Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 1964
November 2023

Rethinking revealed comparative advantage with micro and macro data

Hanwei Huang
Gianmarco I.P. Ottaviano

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Abstract
The Balassa's index of revealed comparative advantage does not necessarily reveal Ricardian comparative advantage. We propose an alternative sufficient statistics approach based on a quantitative standard trade model incorporating firm and product selection. We show that the model's micro foundations do not necessarily imply that the relevant data for the proposed sufficient statistics must include micro information, but its micro structure is needed to understand how only macro information can be used instead. Applying our approach to Chinese micro data and cross-country macro data, we find that firm behavior has far-reaching implications for understanding aggregate productivity and revealed comparative advantage.

Keywords: revealed comparative advantage, sufficient statistics, firm heterogeneity, multi-product firms

This paper was produced as part of the Centre's Trade Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

For helpful comments, we would like to thank Eric Bartelsman, Flora Bellone, Andy Bernard, Lorenzo Caliendo, Dave Donaldson, Peter Egger, Alejandro Graziano, James Harrigan, Elhanan Helpman, Zhiyuan Li, Mengxiao Liu, Hong Ma, Kalina Manova, Thierry Mayer, Marc Melitz, Peter Neary, Albert Park, Steve Redding, Ariell Reshef, Gerard Roland, Pete Schott, Davide Suverato, Cristina Terra, Miaojie Yu, Ben Zissimos, and participants at several conferences, seminars and workshops. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement n. 789049-MIMAT-ERC-2017-ADG).

Hanwei Huang, City University of Hong Kong, CIFER and Centre for Economic Performance at London School of Economics. Gianmarco Ottaviano, Bocconi University, BAFFI, IGIER, CEPR and Centre for Economic Performance at London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

The famous "four numbers" paragraph on comparative advantage in Ricardo (1817) is one of the oldest analytical results in economics (Bernhofen and Brown, 2018). Interpreted as unit labor requirements in the wake of James Mill (1821), those numbers have provided the basis of the Ricardian model of international trade from John Stuart Mill (1852) to Eaton and Kortum (2002).[1] The "four numbers" refer to the units of labor required to produce a unit of output in two industries that are active in two countries. A country has a Ricardian comparative advantage in an industry if in the country the unit labor requirement of that industry relative to the other industry is lower than in the other country. Equivalently, a country has a Ricardian comparative advantage in an industry if the country can supply that industry's product at a lower relative opportunity cost or relative autarky price, that is, at a lower relative marginal cost prior to trade given the assumption of perfect competition.

In the traditional Ricardian model, the unit labor requirements are exogenously determined, and the prediction is that a country exports the product of the industry in which it has a comparative advantage and imports the product of the other industry. Hence, the endogenous trade patterns "reveal" the exogenous Ricardian comparative advantage.

There are, however, two important complications with this simple "revelation principle." The first is readily understood with reference to the Heckscher-Ohlin model (Haberler, 1930). When labor is not the only input and relative factor intensities vary across sectors, opportunity costs are also determined by relative factor endowments, and comparative advantage in terms of opportunity costs is not necessarily aligned with Ricardian comparative advantage in terms of relative unit input requirements. The opportunity cost formulation of comparative advantage is more general, and the Ricardian and Heckscher-Ohlin models can be viewed as special cases highlighting relative unit input requirements and relative factor endowments as different sources of comparative advantage. For this reason, in defining his celebrated index of "revealed comparative advantage" based on the specialization of a country's exports, Balassa (1965) remains agnostic about the source of such advantage, also discussing how revealed comparative advantage may not necessarily coincide with comparative advantage based on relative costs, as recently argued by Costinot et al. (2012) and French (2017).

The second complication is related to recent developments in the theory of international trade with imperfect competition and firm heterogeneity following Melitz (2003), which highlights the distinction between exogenous and endogenous comparative advantage. Firm heterogeneity complicates the picture due to endogenous firms' and products' selection at the extensive margin with survival depending on their productivity (i.e. the inverse of their unit input requirement) as in Melitz (2003) and Bernard et al. (2011) respectively, as well as reallocation of market shares from less to more productive surviving firms at the intensive margin as in Melitz and Ottaviano (2008), and from less to more productive products within their surviving products as in Mayer et al. (2014). For example,

[1] While Bernhofen and Brown (2018) highlight that the labor unit interpretation of the "four numbers" obscures the logic of Ricardo's original exposition, we follow that interpretation in continuity with the referenced literature.

Bernard et al. (2007) show that, because of fiercer trade-induced selection of more productive firms in industries that are relative intensive in the relatively abundant factor, firm heterogeneity amplifies the exogenous Heckscher-Ohlin comparative advantage through an industry composition effect that works as a source of additional endogenous Ricardian comparative. Though not explicitly discussed, the distinction between the exogenous and endogenous aspects of comparative advantage is also salient in the Ricardian setup of Melitz and Ottaviano (2008), where selection affects the translation of a country's exogenous "state of technology" (Eaton and Kortum, 2002) in a given industry into the corresponding endogenous industry unit input requirement (Corcos et al. 2012).[2]

On account of both complications, one may say in modern terminology (Chetty, 2009) that the Balassa index of revealed comparative advantage based on the degree of a country's export specialisation is not a "sufficient statistic" for identifying the existence, the magnitude and the source of Ricardian comparative advantage. Cutting new ground in terms of sufficient statistics for this advantage is the goal of the present paper. This is pursued through a structural approach based on a quantitative trade model with firm and product selection where both relative unit input requirements and relative factor endowments play a role. We show that the model's micro foundations do not necessarily imply that the relevant data for the proposed sufficient statistics must include micro information (i.e. at the firm or product level), but its micro structure is needed to understand how only macro information (i.e. at the sector or country level) can also be used instead.

While we are not aware of any existing study revisiting the concept of revealed comparative advantage from this angle, we still build on the existing literature on comparative advantage with heterogeneous firms, which is nevertheless surprisingly scant. Despite two decades of studies on firm heterogeneity and growing interest in its macro implications for productivity and welfare (see, e.g., Melitz, 2003; Arkolakis et al., 2012; Melitz and Redding, 2015; Arkolakis et al., 2019), we still know relatively little about its interactions with comparative advantage. The little we know is essentially limited to two studies. The first is the aforementioned paper by Bernard et al. (2007), who embed heterogeneous monopolistically competitive firms à la Melitz (2003) in a standard Heckscher-Ohlin-Samuelson model.[3] The second is the paper by Gaubert and Itshoki (2021), who consider a Ricardian model with a continuum of industries à la Dornbusch et al. (1977), allowing for market structure to be oligopolistic. They show that with oligopolistic competition, firm granularity explains a non-trivial amount of variation in sectoral export participation and that firm dynamics play a crucial role in shap-

[2]Eaton and Kortum (2002) assume that firms draw their total factor productivity from Fréchet distributions featuring the same shape parameter but different means across countries. A country with a larger mean than another country is said to have a better "state of technology" in that the distribution of its firms' productivity draws first-order stochastically dominates the other country's distribution. Corcos et al. (2012) adopt a similar concept within the framework of Melitz and Ottaviano (2008), assuming that monopolistically competitive firms draw their total factor productivity from a common Pareto distribution with the same shape parameter and different lower bound of the support across countries.

[3]Burstein and Vogel (2017) and Huang et al. (2017) also consider multi-sector models with heterogeneous firms. More recently, Bai et al. (2022) propose an alternative way to introduce heterogeneous firms under perfect competition by assuming that a firm is a collection of productivity draws, each allowing for producing a homogeneous output. They show that trade need not make selections in the comparative advantage sector stricter as long as the entry costs are incurred in different combinations of inputs than production costs. See Footnote 12 in Bernard et al. (2007) for a similar point.

ing comparative advantage.[4] However, neither Bernard et al. (2007) nor Gaubert and Itshoki (2021) connect their findings to sufficient statistics for revealed comparative advantage.

Differently from these studies, we investigate how the notion of revealed comparative advantage can be operationalized through sufficient statistics if one is interested in the Ricardian source of comparative advantage. We do so by developing a multi-sector Ricardian-Heckscher-Ohlin model with heterogeneous multi-product firms. The model features two countries and a continuum of monopolistically competitive industries, with Ricardian and Heckscher-Ohlin sources of comparative advantages as Dornbusch et al. (1977) and (1980), respectively. Within each industry, firms are heterogeneous in terms of productivity as in Melitz (2003) with single-product firms and Bernard et al. (2010) with multi-product firms, and average industry efficiency is determined by an endogenous survival cutoff for firm efficiency. For our purposes, an important limitation of these works is that the CES ("constant elasticity of substitution") demand system they use implies that the passthrough of the exogenous component of Ricardian comparative advantage to autrakic relative industry prices is one-to-one (hence complete) as selection is immaterial in closed economy (Melitz, 2003), which removes by assumption a potentially important aspect of its transmission to observed outcomes.[5]

To relax the complete passthrough straightjacket, one could adopt a simple linear demand system – as in Melitz and Ottaviano (2008) with single-product and Mayer et al. (2014) with multi-product firms. To ensure wider applicability, we prefer to use a more flexible family of VES ("variable elasticity of substitution") demand systems (Bulow and Pfleiderer, 1983; Atkin and Donaldson, 2015; Mrázová and Neary, 2017; Mayer et al., 2021), leaving it to the data to nail down a specific parametrization. This family encompasses CES and linear demands as special cases, and a continuum of other options, enabling us to evaluate the empirical appeal of alternative demands.

With our model, we do four things, leveraging detailed manufacturing data on China for proof of concept. First, we characterize the conditions that the model fundamentals must satisfy regarding the production technology and consumer preferences to match three key stylized facts on Chinese exports: (i) the export propensity and the export intensity of Chinese industries rise with their labor intensity and productivity; (ii) firms' export product scope (i.e. the number of exported products) increases with their labor intensity and productivity; (iii) the export product mix (i.e. the distribution of exports across products) is more skewed for capital-intensive and low-productivity firms than for labor-intensive and high-productivity firms. The first two facts can be rationalized by extending multi-product firm CES models, such as those by Bernard et al. (2011) and Arkolakis et al. (2021), to multiple industries with different factor intensities and unit input requirements. However, as the CES assumption imposes an exogenously fixed markup, differences in market conditions across industries

[4] Gaubert and Itshoki (2021) depart from Bernard et al. (2007) by breaking both the assumptions of constant markups and atomistic firms. We retain the assumption of atomistic firms but allow for variable markup.

[5] In general, with selection, there are two types of passthrough at the industry level holding the unit input price constant: from the unconditional average unit input requirement to the conditional unit input requirement and thus to the marginal cost, and from the latter to the average industry price. With CES, both types are complete. At the firm level, there is only one type of passthrough from marginal cost to price, which is also complete with CES.

do not affect firms' export product mix. Hence, the third stylized fact calls for variable markups, which is the case with VES but not with CES demand. We also directly test that the demand system is VES rather than CES by estimating firms' export pricing equations. Crucially, we argue that, for the model to match those facts, selection into the *export* market has to be weaker in comparative advantage industries. This tends to dampen their comparative advantage and works against the amplifying effect due to selection into the *domestic* market, with the balance between the two effects depending on the degree of firm heterogeneity in the presence of variable markup and incomplete passthrough.

Second, we use the model to discuss how to compute theoretically consistent measures of Ricardian comparative advantage through a sufficient statistics approach, and how either micro or macro data can be used to construct such statistics. Third, we apply the sufficient statistics approach to Chinese data and quantify the exogenous and endogenous components of relative industry productivity of China with respect to the rest of the world. We find that, as the endogenous amplifying and dampening components do not offset each other, the observed differences in relative input requirements across industries provide biased estimates of China's Ricardian comparative advantage.

Finally, we conduct a counterfactual analysis to investigate how rising labor costs may change the ranking of Chinese industries in terms of the Balassa index by making their Ricardian comparative advantage more salient. We find that higher labor costs dent China's top-ranked industries but bolster the bottom-ranked ones to an extent that depends on the degree of firm heterogeneity and passthrough.

