estimates (see Section 5.5 for details) on the share of population living in the largest as compared to smallest decile of cities (col 1), aggregate exports (col 2), aggregate sales (col 3), export intensity (col 4), and aggregate productivity (col 5) for our baseline quantitative model.