Our paper contributes to the following strands in the literature. It is closely related to recent studies of the macro implication of firm heterogeneity. Arkolakis et al. (2012) find that, for a class of popular models satisfying certain restrictions, with or without firm heterogeneity, a country's welfare gains from trade can be evaluated through sufficient statistics consisting of the trade elasticity and the share of expenditure on domestic goods.[6] Similarly, we show that sufficient statistics can also be used to assess Ricardian comparative advantage, and that, despite firm heterogeneity, firm-level data is not necessary to gauge the importance of export selection. As already discussed, Bernard et al. (2007) highlight the amplifying effect of firm selection with respect to comparative advantage. We complement their analysis by showing that there is another channel through which firm heterogeneity shapes comparative advantage: softer competition in the export market induces within-industry reallocations that dampen the comparative advantage.

We also contribute to the literature on the measurement of comparative advantage. There has been a renaissance in the quantification of Ricardian comparative advantage since the seminal contribution by Eaton and Kortum (2002).[7] Differently from existing works, we provide sufficient statistics re-

[6]The restrictions include CES preferences and a constant trade elasticity. Arkolakis et al. (2019) depart from these two restrictions and study welfare gains from trade in models with variable markups. Melitz and Redding (2015) show that the Melitz model with firm heterogeneity implies higher welfare gains from trade than the Krugman model with homogeneous firms.

[7]Costinot et al. (2012) estimate the importance of Ricardian comparative advantage on trade patterns and welfare using an extended Eaton-Kortum model. Levchenko and Zhang (2016) use the gravity equation to infer comparative advantage from trade flows and their evolution over time. Costinot et al. (2016) focus on the agriculture sector, for which parcel-level land productivity can be precisely estimated for different crops. Redding and Weinstein (2018) develop a method based

sults that identify Ricardian comparative advantage as the exogenous component of observed relative industry productivity. We also show that the exact productivity measures matter when measuring productivity gains from trade. Measures that consider only export selection at the extensive margin and neglect export reallocation at the intensive margin miss an important endogenous determinant of industries' relative unit input requirements and bias their estimation upwards.

Finally, we contribute to the theoretical and empirical literature on multi-product firms.[8] Our analysis highlights how comparative advantage affects factor reallocation along the within-firm extensive and intensive margins, and how it feeds back to industry relative productivity. The mechanism is similar to that of Mayer et al. (2014, 2021). While their primary emphasis lies in competition driven by market size, our focus centers on the implications for comparative advantage. By doing so, we provide a finer characterization of multi-product exports in a world with many industries.

The rest of the paper is organized as follows. Section 2 describes the data and the three stylized facts that motivate our analysis. Section 3 presents the model, derives the implications for Ricardian comparative advantage, and develops the sufficient statistic approach. Section 4 implements this approach using Chinese micro data and cross-country macro data. Section 5 concludes.

2 Motivating Evidence

This section presents three stylized facts on how exports vary with labor intensity and productivity across Chinese manufacturing industries, emphasizing export participation, exporters' product scope, and product mix. These stylized facts motivate and discipline the model in the subsequent section.

2.1 Data

The facts are generated using the following two datasets. The first dataset is the Annual Survey of Industrial Firms (ASIF), which covers all State-Owned Firms (SOEs) and non-SOEs with sales above 5 million Chinese Yuan. It provides rich information on firms' financial statements and forms of identification such as name, address, ownership, etc. The other dataset is Chinese Customs Trade Statistics (CCTS), which covers all Chinese import and export transactions. For each transaction, we know the identity of the Chinese firm, the product code at the 8-digit Harmonized System (HS) level, value, quantity, origin/destination, etc. As no firm identifier exists between the ASIF and CCTS, we match the two datasets by firms' names, addresses, telephone numbers, and zip codes.[9] We focus on

on a CES model to account for trade patterns and construct indexes of comparative advantage using micro trade data.

[8]Feenstra and Ma (2009), and Eckel and Neary (2010) examine the cannibalization effect for multi-product oligopolistic firms. Arkolakis et al. (2021) and Bernard et al. (2011) emphasize selection along the extensive margin, while Mayer et al. (2014, 2021) focus on selection along the intensive margin. Manova and Yu (2017) consider product selection along the quality margin instead. Bernard et al. (2010) and Iacovone and Javorcik (2010) investigate product churning over time in response to changes in market conditions.

[9]This matching method has been commonly used in the literature. See, Yu (2015), and Manova and Yu (2017). The matched sample represents about 37% of all Chinese exports reported in the customs data in 2000 and 52% in 2006. Trade

Chinese manufacturers and exclude firms from the recycling, mining, and utility sectors in ASIF, as well as wholesalers and trade intermediaries in CCTS.

We measure the inverse of unit input requirements in terms of total factor productivity (TFP), which we estimate at the firm level using the method by Ackerberg et al. (2015) and a panel of ASIF firms for 1999-2007 constructed following Brandt et al. (2012) and Brandt et al. (2017). Under the assumptions that the technology is Cobb-Douglas and 4-digit industries exhibit the same factor shares within the same 2-digit industry, we first estimate value-added production functions by 2-digit industry, relying on the Chinese Industry Classification (CIC) while accounting for China's accession to the WTO in 2001, for firm ownership and for export and processing trade status (Yu, 2015; Brandt et al., 2017). Industry productivity for each 4-digit CIC industry is then obtained as the simple average of the logarithm of its firms' productivities within the same 4-digit industry.

As for industry factor intensity, we first measure firms' labor intensity by labor costs relative to value-added, with firms' labor costs including payable wages, labor and employment insurance fees, and the total of employee benefits payable.[10] We then take the average labor intensity across firms within each 4-digit CIC industry to obtain the industry labor intensity. Given these measures, we find the following facts for 2004, the year we will also use for the quantification analysis in Section 4.[11]

2.2 Stylized Facts

Fact 1: *Industries' export propensity and export intensity rise with labor intensity and productivity.*

An industry's export propensity is defined as the ratio of the number of exporters to the total number of firms, and export intensity as the ratio of total exports to total sales. Panels (a) and (b) in Figure 1 present 3-dimensional scatter plots of export propensity and intensity respectively, against industry labor intensity and productivity together with the best-fit linear plane. The results indicate that both export propensity and intensity increase with industry labor intensity and productivity.[12]

We conduct formal regression analyses in Panel A of Table 1 to control for confounding factors influencing firms' export participation. In columns (1) and (5), we replicate the results of Figure 1 by regressing export propensity and intensity on industry labor intensity and productivity, respectively. These columns confirm that export propensity and intensity increase with industry labor intensity and productivity. In columns (2)-(4) and (6)-(8), we progressively add controls for shares of state-owned enterprises and foreign-invested enterprises, average firm age, and the share of processing trade firms in each industry. The positive association between export propensity/intensity and industry labor intensity/productivity remains strong.

intermediaries carried a significant amount of the remaining exports (Ahn et al., 2011).

[10]We exclude firms with labor intensities that are negative or greater than 1. Their presence is likely due to misreporting or data input errors. We also exclude firms with negative value-added, employment, or assets. Firms with fewer than eight employees are also excluded since they are under a different legal regime.

[11]We also access the 2004 Industrial Census of China, which complements the quantitative analysis later.

[12]A similar pattern emerges from the best-fit linear plane in an unreported 3-dimensional scatter plot of total exports against industry labor intensity and productivity.

Fact 2: *Firms' export product scope increases with industry labor intensity and productivity.*

A firm's export product scope is defined as the number of distinct HS 8-digit products exported to all destinations. ASIF firms unmatched with CCTS are assumed to have zero export product scope. Panel B of Table 1 examines how firms' export product scope varies with industry labor intensity and productivity. In columns (1)-(4), the outcome is a dummy for multi-product exporters, defined as firms with export product scope greater than one. Columns (5)-(8) directly analyze the export product scope, using Poisson pseudo-maximum likelihood (PPML) to account for zeros in the outcome variable and potential heteroskedasticity (Silva and Tenreyro, 2006). Results indicate that firms in labor-intensive and high-productivity industries are more likely to be multi-product exporters and export more products than firms in capital-intensive and low-productivity industries. These findings hold for the full ASIF and matched ASIF-CCTS samples, with and without controls for firm age, size, ownership, location, and processing trade engagement.

Fact 3: *The skewness of firms' export product mix decreases with industry labor intensity and productivity.*

A firm's product mix refers to its sales distribution across its products. In Panel C of Table 1, we measure the skewness of the export product mix using two indicators. Columns (1)-(4) adopt a local measure: the export sales ratio of the second best-selling to the first best-selling ("core") export products. For single-product exporters, this measure is zero. A smaller ratio indicates the exporter skews sales more toward the core product. Columns (5)-(8) use the Hirschman-Herfindahl Index (HHI), which is a global measure of sales concentration across all exported products. A higher HHI corresponds to greater skewness. The table shows that firms in labor-intensive and high-productivity industries systematically exhibit a larger sales ratio of their second to first best-selling products. Additionally, their HHI is lower, indicating less skewed export product mixes. These patterns hold with and without controls for firm age, size, ownership, location, and processing trade engagement.

2.3 Discussion

A first aspect worth deeper scrutiny concerns the validity of the stylized facts over the full sample of 2000-2006. This is confirmed as we find that all three persist over time. See appendix Table B1 for the first fact, and columns (1), (3), (5), and (6) of Table B2 for the second and third facts.

Second, a non-trivial fraction of the unmatched ASIF firms with CCTS reported positive total exports. This is probably because some firms export indirectly via trade intermediaries (Ahn et al., 2011; Bai et al., 2017). We have assumed that these firms have zero export product scope. Dropping them altogether strengthens the positive correlations of industry labor intensity/productivity with export product scope (see columns (2) and (4) of Table B2), suggesting that our assumption introduced a downward bias in the estimates.

Finally, the third stylized fact might be sensitive to the way we measure skewness. As alternative measures, we try the sales ratio of the third best-selling product to the core one, the sales of the third to the second best-selling product, and the core product's export share. Using these alternative measures, we continue to find that the export product skewness falls with industry labor intensity and productivity (see columns (7)-(9) in Table B2).

3 Theory

To rationalize the stylized facts highlighted in the previous section, a model is needed in which relative productivity and factor intensity determine the industry variation of firms' export propensity, export intensity, product scope, and product mix. Such a model is not available off the shelf from the literature. Rationalizing the behavior of the product mix is particularly challenging as it requires to relax the widespread assumption of constant elasticity of substitution (CES), which implies that the relative sales of different products within an industry depend only on their relative productivity (see, e.g., Bernard et al., 2007 and 2011). It is, therefore, important that the model to be used is flexible enough to allow for variable elasticity of substitution (VES), while leaving the final assessment of the practical relevance of the departure from CES to the quantification exercise.

The model we propose builds on the setups of Dornbusch et al. (1977, 1980) and Mayer et al. (2014). In Dornbusch et al. (1977), the economy consists of two countries differ in relative productivity across a continuum of perfectly competitive industries. There is only one factor (labor), which is freely mobile between industries within the same country and immobile between countries. Within countries, all industries, therefore, pay the same remuneration to that factor, and a country pays a higher remuneration if it is more productive than the other country in all sectors. In this situation, the country has a "Ricardian absolute advantage." International trade allows a country to specialize in producing and exporting goods supplied by industries whose relative productivity is higher than in the other country. These are its industries of "Ricardian comparative advantage." In contrast, in Dornbusch et al. (1980), there are two factors (capital and labor), and in autarky the two countries have different relative factor prices due to different relative factor endowments. In addition, industries have different factor intensities. So trade leads a country to specialize in producing and exporting goods supplied by industries that are relatively intensive in its relatively abundant factor. These are its industries of "Heckscher-Ohlin comparative advantage." Each factor is paid the same remuneration in all the industries of a given country due to free internal factor mobility across them, and the same remuneration in both countries unless these are fully specialized in different industries or face trade frictions.

We merge Dornbusch et al. (1977) with Dornbusch et al. (1980), and enrich them in terms of market structure, allowing for monopolistic competition among multi-product firms of heterogeneous productivity as in Mayer et al. (2014). However, while in Mayer et al. (2014) demand is linear, and there are no income effects, we allow for these effects and adopt a more flexible demand system embedding CES and linear demands as special cases, whose relative relevance can then be assessed empirically. Finally, the introduction of empirically relevant trade frictions prevents international factor price equalization above and beyond the implications of Ricardian absolute advantage.

In addition to rationalizing the stylized facts described in the previous section, the resulting model makes novel predictions about the interactions among firm heterogeneity, industry productivity, and factor intensity, and provides a gateway to assessing a country's Ricardian comparative advantage from available data through sufficient statistics.

We start the model's exposition in a closed economy as the autarkic equilibrium is needed for the definition of Ricardian comparative advantage.

3.1 Closed Economy

Suppose there are two countries, Home (China) and Foreign (Rest of the World). The consumers in each country have identical preferences and derive their utility from consuming the output of a continuum of industries, indexed by $z \in [0, 1]$, each supplying varieties of its own horizontally differentiated good. We focus on Home as symmetric results can be readily applied to the foreign country.

Preferences and Demand

In Home there are L identical consumers with individual utility given by

$$U = \int_0^1 b(z) \ln \left(\alpha \int_{i \in \Omega(z)} q_i^c(z) di - \frac{\gamma}{1-\delta} \int_{i \in \Omega(z)} \left(q_i^c(z)\right)^{1-\delta} di \right) dz, \tag{1}$$

where $b(z) > 0$ is the expenditure share of consumption on goods from industry z satisfying $\int_0^1 b(z)dz = 1$, $q_i^c(z)$ denotes the consumption of the differentiated variety i in industry z, and $\Omega(z)$ is the set of differentiated varieties in industry z. The industry sub-utility combines a linear component (with coefficient $\alpha \geq 0$) whereby only total consumption matters, with a CES component (with coefficient $-\gamma/(1-\delta)$) whereby also the dispersion of total consumption across varieties matters. While the importance of the CES component is regulated by γ, the 'love of variety' it embeds is measured by δ. Varieties are perfect substitutes in two extreme cases: when γ goes to zero, there is no CES component; when δ goes to zero, the CES component's elasticity of substitution $(1/\delta)$ limits infinity.

The budget constraint faced by a Home consumer is given by

$$\int_0^1 \int_{i \in \Omega(z)} p_i^c(z) q_i^c(z) di dz = I,$$

where I is income. Solving the consumer's utility maximization problem delivers the following inverse individual demand for the differentiated variety i in industry z:[13]

$$p_i(z) = p_{max}(z) - \frac{\gamma}{\lambda(z)} q_i^c(z)^{-\delta}, \tag{2}$$

[13] Appendix A2 provides details about the solutions of the consumers' and firms' optimization problems.

where $p_{max}(z) = \alpha/\lambda(z)$ is the endogenous "choke price" , below which $p_i(z)$ has to fall for quantity demanded to be positive, and $\lambda(z)$ is proportional to the marginal utility of income.[14]

Aggregating (2) determines the market demand for variety i in industry z as

$$q_i(z) = L\left(\frac{\gamma}{\lambda(z)}\right)^{\frac{1}{\delta}} (p_{max}(z) - p_i(z))^{-\frac{1}{\delta}}, \tag{3}$$

given that L is the number of identical consumers ("market size"). For $\alpha = 0$, $\delta > 0$ and $\gamma < 0$, demand (3) is CES; for $\delta = -1$ and $\gamma > 0$, it is linear.[15] In general, it can be viewed as a mixture of CES and linear demands. This can be usefully seen in terms of the demand elasticity

$$\varepsilon_{q_i(z)} = \frac{1}{\delta}\frac{p_i(z)}{p_{max}(z) - p_i(z)},$$

where the first factor $(1/\delta)$ on the right-hand side is the CES demand elasticity component, whereas the second is the linear demand one. As $|\delta|$ increases, the demand becomes less elastic. In particular, as in the CES case, if δ goes to infinity, then $\varepsilon_{q_i(z)}$ goes to zero, and the demand is perfectly inelastic; if δ goes to zero, then $\varepsilon_{q_i(z)}$ goes to infinity and the demand is perfectly elastic. On the other hand, as in the linear case, the demand elasticity is a decreasing function of the choke price $p_{max}(z)$, and firms face different demand elasticity depending on their prices. Higher prices are associated with higher elasticity, which in the limit becomes infinite for a price equal to $p_{max}(z)$. In what follows, we will focus on the case where $\alpha > 0$, $\gamma > 0$, and $\delta < 0$ hold, which is neither a CES nor a linear demand, and allow the data to tell the relative importance of the two components.[16]

Production and Firm Behaviour

Production requires labor and capital as factors. Factor markets are perfectly competitive, and factor shares vary across industries, with firms in the same industry sharing the same factor proportions as in Dornbusch et al. (1980). Each industry employs its own composite input with a production function that is homogeneous of degree one, and we use $\omega(z)$ to denote the exact price index of the composite input of industry z, which we call the industry's "unit input price."

Firms differ in terms of total factor productivity (TFP), and we characterize their heterogeneity by its inverse, which we refer as "unit input requirement" (UIR) as it measures the amount of composite input a firm needs to produce a unit of output. The product market structure is monopolistically competitive. Firms can produce multiple varieties starting from a "core competency," which is the variety a firm supplies at the lowest UIR. Other varieties are variants of the core competency. Their production requires additional amounts of the composite input that increase with their difference ("distance")

[14] Bulow and Pfleiderer (1983) study a similar inverse demand without an endogenous choke price and income effect.

[15] The choke price here depends on the income effect. This dependence is absent in Mayer et al. (2014) due to the inclusion of an outside good that enters the utility function linearly. On the other hand, their utility function features an additional cross-variety effect that makes firms interact despite the absence of the income effect.

[16] See Mrázová and Neary (2017) for a detailed discussion of the families of demand systems generating variable markup while nesting CES demand as a special case.

from the core competency. Specifically, varieties are ranked in increasing order of distance from the core competency and indexed by m. Variety m's UIR is an increasing function of its distance from the core competency $\nu(m,c) = \xi^{-m}c$, with $\xi \in (0,1)$ and $m = 0$ corresponding to the core competency with UIR c.[17] The parameter ξ captures the flexibility of the multi-product technology. The larger ξ, the easier it is for firms to expand their product scope. As ξ tends to zero, firms become single-product firms as the UIR of additional varieties other than the core competency approaches infinity.

To maximize profits, a firm solves the following problem for each variety v it considers producing:

$$\max_{p(z,v)} (p(z,v) - \omega(z)\nu)q(z,v),$$

with demand $q(z,v)$ satisfying (3). Solving the problem gives variety v's profit-maximizing price

$$p(z,\nu) = \frac{-\delta}{1-\delta} p_{max}(z) + \frac{1}{1-\delta}\omega(z)\nu. \tag{4}$$

Equation (4) highlights the implications of allowing for the demand mixture. If $\delta = -1$, the price is a simple arithmetic average of the choke price and the marginal cost as in the linear demand system used by Mayer et al. (2014). Otherwise, the price is a weighted arithmetic average of the choke price and the marginal cost with weights determined by $\delta < 0$. The smaller $|\delta|$, the smaller the weight of the choke price relative to the marginal cost. As in the CES case, if δ goes to infinity, demand is perfectly inelastic with firms charging the highest possible price $p_{max}(z)$ and markup $p_{\max}(z) - \omega(z)\nu$. If δ goes to zero, demand is perfectly elastic, and firms price at the marginal cost $\omega(z)\nu$ charging zero markup. On the other hand, as in the linear case, a higher price $p(z,\nu)$ is associated with higher elasticity, and thus with lower markup. Moreover, the price and the markup are decreasing functions of $p_{max}(z)$ as in the linear case. They are increasing and decreasing functions of the marginal cost $\omega(z)\nu$, respectively. In the limit, a marginal cost equal to the choke price commands zero markup.

Firm Entry and Domestic Selection

Firms discover the UIR of their core competency upon entering the market. Entry is costly as firms need to pay upfront a sunk cost for product development. Each industry has a large pool of potential entrants. Entrants in industry z pay a common sunk cost $\omega(z)f_E$ and draw their core competency's UIR from an inverse Pareto distribution with c.d.f.

$$G(z,c) = \left(\frac{c}{C_M(z)}\right)^k, \; c \in [0, \, C_M(z)], \tag{5}$$

where the support's upper bound $C_M(z)$ captures the industry's "state of technology" (Eaton and Kortum, 2002). As larger $C_M(z)$ implies higher probability of high UIR draws, it indicates a worse state of technology.[18]

[17] Eckel and Neary (2010) term this modeling approach for cost-side product asymmetries "flexible manufacturing."

[18] A Pareto distribution with a given shape parameter first-order stochastically dominates another Pareto distribution with the same shape parameter but a larger upper bound of the support. The assumption that the shape parameter k is

There is firm and product selection. Firms drawing c larger than the cutoff UIR $C_D(z) = p_{\max}(z)/\omega(z)$ exit the market without producing as they cannot generate enough revenue to cover their marginal cost with any of their products. Firms drawing $c \leq C_D(z)$ supply only varieties with UIR $\nu(m,c) \leq C_D(z)$. Then, as firms enter under a veil of ignorance about their c, free entry implies that the expected profit must match the sunk entry cost:

$$\int_0^{C_D(z)} \Pi_D(z,c)dG(z,c) = \omega(z)f_E, \tag{6}$$

where $\Pi_D(z,c) = \sum_{m=0}^{M_D(z,c)-1} \pi_D(z,\nu(m,c))$ is the total profit from all varieties that a firm with core competency c produces, $\pi_D(z,v(m,c))$ is the firm's maximized profit from variety m, and $M_D(z,c)$ is its "product scope," which is the number of varieties it produces and satisfies:

$$M_D(z,c) = \begin{cases} 0 & \text{if } c > C_D(z), \\ \max\{m \mid \xi^{-m}c \leq C_D(z)\} + 1 & \text{if } c \leq C_D(z). \end{cases} \tag{7}$$

Evaluating the free entry condition (6) under assumption (5) yields the autarky cutoff UIR

$$C_D^A(z) = \left[C_M(z)^k \frac{f_E}{\beta \Psi L} \left(\frac{\alpha}{\gamma} \right)^{\frac{1}{\delta}} \right]^{\frac{1}{k+1}}, \tag{8}$$

where $\Psi = (1-\xi^k)^{-1}$ is a bundling parameter reflecting multi-product flexibility, with larger Ψ associated with more flexibility; $\beta = -\delta k B(k, 2-\frac{1}{\delta})/(1-\delta)^{1-\frac{1}{\delta}}$ is a constant, with $B(\cdot)$ being the Beta function.[19] Note that the unit input price does not appear in expression (8).[20]

The Foreign is assumed to differ from Home in three dimensions only: its state of technology, unit input prices, and market size. Hence, by symmetry, Foreign autarky cutoff UIR evaluates to

$$C_D^{A*}(z) = \left[C_M^*(z)^k \frac{f_E}{\beta \Psi L^*} \left(\frac{\alpha}{\gamma} \right)^{\frac{1}{\delta}} \right]^{\frac{1}{k+1}}, \tag{9}$$

where the asterisk is used henceforth to label Foreign parameters and variables. Expressions (8) and (9) show that selection is tougher (i.e. a smaller cutoff) in the country with a better state of technology and larger market size.

common across industries is made to simplify the quantitative exercise that we will run as proof of concept. In this respect, allowing the shape parameter to vary across industries would complicate the analysis without adding much insight. See Ottaviano and Suverato (2023) for a model with linear demand where k varies across industries.

[19]The Beta function is defined as $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ for $x > 0$ and $y > 0$. It is related to the Gamma function $\Gamma(\cdot)$ in Eaton and Kortum (2002) and satisfies $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$.

[20]This is because we assume the fixed cost and variable cost share the same factor proportion. This assumption is also adopted by Romalis (2004) and Bernard et al. (2007).

3.1.1 Ricardian Comparative Advantage

As Dornbusch et al. (1977), we define Ricardian comparative advantage in terms of relative industry UIR or equivalently in terms of relative inverse industry TFP. In this respect, the presence of firm heterogeneity requires averaging across all firms within each industry. We consider quantity-based TFP (TFPQ), defined as industry output per unit of composite input. This is the inverse of the amount of industry composite input per unit of output (i.e. output-weighted average firm UIR) and is thus consistent with our use of UIR. Under autarky, Home industry z's TFPQ is given by

$$\overline{\Phi}^A(z) = \frac{\int_0^{C_D^A(z)} Q(z,c)dG(z,c)/G(C_D^A(z))}{\int_0^{C_D^A(z)} T(z,c)dG(z,c)/G(C_D^A(z))}, \tag{10}$$

where $Q(z,c) = \sum_{m=0}^{M_D(z,c)-1} q(z, v(m,c))$ and $T(z,c) = \sum_{m=0}^{M_D(z,c)-1} v(m,c)q(z,v(m,c))$ are total output and the total variable amount of composite input respectively for a firm with a core competency UIR c producing $M_D(z,c)$ varieties, which are then aggregated across all active firms for $c \in [0, C_D^A(z)]$. For comparison, we also consider the unweighted average firm UIR given by

$$\overline{c}^A(z) = \int_0^{C_D^A(z)} \sum_{m=0}^{M_D(z,c)-1} \nu(m,c)dG(z,c)/G(C_D^A(z)). \tag{11}$$

The following lemma discusses their relationship.

Lemma 1. ***(Autarkic Average Firm UIR)*** *Under autarky, Home industry z's TFPQ, or the inverse of output-weighted average firm UIR, satisfies*

$$\overline{\Phi}^A(z) = \frac{k+1-1/\delta}{k}\frac{1}{C_D^A(z)}, \tag{12}$$

while the unweighted firm UIR satisfies

$$\overline{c}^A(z) = \frac{k}{k+1}\Psi C_D^A(z). \tag{13}$$

Proof. See Appendix A1.1. □

Analogously, for Foreign country, we have $\overline{\Phi}^{A*}(z) = \frac{k+1-1/\delta}{k}\frac{1}{C_D^{A*}(z)}$ and $\overline{c}^{A*}(z) = \frac{k}{k+1}\Psi C_D^{A*}(z)$. Clearly, the unweighted and weighted average firm UIR differ only by a constant. Therefore, weighting is immaterial for the following definition of Ricardian comparative advantage.

Definition 1. ***(Ricardian Comparative Advantage)*** *Home has a "Ricardian comparative advantage" in industry z with respect to industry z' whenever, compared with Foreign, its autarky average firm UIR is relatively lower in z, i.e., $\frac{C_D^A(z)}{C_D^A(z')} < \frac{C_D^{A*}(z)}{C_D^{A*}(z')}$, or equivalently, its autarky average TFPQ is relatively higher, i.e., $\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^A(z')} > \frac{\overline{\Phi}^{A*}(z)}{\overline{\Phi}^{A*}(z')}$.*

Given Definition 1, comparing (8) with (9) highlights that, whereas Ricardian absolute advantage depends on the countries' states of technology as well as on market sizes, only the different states of technology matter for Ricardian comparative advantage.[21] Specifically, we can write:

Proposition 1. ***(Ricardian Comparative Advantage and State of Technology)*** *Home has a "Ricardian comparative advantage" in industry z with respect to industry z' if and only if, compared with Foreign, in that industry its state of technology is relatively better, i.e.,* $\frac{C_M(z)}{C_M(z')} < \frac{C_M^*(z)}{C_M^*(z')}$*. However, the Ricardian comparative advantage does not fully reflect the gap in the state of technology; and the more so, the higher the degree of firm heterogeneity as inversely measured by k.*

Proof. By inspection of (8) with (9). □

To understand why a higher degree of firm heterogeneity dampens the transmission of the relative state of technology to Ricardian comparative advantage, we first note that when Home has a Ricardian comparative advantage in industry z with respect to z', (8) with (9) imply that the following holds: $\frac{C_M(z)/C_M(z')}{C_M^*(z)/C_M^*(z')} < \left(\frac{C_M(z)/C_M(z')}{C_M^*(z)/C_M^*(z')}\right)^{\frac{k}{k+1}} = \frac{C_D^A(z)/C_D^A(z')}{C_D^{A*}(z)/C_D^{A*}(z')} < 1$. To interpret the exponent $\frac{k}{k+1}$, we can examine (8), which implies that the passthrough from $C_M(z)$ to $C_D^A(z)$ is incomplete as we have

$$\frac{\partial \ln C_D^A(z)}{\partial \ln C_M(z)} = \frac{k}{k+1} < 1, \tag{14}$$

with the incompleteness regulated by firm heterogeneity: the more heterogeneous firms are (i.e. the smaller k is), the more incomplete the passthrough is (i.e. the smaller $\frac{k}{k+1}$ is). The reason is that a firm's passthrough from its products' marginal costs to prices is an increasing function of the firm's core competency UIR. When k is smaller, there is a larger density of firms with lower core competency UIR in the industry and thus with more incomplete passthrough. This is an important implication of VES demand and variable markup. In contrast, all firms and thus the industry would feature complete passthrough with CES demand and constant markup.

3.2 Open Economy

Ricardian comparative advantage is defined in autarky, which is typically a counterfactual situation, as in our case. Can we elicit it from data available in the factual situation where countries trade? To answer this question, we open up the economy and assume that all industries $z \in [0, 1]$ are active in both countries, so that specialization is incomplete as this is the empirically relevant scenario for us. We further assume that firms face an iceberg trade friction common to all industries: $\tau > 1$ units of output have to be shipped for a unit to reach the destination.[22] Finally, in the quantitative exercise we will lift factor prices from the data. Hence, we take them here as given under the assumption that the countries' relative factor endowments are such that they make the observed factor prices aligned with the model's predictions.

[21] Home has a "Ricardian absolute advantage" in industry z whenever it exhibits tougher selection ($C_D^A(z) < C_D^{A*}(z)$).

[22] This assumption is made to simplify the quantitative exercise we run for proof of concept. Allowing the iceberg friction to vary across industries would complicate the analysis a lot, without adding much insight.

3.2.1 Export Selection

The delivered marginal cost of a Home firm with core competency UIR c in industry z selling its product m abroad is $\tau\omega(z)\nu(m,c)$. Given demand (3), only varieties with UIR $v(m,c)$ below the export cutoff UIR $C_X(z) = \frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)}$ generate positive sales abroad, and only firms with core competency UIR c below that threshold can profitably export at all. Therefore, the number of varieties a Home firm exports to Foreign is

$$M_X(z,c) = \begin{cases} 0 & \text{if } c > C_X(z), \\ \max\{m|\tau\omega(z)v(m,c) \leq \omega^*(z)C_D^*(z)\} + 1 & \text{if } c \leq C_X(z), \end{cases} \tag{15}$$

which is the "export product scope," whereas the domestic product scope is still determined by (7).

Free entry implies that the sum of expected profits from the domestic and export markets equals the sunk entry cost. The free entry condition can be written as

$$\int_0^{C_D(z)} \Pi_D(z, v(m,c))dG(z,c) + \int_0^{C_X(z)} \Pi_X(z, v(m,c))dG(z,c) = \omega(z)f_E, \tag{16}$$

where a firm's profits from domestic sales and exports are the sum of the profits from all varieties sold by the firm in the Home and Foreign markets: $\Pi_D(z, v(m,c)) = \sum_{m=0}^{M_D(z,c)-1} \pi_D(z, v(m,c))$ and $\Pi_X(z, v(m,c)) = \sum_{m=0}^{M_X(z,c)-1} \pi_X(z, v(m,c))$, with $\pi_D(z, v(m,c))$ and $\pi_X(z, v(m,c))$ being the profits the firm earns from selling variety m domestically and abroad, respectively. Under the distributional assumption (5), condition (16) evaluates to

$$LC_D(z)^{k+1} + \rho\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} L^*C_D^*(z)^{k+1} = \frac{f_E}{\beta}\left(\frac{\alpha}{\gamma}\right)^{1/\delta} \frac{C_M(z)^k}{\Psi}, \tag{17}$$

where $\rho = \tau^{-k}$ is an index of trade freeness ranging between 0 (autarky) and 1 (free trade). There is an analogous equation for Foreign. We solve the two equations by Cramer's rule and find the open-economy domestic cutoff UIRs in the two countries:

$$C_D(z) = C_D^A(z)\left[\frac{1-\rho\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\left(\frac{C_M^*(z)}{C_M(z)}\right)^k}{1-\rho^2}\right]^{\frac{1}{k+1}},$$

$$C_D^*(z) = C_D^{A*}(z)\left[\frac{1-\rho\left(\frac{\omega(z)}{\omega^*(z)}\right)^{k+1}\left(\frac{C_M(z)}{C_M^*(z)}\right)^k}{1-\rho^2}\right]^{\frac{1}{k+1}}. \tag{18}$$

These expressions highlight the role of relative unit input price, $\frac{\omega(z)}{\omega^*(z)}$, and relative state of technology, $\frac{C_M(z)}{C_M^*(z)}$, for firm entry and domestic selection. There is entry in both countries whenever both cutoffs are positive, with the necessary and sufficient condition given by

$$\rho < \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\left(\frac{C_M^*(z)}{C_M(z)}\right)^k < \frac{1}{\rho},$$

which is also necessary and sufficient for both countries to experience tougher selection than in autarky ($C_D(z) < C_D^A(z)$ and $C_D^*(z) < C_D^{A*}(z)$). Hence, as long as there is entry in both countries, international specialization is incomplete, and domestic selection is tougher with trade than in autarky.

To better understand the implication of (18) for selection, for the sake of argument, consider an initial situation in which $(\frac{\omega^*(z)}{\omega(z)})^{k+1}(\frac{C_M^*(z)}{C_M(z)})^k = 1$ holds so that trade changes autarkic cutoffs by the same factor in both countries. Let Home's unit input price $\omega(z)$ fall and the state of technology improve (i.e. $C_M(z)$ decrease), which fosters entry and domestic selection in Home as $C_D(z)$ falls but hampers entry and domestic selection in Foreign as $C_D^*(z)$ rises. In contrast, exports become less selective in Home as $C_X(z) = \frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)}$ rises. Hence, Home varieties sold in domestic markets are more efficiently produced than before, while the opposite happens to those sold in Foreign markets.

3.2.2 Matching the Stylized Facts

So far, we have put very few restrictions on model fundamentals. We now provide conditions that the fundamentals should satisfy to match the stylized facts presented in Section 2.

Our model implies that Home export propensity is the ex-ante probability of domestic firms to export $\chi(z) = \left(\frac{C_X(z)}{C_D(z)}\right)^k$, which is also the ex-post share of firms that export. Without loss of generality, we rank domestic industries in increasing order of $\chi(z)$, just like Dornbusch et al. (1977, 1980) rank them according to the strength of comparative advantage. We first examine the cross-industry relation between Home export propensity and export intensity (i.e. the share of sales exported, $\theta(z)$):

Lemma 2. ***(Export Propensity and Intensity)*** *Home export propensity $\chi(z)$ increases with z if and only if Home export intensity $\theta(z)$ increases with z for $z \in [0,1]$:*

$$\frac{\partial\theta(z)}{\partial z} \geq 0 \Leftrightarrow \frac{\partial\chi(z)}{\partial z} \geq 0.$$

Proof. See Appendix A1.2. □

Therefore, industry rankings by export propensity and export intensity are the same. This is consistent with the first stylized fact that China's export intensity tends to be higher in industries with higher export propensity. Next, to fully rationalize that fact, we look at the relation of Home export propensity with Ricardian comparative advantage and relative unit input price across industries, which we summarize by:

Proposition 2. ***(Export Propensity, Ricardian Comparative Advantage and Input Prices)*** *The cross-industry relation of Home export propensity $\chi(z)$ with its Ricardian comparative advantage ($C_M^*(z)/C_M(z)$) and relative unit input prices ($\omega(z)/\omega^*(z)$) is such that:*
(a)

$$\frac{\partial\chi(z)}{\partial z} \geq 0 \Leftrightarrow \frac{\partial\left[\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\left(\frac{C_M^*(z)}{C_M(z)}\right)^k\right]}{\partial z} \geq 0,$$

or equivalently (b)

$$\frac{\partial \chi(z)}{\partial z} \geq 0 \Leftrightarrow \varepsilon_\omega(z) + \frac{k}{k+1}\varepsilon_c(z) \geq 0,$$

where $\varepsilon_\omega \equiv \frac{\partial \ln(\omega^*(z)/\omega(z))}{\partial \ln(z)}$ *and* $\varepsilon_c \equiv \frac{\partial \ln\left(C_M^*(z)/C_M(z)\right)}{\partial \ln(z)}$ *are the elasticities of relative unit input price and relative state of technology to the industry index* z*.*

Proof. See Appendix A1.3. □

Given Lemma 2, Proposition 2 implies that, for the model's predictions to fully match the first stylized fact that export propensity increases with both industry labor intensity and industry productivity, we need both $\omega^*(z)/\omega(z)$ and $C_M^*(z)/C_M(z)$ are increasing function of z: China has better relative state of technology and lower relative input prices in higher z industries. Hence, based on the above discussion of (18), industries with higher export propensity exhibit relatively tougher selection in the domestic market than in the export market. Another implication is that export propensity is not necessarily aligned with the Ricardian comparative advantage in the presence of Heckscher-Ohlin forces. The more so the more heterogeneous firms are (i.e. the smaller k and thus $\frac{k}{k+1}$ are) as more heterogeneity reduces the passthrough from the relative state of technology to comparative advantage.

For the model to be also in line with the other stylized facts, we need to impose additional restrictions. Specifically, with respect to the second stylized fact, we have to discipline the way the export cutoff $C_X(z)$ varies across industries as the following holds:

Proposition 3. ***(Export Propensity and Product Scope)*** *Firms' export product scope increases weakly with* z *if and only if*

$$\varepsilon_{C_X} \geq 0, \tag{19}$$

where $\varepsilon_{C_X} \equiv \frac{\partial \ln C_X(z)}{\partial \ln(z)}$ *is the elasticity of the export UIR cutoff to the industry index* z*.*

Proof. See Appendix A1.4. □

This proposition implies that export selection is less intense in Home industries with higher export propensity as exporters find it easier to expand their export product scope when competition in the export market is softer. Hence, based on (18), comparing two industries z' and z with $z' > z$ such that $\frac{\omega^*(z)}{\omega(z)} < \frac{\omega^*(z')}{\omega(z')}$ and $\frac{C_M^*(z)}{C_M(z)} < \frac{C_M^*(z')}{C_M(z')}$, the former industry with higher export propensity is characterized by softer selection in the export market. Given $C_X(z) = \frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)}$, (9) and (18), a sufficient condition for Home firms' export product scope to increase with industry labor intensity and productivity (the second stylized fact) is that not only $\omega^*(z)/\omega(z)$ and $C_M^*(z)/C_M(z)$, but also $C_M^*(z)$ rise with z: Foreign has worse state of technology in industries where Home has higher export propensity than in those where Home has a lower export propensity. This is not, however, necessary when Home relative unit input prices fall fast enough or its Ricardian comparative advantage rises fast enough with z.

As for the third stylized fact, we have to discipline firms' ability to be multi-product. In this respect, we can state the following:

Proposition 4. ***(Export Propensity and Product Mix)*** *As long as* $\varepsilon_{C_X} \geq 0$ *and*

$$(1+\xi)\chi(z)^{1/k} \geq 1 + \frac{1}{\delta}, \tag{20}$$

firms' export revenues from the first best-selling variety relative to those from the second best-selling variety decreases weakly with z *for all* $z \in [0, 1]$.

Proof. See Appendix A1.5. □

Taken together with the first, the third stylized fact implies that the export product mix is less skewed in industries of higher export propensity. Proposition 4 provides sufficient conditions for this to hold. We first need export selection to be more lenient in industries of higher export propensity, as granted by $\varepsilon_{C_X} \geq 0$. In addition, we need constraints on the supply and demand parameters ξ and δ appearing in (20). Since $\chi(z) \geq 0$, inequality (20) holds for any $\xi \in (0, 1)$ when $-1 \leq \delta < 0$. However, as δ decreases further, the demand becomes less elastic, making it more difficult for firms to generate additional revenues from varieties away from the core competency. To counteract this effect, larger multi-product flexibility ξ is needed for (20) to hold. The larger ξ is, the easier it is for a firm to proliferate its varieties, facilitating the reallocation of sales away from the core competency. Finally, given ξ and δ, as the export propensity $\chi(z)$ increases with z, inequality (20) is more likely to hold in Home industries of lower relative input prices or better relative state of technology.

3.2.3 Relative Industry Productivity

As in the closed economy, we define industry TFPQ as industry output per unit of composite input. For Home industry z's, TFPQ is thus given by

$$\overline{\Phi}(z) = \frac{\int_0^{C_D(z)} Q_D(z,c)dG(z,c)/G(C_D(z)) + \int_0^{C_X(z)} Q_X(z,c)dG(z,c)/G(C_D(z))}{\int_0^{C_D(z)} T_D(z,c)dG(z,c)/G(C_D(z)) + \int_0^{C_X(z)} T_X(z,c)dG(z,c)/G(C_D(z))},$$

where we distinguish between the amounts of output sold in the domestic market (D) and in the foreign market (X) as well as between the associated amounts of composite input. Under the distributional assumption (5), the open-economy TFPQ of industry z is characterized by:

Lemma 3. ***(Open-economy TFPQ)*** *The open-economy TFPQ of Home industry* z *is*

$$\overline{\Phi}(z) = \left(\eta_D(z)\frac{1}{C_D(z)} + \eta_X(z)\frac{1}{\tau C_X(z)}\right)\frac{k+1-1/\delta}{k}, \tag{21}$$

where $\eta_D(z)$ *and* $\eta_X(z)$ *are the shares of the total amount of composite input embodied in output sold domestically and exported, respectively, with* $\eta_D(z) = \frac{LC_D(z)^{k+1}}{LC_D(z)^{k+1} + L^*\tau C_X(z)^{k+1}}$ *and* $\eta_X(z) = 1 - \eta_D(z)$.

Proof. See Appendix A1.6. □

This lemma highlights the channels whereby a Home industry's TFPQ with trade deviates from its autarkic TFPQ, on which Definition 1 is based. Intuitively, $\overline{\Phi}(z)$ is a weighted average of the industry's inverse UIRs in the domestic and export markets ($1/C_D(z)$ and $1/C_X(z)$ respectively), with weights given by the composite input shares of output sold in the two markets ($\eta_D(z)$ and $\eta_X(z)$ respectively). We have already discussed, with reference to Proposition 3, how industries with higher export propensity exhibit softer selection in the export market, which makes $C_X(z)$ relatively larger in higher z industries. In this respect, with trade, there is an extensive margin adjustment on industry productivity that tends to make $\overline{\Phi}(z)$ decrease with export propensity through weaker export selection. In addition, it is easy to verify that industries with higher export propensity exhibit larger export shares of their composite inputs: $\partial\eta_D(z)/\partial\chi(z) < 0$ and $\partial\eta_X(z)/\partial\chi(z) > 0$.[23] This implies that, since exporters are more productive than non-exporters, there is also an intensive margin adjustment that tends to raise $\overline{\Phi}(z)$. We elaborate on these points in the following:

Proposition 5. ***(Relative TFPQ Decomposition)*** *In the open economy, Home industry z's relative TFPQ can be decomposed as the product of an ex-ante component before trade and two ex-post components after trade:*

$$\frac{\overline{\Phi}(z)}{\overline{\Phi}^*(z)} = \underbrace{\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}}_{ex\text{-}ante} \underbrace{\left(\frac{\frac{L}{L^*}\rho + \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L^*}{L}\rho + \left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}} \right)^{\frac{k}{k+1}}}_{ex\text{-}post\ amplifying\ (XPA(z))} \underbrace{\frac{\rho\chi(z)^{-1} + \frac{L^*}{L}\rho}{\rho^{-1}\chi(z) + \frac{L}{L^*}\rho}}_{ex\text{-}post\ dampening\ (XPD(z))}, \tag{22}$$

with the ex-post amplifying (dampening) component increasing (decreasing) with $\chi(z)$.

Proof. See Appendix A1.7. □

To understand how Proposition 5 relates to the results in Bernard et al. (2007), it is useful to compute the Home and Foreign industry unweighted average TFPQs, $\overline{c}(z)^{-1}$ and $\overline{c}^*(z)^{-1}$, and take their ratio.[24] This leads to:

Proposition 6. ***(Relative Unweighted Average Productivity)*** *In the open economy, Home to Foreign industry z's unweighted average TFPQ can be decomposed as the product of an ex-ante component before trade and an ex-post amplifying component after trade*

$$\frac{\overline{c}(z)^{-1}}{\overline{c}^*(z)^{-1}} = \frac{C_D^*(z)}{C_D(z)} = \underbrace{\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}}_{ex\text{-}ante} \underbrace{\left(\frac{\frac{L}{L^*} + \rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L^*}{L} + \rho\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}} \right)^{\frac{1}{k+1}}}_{ex\text{-}post\ amplifying (UXPA(z))}, \tag{23}$$

with the ex-post amplifying component increasing with $\chi(z)$.

[23] Note that $\eta_D(z)$ and $\eta_X(z)$ can both be expressed in terms of $\chi(z)$.

[24] $\overline{c}(z)$ and $\overline{c}^*(z)$ are defined by replacing $C_D^A(z)$ in Equation (11) with $C_D(z)$ and $C_D^*(z)$, respectively.

Proof. See Appendix A1.8. □

Intuitively, as Bernard et al. (2007) point out, the higher expected export profits in comparative advantage industries induce tougher entry selection than in comparative disadvantage industries. This tends to enlarge the relative productivity differences across industries and amplify comparative advantage. Our model preserves this channel, as revealed by the ex-post component in (23).

A few remarks are in order. First, the above two propositions reveal that the ex-post amplifying and dampening components are related to trade-induced adjustments at the extensive and intensive margins, respectively, as the latter is contemplated in (22) but not in (23). Second, the origin of the dampening component can be understood by considering the implication of propositions 2 - 4 on industry TFPQ. In particular, in industries with Ricardian comparative advantage and lower relative unit input prices, a larger fraction of producers are exporters, which implies that relatively more low-productivity firms become exporters and expand relatively more, absorbing a larger share of the composite input. Moreover, as exporters from those industries tend to have larger export product scope and less skewed export product mix, they allocate more composite input to products they are less efficient in supplying. Such within- and between-firm reallocations dent relative industry TFPQ. Third, both firm heterogeneity and export selection are necessary to give rise to the ex-post components. In particular, if k limited infinity and the firm distribution degenerated, then the ex-post components in equations (22) and (23) would both converge to 1. Even if k were bounded but there were no export selections, i.e., $\chi(z) = 1$ so that all firms exported, then the ex-post components become constants, and relative productivity is perfectly correlated between the open economy and autarkic economy.[25]

Last but not least, the ex-post dampening effect of export selection is not specific to our VES model. For example, it also appears in the CES model of Bernard et al. (2007) as industry export productivity cutoffs are relatively closer to the domestic productivity cutoffs in comparative advantage industries.[26] If they had adopted productivity measures that contained trade-induced adjustments at the intensive margins as we do, they would have also captured the dampening aspect of export selection.[27] What is specific to the VES model is that the dampening components include an additional effect at the extensive margin due to changes in industry and firm product mixes.

3.3 Sufficient Statistics

The sufficient statistics approach allows us to express the theoretical objects of interest as functions of reduced-form elasticities and observables, thus allowing for counterfactual analysis without having to solve, calibrate, or simulate the whole general equilibrium of the model (Chetty, 2009; Costinot et al., 2012; Costinot et al., 2019). This subsection uses the sufficient statistics approach to express Ricardian comparative advantage in terms of parameters and variables that can be readily retrieved from

[25]All firms export if $C_X(z) \geq C_D(z)$, which implies $\tau \leq \frac{\omega^*(z)C_D^*(z)}{\omega(z)C_D(z)} = \frac{p_{max}^*(z)}{p_{max}(z)}$, i.e., the iceberg trade cost is small compared to the choke price of Foreign relative to Home.

[26]See Proposition 4 and Figure 2 of their paper.

[27]We formally prove it in Appendix A4.

our datasets. It also looks into the relation between Ricardian comparative advantage and Balassa's revealed comparative advantage through the lens of the model, showing how the sufficient statistics approach can be usefully applied to understand the response of the index to counterfactual relative unit input prices while keeping the relative state of technology unchanged.

Revealing Ricardian Comparative Advantage

There are several challenges if one wants to measure Ricardian comparative advantage. First, in general, Ricardian comparative advantage is not directly observable as it depends, by definition, on relative productivity in autarky, while the data concern the open economy.[28] Second, measuring relative productivity remains challenging even for the observed open economy. A practical obstacle is that researchers typically do not have simultaneous access to the data needed to estimate firm productivity across countries in a harmonized way. Third, even if they had access to the relevant data for estimating productivity in the open economy, firm selection would pose a significant challenge connecting the open economy and autarkic productivity (Costinot et al., 2012). Finally, while making restrictive assumptions on the firm productivity distribution may help, it remains challenging to estimate the distribution's primitives. For instance, under the inverse Pareto distribution assumption, Ricardian comparative advantage depends on the bounds of the support, which do not have clear empirical counterparts.

To overcome these challenges, we establish two identification results that allow us to express the different components of (22) as functions of a small number of sufficient statistics:

Proposition 7. ***(Sufficient Statistics for Ricardian Comparative Advantage)*** *(a) The Pareto shape k, trade freeness ρ, export propensity $\chi(z)$ and intensity $\theta(z)$ are sufficient statistics for the ex-post amplifying component ($XPA(z)$) and dampening component ($XPD(z)$) of industry z's relative TFPQ. (b) The sufficient statistics for the ex-ante component further include $\omega(z)/\omega^*(z)$, the relative unit input prices, as this component can be rewritten as*

$$\frac{\overline{\Phi}^{A}(z)}{\overline{\Phi}^{A*}(z)} = \left(\frac{C_M^*(z)}{C_M(z)}\right)^{\frac{k}{k+1}} \left(\frac{1-\theta(z)}{\theta(z)}\right)^{\frac{1}{k+1}} \chi(z)^{\frac{1}{k}} \rho^{-\frac{1}{k(k+1)}}, \tag{24}$$

with the relative state of technology given by

$$\frac{C_M^*(z)}{C_M(z)} = \left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{k+1}{k}} \left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right]^{\frac{1}{k}}.$$

Proof. See Appendix A1.9. □

Result (a) allows us to evaluate by how much relative TFPQ differs between the open and the closed economies as (22) implies $\frac{\overline{\Phi}(z)}{\overline{\Phi}^*(z)} / \frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = XPA(z) \cdot XPD(z)$. As for result (b), this indicates

[28] Most modern economies are far from autarky. Bernhofen and Brown (2004) investigate the sudden opening-up of Japan in the 1860s to test the law of comparative advantage (Deardorff, 1980). Their test requires information about the autarky economy and assumptions about how the economy would have behaved had Japan remained autarky.

that we must net out the influence of the relative unit input price before inferring Ricardian comparative advantage from export propensity and intensity, as they also depend on that price. In particular, the higher the observed Home relative unit input price is, the higher its unobserved autarkic relative TFPQ must be in order to generate the observed export propensity and intensity given the observed trade freeness and firm heterogeneity. Moreover, result (b) shows how we can also retrieve the unobserved relative state of technology $C_M^*(z)/C_M(z)$ from observed relative unit input prices, trade freeness, import intensity, and the Pareto shape parameter.[29]

What is remarkable about Proposition 7 is that the only piece of Foreign information needed to compute all the different components of Home's relative TFPQ is the unit input price $\omega^*(z)$. Its empirical application requires micro firm-level data in order to estimate the shape parameter of the firm UIR distribution (k) and assess the share of exporters among producers ($\chi(z)$). However, given that k can also be estimated as (the absolute value of) the trade elasticity in gravity regressions using aggregate trade data (Head and Mayer, 2014), the following corollary shows that the model's structure allows us to bypass the micro firm-level data and only rely on macro industry- and country-level data:

Corollary 1. ***(Sufficient Statistics Using Macro Data)*** *(a) Trade elasticity k, trade freeness ρ, export intensity $\theta(z)$ and relative market size L/L^* are sufficient statistics for the ex-post components of Home industry z's relative TFPQ.*
(b) The sufficient statistics for the ex-ante component further include relative unit input prices $\omega(z)/\omega^*(z)$ as this component can be equivalently rewritten as

$$\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = \frac{\omega(z)}{\omega^*(z)} \left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right]^{\frac{1}{k+1}} \left(\frac{L}{L^*}\right)^{\frac{1}{k+1}}. \tag{25}$$

Proof. See Appendix A1.10. □

Therefore, according to (b), even if we could not access firm-level data, we would still be able to reveal Home Ricardian comparative advantage by including market size as an additional Foreign piece of information. Intuitively, the higher the observed Home relative unit input price is, the higher its unobserved autarkic relative TFPQ must be in order to generate the observed export intensity given the observed trade freeness, firm heterogeneity and relative market size.

Dissecting Balassa's RCA Index

The Balassa index of revealed comparative advantage (RCA), based on the degree of countries' export specialization, is the standard tool used in the literature to identify their industries of comparative advantage. Existing works discuss the index's shortcomings, but they neglect selection in the presence of heterogeneous firms (see, e.g., Yi, 2003; French, 2017). We supplement their discussions by studying the RCA index in the context of our model, where firm heterogeneity plays a key role.

[29] In Appendix A4, we show that there exists an analogous set of sufficient statistics for the relative state of technology $C_M^*(z)/C_M(z)$ in the Melitz model with CES demand.

By definition, a country's RCA index for a given industry equals the proportion of the country's exports the industry accounts for divided by the proportion of world exports the industry accounts for. A comparative advantage (disadvantage) for the country is "revealed" in the industry under consideration if the corresponding RCA index is larger (smaller) than 1. Accordingly, in the case of our model, Balassa's RCA index for industry z is given by

$$RCA_B(z) = \frac{Exp(z)/\int_0^1 Exp(z)dz}{(Exp(z) + Exp^*(z))\,/\left(\int_0^1 Exp(z)dz + \int_0^1 Exp^*(z)dz\right)}, \tag{26}$$

where $Exp(z)$ and $Exp^*(z)$ are industry z's Home and Foreign exports respectively. It is obvious that $RCA_B(z)$ increases with $Exp(z)/Exp^*(z)$. Therefore, the ranking of industries according to the RCA index is the same as the ranking determined by relative exports. The following result unpacks the components of relative exports and provides sufficient statistics for them:

Proposition 8. ***(Balassa Index and Comparative Advantage)*** *(a) Home industry z's relative exports can be decomposed as*

$$\frac{Exp(z)}{Exp^*(z)} = \frac{\frac{I}{I^*}\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}} \frac{\rho^{-1} + \frac{L^*}{L}\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}} + \frac{L}{L^*}\rho} \frac{\theta(z)}{1-\theta(z)} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} \left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\right)^{k+1}. \tag{27}$$

(b) The Pareto shape parameter k, trade freeness ρ, export propensity $\chi(z)$, export intensity $\theta(z)$ and relative income I/I^* are sufficient statistics for $Exp(z)/Exp^*(z)$ and its components.
(c) The sufficient statistics for the RCA index $RCA_B(z)$ further include Home exports $Exp(z)$.

Proof. See Appendix A1.11. □

Result (a) unveils what determines relative exports and thus the RCA index. Ricardian and factor prices aspects are recognizable in the last two factors on the right-hand side of expression (27): higher relative TFPQ in autarky and lower relative unit input prices promote Home relative exports. In addition, aspects stressed by new trade theories also matter. Firstly, relative exports increase with Home's relative income, which echoes the "home market effect" emphasized by Krugman (1980): with monopolistic competition, the country with larger local demand in a given industry tends to be a net exporter in that industry. Secondly, relative exports increase with relative export selection as Home exports more in industries with relatively higher export propensity and intensity. Finally, export selection and the home market effect complement each other as Home's relative exports increase more with Home's relative income if Home has a relatively higher export propensity.

Clearly, result (a) also implies that relative exports are not a sufficient statistic for relative autarkic TFPQ, and thus $RCA_B(z)$ does not contain enough information to reveal Ricardian comparative advantage. In this respect, result (b) provides sufficient statistics for Home relative exports without using foreign trade data. Once we know relative exports, we can rank industries according to $RCA_B(z)$.

However, according to result (c), if we want to pin down the exact value of $RCA_B(z)$, we need the export volume of industry z.

Proposition 8 requires micro data, in particular, export propensity $\chi(z)$. The following result shows that we can alternatively rely on macro data only:

Corollary 2. ***(Macro Sufficient Statistics for Relative Exports)*** *Trade freeness ρ, export intensity $\theta(z)$ and relative aggregate income $LI/(L^*I^*)$ are alternative sufficient statistics for Home industry z's relative exports and their components.*

Proof. See Appendix A1.12. □

Overall, as summarized in Table 2, propositions 7 and 8 together with corollaries 1 and 2 allow us to quantify Ricardian comparative advantage and Balassa's RCA index by sufficient statistics using either micro or macro data. Moreover, the sufficient statistics approach can also be used to evaluate the effect of changes in unit input prices on revealed comparative advantage without fully estimating the model thanks to the following result:

Corollary 3. ***(Equivalent Rankings)*** *(a) The ranking of industries based on Balassa's index $RCA_B(z)$ is the same as the ranking based on export propensity $\chi(z)$ or export intensity $\theta(z)$. (b) It is also the same as the ranking based on $(\omega^*(z)/\omega(z))^{k+1}\,(C_M^*(z)/C_M(z))^k$.*

Proof. See Appendix A1.13. □

We already know from Lemma 2 that industry rankings by export propensity and export intensity are the same. Result (a) states that the industry ranking by Balassa's RCA index is also the same. We have further discussed how Proposition 2 implies that, for our model's predictions to match the first stylized fact, we have to impose that China has a better relative state of technology and lower relative input prices in higher z industries so that both $\frac{\omega^*(z)}{\omega(z)}$ and $\frac{C_M^*(z)}{C_M(z)}$ are increasing function of z. Result (b) states that in this case, $RCA_B(z)$ is also increasing in z. In addition, it suggests how to perform a counterfactual analysis of the effects of changes in unit input prices on Balassa's RCA.[30] Specifically, for any counterfactual change in the relative unit input price from factual $\frac{\omega^*(z)}{\omega(z)}$ to counterfactual $\frac{\omega^{*\prime}(z)}{\omega^{\prime}(z)}$, we can rely on result (b) to determine the resulting RCA ranking according to $\frac{\omega^*(z)'^{k+1}C_M^*(z)^k}{\omega(z)'^{k+1}C_M(z)^k}$.[31]

4 Quantification

This section applies the sufficient statistics approach developed in the previous section to quantify Ricardian comparative advantage and relative productivity gains from trade, as well as Balassa's

[30]Differently from us, Chor (2010) and Costinot et al. (2012) conduct counterfactual analyses on comparative advantage within a Ricardian framework based on Eaton and Kortum (2002).

[31]The effects of counterfactual changes in $C_M^*(z)/C_M(z)$ cannot be evaluated in the same way as, in this case, the implied changes in the relative unit input price should also be assessed.

RCA index with both factual and counterfactual relative unit input prices. Given that the approach can be applied using micro or macro data, we provide two sets of results: results for China vis-à-vis the Rest of the World (RoW) leveraging its firm-level data, and results for 41 economies leveraging their sectoral trade data in the World Input-Output Database.

4.1 VES vs. CES

Before implementing the sufficient statistics, we offer direct evidence that a VES demand system is better suited than a CES one to explain firms' pricing behavior. Specifically, we rely on the export pricing equation $p_X(z,\nu) = \frac{-\delta}{1-\delta} p^*_{max}(z) + \frac{1}{1-\delta}\tau\omega(z)\nu(m,c)$, which is stated by analogy to (4). This equation suggests the following empirical specification that also allows us to estimate δ by ordinary least squares using matched ASIF and CCTS data for 2004:

$$p_n^{jt} = \phi p_{\max}^{jt} + \sum_m a_m X_m^n + \sum_l b_l Y_l^j + \sum_s d_s Z_t^s + \epsilon_n^{jt}, \tag{28}$$

where p_n^{jt} is the price charged by a firm n at destination j for product t, $p_{\max}^{jt}$ is the associated choke price, and ϵ_n^{jt} is the error term. We assume that firms' marginal cost selling product t to market j depends on firm (X_m^n), destination (Y_l^j), and product (Z_s^t) characteristics.[32] As with CES demand firms price at a constant markup over marginal cost, the choke price does not appear in their pricing equations, which would be the case for ϕ=0.[33] Accordingly, a positive estimate of ϕ would support VES against CES. Moreover, as the export pricing equation implies $\phi = -\delta/(1-\delta)$, we can readily retrieve the point estimate of δ from the point estimate of ϕ as $\delta = \phi/(\phi - 1)$.

The estimation results are shown in Appendix Table T1. The dependent variable is the price Chinese exporters charge for each HS 8-digit product in each destination market, measured by the unit value (value/quantity). We measure the choke price p_{max}^{jt} using the maximum price among all CCTS exporters of the same product in each market. Column (1) includes only the choke price. In column (2), we further control for the gravity variables. As expected, prices increase with distance and decrease with common language or shared border. Column (3) adds the global rank of a product in an exporter's total exports. The positive coefficient of the product rank variable implies that the top-ranked products tend to have lower prices than the bottom-ranked products. Column (4) adds the destination fixed effect, which absorbs the gravity variables. Columns (5) and (6) further include firm and product category fixed effects. As ϕ differs significantly from 0 in all columns, a VES demand system is better than a CES one in explaining firms' pricing behavior. Our preferred value is from column (6), where we add all fixed effects. It implies $\delta = -2.46$ is less than -1. Hence, on average,

[32] As an alternative, we also estimate δ by non-linear least squares (NLLS), taking into account that marginal cost enters the price with a coefficient of $1/(1-\delta)$. Appendix Table B3 presents the corresponding results. While the estimated δ is close to the OLS estimate, we prefer OLS to NLLS for two reasons: (i) the NLLS results are sensitive to initial guesses of δ; (ii) with NLLS, it is computationally difficult to add high-dimension firm fixed effects.

[33] As shown by Melitz (2003), to have selection into export status with CES demand, one has to assume the existence of a fixed export cost. As there is no choke price with CES demand, the $p_{max}(z)$ is the maximum price that allows an exported product to generate enough operating profits to cover the fixed export cost.

Chinese exporters face a demand curve that is slightly less elastic than the linear demand and put a higher weight on the choke price than their own marginal cost in pricing.

4.2 Micro Data Analysis

We start by estimating the parameters needed to compute the sufficient statistics, and then use them to apply propositions 7 and 8 and corollary 3. In doing so, we rely on Chinese micro data for the year 2004 as described in Section 2.1.

4.2.1 Parameter Estimation

We first estimate the Pareto shape parameter k from the distribution of firms' TFP, which is the inverse of their UIR, obtained as described in Section 2.1. In particular, we run a log-log regression of $G(c)$ on c. The results are shown in Appendix Table T2. Column (1) reports the ordinary least squares results, while column (2) reports those with industry fixed effects.[34] Column (3) uses an M-estimator, which reduces the influence of outliers and gives our preferred estimate of $k = 1.310$.[35]

We next estimate trade freeness ρ. We use the method of Head and Ries (2001), who show that if aggregate bilateral trade flows follow the gravity equation, under the assumptions of symmetric trade costs and free domestic trade, one can write $\rho_{ij} = \sqrt{X_{ij}X_{ji}/\left(X_{ii}X_{jj}\right)}$ with X_{ij} being the aggregate exports from region i to region j. In the same vein, in Appendix A3, we show that our model implies

$$\rho = \sqrt{\frac{Exp}{S_D}\frac{Exp^*}{S_D^*}},$$

where Exp and S_D (Exp^* and S_D^*) are Home (Foreign) export and domestic sales. To operationalize this result, we need local sales data S_D^* in the RoW, which are unavailable from the Chinese firm survey or customs data. We instead use the World Input-Output Database, from which we retrieve exports and local sales for China and RoW in 2004.[36] We estimate the trade freeness of all manufacturing industries as an aggregate. As shown in Appendix Table T3, the resulting estimate is $\rho = 0.0754$.

To quantify the ex-ante component of relative TFPQ, we also need the relative unit input price $\omega(z)/\omega^*(z)$. We assume that industry z's composite input is a Cobb-Douglas combination of labor and capital with labor share $\iota(z) \in (0, 1)$ where w (w^*) and r (r^*) are wage and rental rates in Home and Foreign respectively. Therefore, the relative unit input price is given by

$$\omega(z)/\omega^*(z) = (w/w^*)^{\iota(z)}(r/r^*)^{1-\iota(z)}, \tag{29}$$

where $\iota(z)$ is measured as average firm labor cost relative to value added. We estimate China's wage rate w/w^* and interest rate r/r^* relative to the RoW using data from the Penn World Table

[34] This is consistent with our assumption that the industries share the same Pareto shape of the UIR distribution but differ in the UIR upper bound.

[35] Head et al. (2014) estimated a Pareto shape parameter of 1.367 using Chinese exports to Japan for the year 2000.

[36] We use the 2013 release (Timmer et al., 2015) and aggregate all regions except China as RoW.

10.0 (Feenstra et al., 2015). The wage rate is estimated as real GDP multiplied by labor share and then divided by total employment. The interest rate is measured by the real internal rate of return. Measures for the RoW are the weighted average across all countries without missing data, excluding China, with wage rate weighted by employment and interest rate weighted by capital stock. As Table T3 indicates, the estimated relative wage rate is $w/w^* = 0.367$, and the relative interest rate is $r/r^* = 1.264$, which are consistent with the common view that in 2004 China was a labor-abundant and capital-scarce country. Table T3 also displays the estimated consumer income of China relative to RoW as measured by relative GDP per capita, $I/I^* = 0.488$, which is needed for the estimation of Balassa's RCA index.

Finally, ASIF covers large firms in China. It, therefore, tends to overstate the export propensity and intensity of the entire firm population. For example, in the 2004 Industrial Census, which covers both large and small firms, the export propensity and intensity of manufacturing firms were 10.0% and 21.7%, respectively. In contrast, according to ASIF, export propensity and intensity were 30.0% and 23.3%, respectively. To correct this sample selection problem, we scale the export propensity for each industry by $10.0/30.0$ and export intensity by $21.7/23.3$.[37] As a robustness check we will also report results using export propensity and intensity observed in the census.

4.2.2 Results

We are now ready to look into Ricardian comparative advantage, relative productivity gains from trade and Balassa's RCA index.

Ricardian Comparative Advantage We first quantify the ex-post components of relative productivity using sufficient statistics as per Proposition 7 (a). Figure 2 (a) plots the endogenous amplifying component of the relative TFPQ, $XPA(z)$, which tends to rise with export propensity. Therefore, China's Ricardian comparative advantage is amplified in industries with higher export propensity. However, as emphasized in Proposition 5, there is an additional dampening aspect of export selection. Figure 2 (b) plots the combined ex-post components $XPA(z) \cdot XPD(z)$ of relative TFPQ in equation (22), thus capturing both the amplifying and dampening aspects. Since $XPA(z) \cdot XPD(z)$ falls whereas $XPA(z)$ rises with the export propensity, it must be that the dampening component $XPD(z)$ dominates.

Next, we compute China's relative unit input prices $\frac{\omega(z)}{\omega^*(z)}$ with respect to the RoW. Consistent with the first stylized fact, Figure 2 (c) shows that export propensity is higher in industries with lower relative unit input prices, which are those with higher labor intensity. With relative unit input prices at hand, we can then quantify the ex-ante component of relative TFPQ, $\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}$, and the relative state of technology, $\frac{C^*_M(z)}{C_M(z)}$, by applying Proposition 7 (b). Figure 2 (d) shows that both the ex-ante component

[37] Without the correction, the estimated average productivity of China relative to RoW is greater than 1, which was unlikely in 2004.

and the relative state of technology rise with export propensity. In other words, China exhibits a Ricardian comparative advantage in industries of higher export propensity. Ricardian comparative advantage and relative unit input prices complement each other by favoring similar industries. The different slopes of $\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}$ and $\frac{C_M^*(z)}{C_M(z)}$ reveal the differential impact of domestic selection in autarky between China and the RoW: as the former has a smaller market size, its autarkic relative TFPQ is smaller than its relative state of technology due to weaker selection.

In addition, we can obtain relative productivity by combining the ex-ante and ex-post components. Its unweighted and weighted average measures are plotted in Figure 2 (e) and (f), respectively. While the former increases monotonically with export propensity, there exists an inverted U-shaped relationship between the latter and export propensity. In particular, when export propensity is high, the dominating dampening effect shown in Figure 2 (b) reduces China's TFPQ relative to the RoW.

Finally, we evaluate the contributions of the ex-ante and ex-post components to the variation of productivity and find sizeable export selection effects. Specifically, we run two types of regressions. First, we regress the ex-ante and ex-post components of relative unweighted average productivity in (23) on this relative productivity measure (in logs). Second, we regress the ex-ante and ex-post components of relative TFPQ in (22) on this weighted relative productivity measure (in logs). The coefficients of these regressions sum up to unity, with the size of each coefficient capturing the explanatory power of the respective component. Appendix Table T4 presents the results. Columns (1) and (2) decompose the variation of relative unweighted productivity: the ex-ante component explains 75.3% of the variation, while the ex-post component accounts for the remaining 24.7%. Columns (3)-(4) decompose the variation of relative TFPQ: the ex-ante component explains 112.8%, and two ex-post components jointly undo the extra 12.8%. Columns (5) and (6) separate the amplifying and dampening components: their combined effect of -12.8% is the net effect of a large positive contribution of the former component (340.3%) and an even larger negative contribution of the latter component (-353.0%). Accordingly, consistent with Figure 2 (b), the dampening component dominates the amplifying component, which results in the compression of the ex-ante component.

Relative Productivity Gains from Trade Export selection generates aggregate productivity gains via the reallocation of resources between and within firms. We have shown that reallocation varies systematically across industries. We next check here whether reallocation matters for "gains from trade" in terms of relative productivity, which for industry z are defined as the ratio of relative productivity between the open and the closed economies.[38]

Once again, we consider both the weighted and unweighted measures of industry productivity. From propositions 5 and 6, the corresponding gains are $\frac{\overline{\Phi}(z)}{\overline{\Phi}^*(z)} / \frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = XPA(z) \cdot XPD(z)$ and $\frac{\overline{c}(z)^{-1}}{\overline{c}^*(z)^{-1}} / \frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = UXPA(z)$, respectively. However, aggregating from industry to country gains from trade also requires a choice of industry weights. We compare three alternative industry weighting

[38]We borrow the terminology from Arkolakis, Costinot, and Rodríguez-Clare (2012), who call (welfare) "gains from trade" the ratio of indirect utility between the open and the closed economies.

schemes: uniform ("none"), by firm number, and by output.

Table 3 shows the corresponding results. Our preferred outcomes use output weights, both within and between industries, i.e., those in row (c) and columns (1), (2), and (4). These cells show that in 2004, China was 85.67% (column (2)) as productive as the RoW, and it would have been 84.47% (column (1)) as productive in autarky. Its productivity gains from trade were, therefore, a tiny 1.42%. This is due to the dampening effect of export selection, as revealed by comparisons with the cells where that effect is muted by alternative weighting. For example, when industry productivity is computed as unweighted average firm TFP, China's productivity gains from trade are 16.78%, 17.11% or 15.16% depending on whether industry weights are uniform, by firm number or by output respectively. Ignoring the dampening component would significantly overstate the relative productivity gains of China from trade.

Overall, the net impact of trade-induced micro reallocations on aggregate productivity might be small, but this is the result of quite sizeable reallocations at different margins that happen to almost offset each other.

Balassa's RCA Index and Counterfactuals We now apply Proposition 8 to quantitatively explore the relation between China's Balassa's RCA index and the exogenous source of its Ricardian comparative advantage, that is, its relative state of technology.

Figure 3 presents the ranking of industries by Balassa's RCA index against the relative state of technology $\frac{C_M^*(z)}{C_M(z)}$ and the relative unit input price $\frac{\omega(z)}{\omega^*(z)}$ in 2004. There are two main observations. First, the correlation between $\frac{C_M^*(z)}{C_M(z)}$ and $\frac{\omega(z)}{\omega^*(z)}$ is negative: Chinese industries with lower unit input prices tend to have relatively better state of technology compared with the RoW. Second, industries with lower $\frac{\omega(z)}{\omega^*(z)}$ and higher $\frac{C_M^*(z)}{C_M(z)}$ tend to get better rankings by Balassa's RCA index. These two observations imply that Ricardian and Heckscher-Ohlin sources of comparative advantage tend to complement each other in shaping Chinese export specialization.

Corollary 3 allows us to go one step further. A growing number of empirical studies highlight how rapidly rising wages are eroding China's "global labor advantage" (Yang et al., 2010; Gan et al., 2016; Hau et al., 2020). Our model predicts that this development will give more prominence to Ricardian comparative advantage, and we can rely on the corollary to figure out how China's export specialization will change as some industries expand and others shrink. As examples, we consider two scenarios: (a) full factor price equalization (FPE) so that the pattern of trade is only driven by relative productivity; (b) partial factor price convergence with factor prices set equal to the ones for 2019. In both cases, we keep the relative state of technology unchanged at its 2004 estimate.

Appendix Table T5 lists the top and bottom 10 industries ranked by Balassa's RCA index. Column (1) is the baseline ranking for 2004. The ranking is broadly consistent with the observation that China was a competitive exporter of electronics, machinery, toys, and hats in 2004. Column (2) presents the ranking under FPE with $\omega(z) = \omega^*(z)$. The numbers in square brackets refer to each industry's ranking in the baseline. We find that labor-intensive industries, such as "Toy manufacturing" and

"Hat making" drop from the top 10 (with new rankings 50 and 51, respectively), while some capital-intensive industries, such as "Manufacturing of slideshow and projection equipment" and "Integrated circuit manufacturing" (i.e. "chips"), get into the top 10 (with baseline rankings 36 and 18, respectively). Column (3) replaces the relative input prices of 2004 with their values in 2019. The estimated relative wage rate w/w^* rises from 0.367 to 0.695, and the relative interest rate r/r^* falls from 1.264 to 0.953. In this scenario, we also find that the RCA of labor-intensive industries is weakened. The fact that the top and bottom 10 industries in columns (2) and (3) are similar suggests that China could be already close to exhausting its "global labor advantage" at its 2004 relative state of technology.[39]

Going beyond the top and bottom 10 industries, we can capture the changes in the ranking of all industries through local polynomial regressions of the counterfactual rankings on the 2004 baseline ranking. The regression results are presented in Figure 4. Figure (a) represents the analysis conducted on the entire sample of industries, while figure (b) focuses on the top 100 industries in the baseline ranking. Across the two counterfactual scenarios, we find that the rankings of industries at the top tend to fall while those at the bottom tend to rise. FPE leads to a slightly larger fall in rankings among the top industries than replacing factor prices with their 2019 values, which confirms the idea that China could be close to exhausting its "global labor advantage," even though it has not happened yet.

It is important to highlight that firm heterogeneity matters for the counterfactual predictions. Figure 4 also depicts the regression results for the 2019 counterfactual with a larger Pareto shape $k' = 10k = 13.1$ instead of the baseline value $k = 1.310$. It shows that the top industries' rankings fall further when firms are less heterogeneous (i.e. when k is larger). This is to be expected because, as discussed in reference to Proposition 2, less firm heterogeneity increases the passthrough from the relative state of technology to Ricardian comparative advantage. Therefore, changes in relative factor prices have a larger impact on revealed comparative advantage.

Robustness Check with Census Data

Our micro data analysis has relied so far on ASIF. As discussed in Section 4.2.1, a problem with ASIF is that it covers only large firms, and we have had to deal with the implied potential selection bias by rescaling the measures of export propensity and intensity.

Alternatively, we could have used the 2004 Industrial Census, which also covers small firms. However, the census does not report small firms' labor and intermediate inputs, which prevents us from estimating their labor intensity and TFP. Nor can we match the small firms with CCTS, which prevents us from examining their export product scope and mix. Hence, relying on ASIF has allowed us to keep the quantification exercise internally consistent.

That said, as a robustness check, appendix figures F1 and F2 report the estimation results obtained by using industry export propensity and intensity from the census data, while keeping other variables and parameters fixed. These results closely resemble the baseline ones in figures 2 and 3, respectively.

[39] This is under the assumption that $C^*_M(z)/C_M(z)$ is fixed at its 2004 value. The counterfactual rankings would have been different if we had allowed the relative state of technology to change over time.

4.3 Macro Data analysis

Corollaries 1 and 2 indicate that we can also quantify the model by implementing a sufficient statistics approach without micro data. We show here how one can use the World Input-Output Database (WIOD) to quantify the patterns of comparative advantage of 41 economies with respect to the RoW.

4.3.1 Parameter Estimation

We again use the 2013 WIOD release and choose 2004 as the baseline year. We aggregate industries into 15 manufacturing industries and one service industry to reduce the number of zeros in trade flows. According to our model, the Pareto shape k corresponds to the trade elasticity of a standard gravity equation, and we lift its value 5.03 from the median estimate in Head and Mayer (2014). For trade freeness, we apply the method again by Head and Ries (2001) and estimate ρ for each economy and industry. For the aggregate income of each economy (LI), we take its total value added across all industries. We then compute the relative market size (L/L^*) as an economy's population relative to the total population of the RoW, using data from the Penn World Table. Finally, an economy's export intensity in each industry ($\theta(z)$) is measured as exports divided by the sum of exports and domestic absorption.

4.3.2 Results

For each industry in each economy, we first apply Corollary 1 to quantify the ex-post TFPQ components relative to the RoW.[40] Then, for each economy, we compute its simple average relative TFPQ gains from trade across industries and average export intensity. Figure 5 (a) plots each economy's average relative TFPQ gains from trade against its average export intensity. It shows that economies with higher average export intensity enjoy relatively larger productivity gains. For example, Luxembourg exhibits an average export intensity of about 78.0% and TFPQ gains of 190%, whereas China features an average export intensity of about 16.0% and TFPQ gains of just 2.6%, which is in the ball park of the unweighted TFPQ gains based on micro data reported in Table 3.

We next use Corollary 2 to estimate each economy's relative exports. We then rank industries according to the estimated relative exports, which gives the same ranking as Balassa's RCA index. Since in the WIOD we also observe each economy's exports in each industry, we can compute the empirical counterpart of relative exports and their rank. Figure 5 (b) plots the correlations of the model-based estimated ranks and the actual ranks of relative exports against each economy's size as measured by total output (in logs). The correlation ranges from -0.11 to 0.96 and averages 0.49. Overall, the correlation tends to be higher for economies with higher output, which implies that our sufficient statistics approach works better for larger economies.

[40]Estimating the ex-ante component requires knowing the input prices of each economy relative to the RoW in each industry (see Proposition 7 (b) indicates). We do not have such data.

5 Conclusion

In a seminal paper, Balassa (1965) introduced the notion of "revealed comparative advantage" as a measure of a country's comparative advantage based on its export specialization across industries relative to the rest of the world. The subsequent literature, however, has highlighted two problems with this approach. First, the sectoral ranking of revealed comparative advantage according to relative sectoral exports may not coincide with the ranking of comparative advantage based on relative costs. Second, if one is interested in a specific source of comparative advantage, the various exogenous and endogenous determinants of the observed export patterns (such as relative unit input requirements, factor endowments, market size, firm, and product selection) may be hard to disentangle as long as some of them are unobserved.

On both accounts, in modern terminology, Balassa's index of revealed comparative advantage is not a sufficient statistic for identifying the existence and the magnitude of any specific source of comparative advantage. We have tackled this issue from a Ricardian perspective by developing a structural approach based on a quantitative trade model with firm and product selection where both relative unit input requirements and relative factor endowments play a role. Through the lens of the model, we have identified sets of sufficient statistics for the endogenous and exogenous components of a country's productivity relative to the rest of the world. Among the exogenous components, we have emphasized the country's relative state of technology as the exogenous determinant of its Ricardian comparative advantage. While the state of technology has important implications for the micro behavior of individual firms, firms' responses have, in turn, far-reaching macro implications for aggregate productivity and revealed comparative advantage.

Applying our sufficient statistics approach to China, we have quantified its Ricardian comparative advantage and its gains from trade in terms of productivity. Through counterfactual analysis, we have also assessed how China's revealed comparative advantage in Balassa's sense would change if it lost its "global labor advantage." Importantly, we have shown that the model's micro foundations do not necessarily imply that the relevant data for the proposed sufficient statistics must include micro information (i.e. at the firm or product level), but its micro structure is needed to understand how only macro information (i.e. at the sector or country level) can also be used instead. Our approach can thus be applied to cross-country analyses, for which harmonized micro data are unavailable, and we have provided an example quantifying the gains from trade in terms of relative productivity across a large set of countries.

References

Ackerberg, D.A., Caves, K. and Frazer, G., 2015. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6), pp.2411-2451.

Ahn, J., Khandelwal, A.K. and Wei, S.J., 2011. The Role of Intermediaries in Facilitating Trade. *Journal of International Economics*, 84(1), pp.73-85.

Arkolakis, C., Costinot, A. and Rodríguez-Clare, A., 2012. New Trade Models, Same Old Gains?. *American Economic Review*, 102(1), pp.94-130.

Arkolakis, C., Costinot, A., Donaldson, D. and Rodríguez-Clare, A., 2019. The Elusive Pro-Competitive Effects of Trade. *Review of Economic Studies*, 86(1), pp.46-80.

Arkolakis, C., Ganapati, S. and Muendler, M.A., 2021. The Extensive Margin of Exporting Products: A Firm-Level Analysis. *American Economic Journal: Macroeconomics*, 13(4), pp.182-245.

Atkin, D. and Donaldson, D., 2015. Who's Getting Globalized? The Size and Implications of Intra-national Trade. *National Bureau of Economic Research*. Working paper No. w21439.

Bai, X., Chatterjee, A., Krishna, K. and Ma, H., 2022. Trade and Selection with Heterogeneous Firms and Perfect Competition. *National Bureau of Economic Research*. Working paper No. w30650.

Bai, X., Krishna, K. and Ma, H., 2017. How You Export Matters: Export Mode, Learning and Productivity in China. *Journal of International Economics*, 104, pp.122-137.

Balassa, B., 1965. Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School*, 33(2), pp.99-123.

Bernard, A.B., Redding, S.J. and Schott, P.K., 2007. Comparative Advantage and Heterogeneous Firms. *Review of Economic Studies*, 74(1), pp.31-66.

Bernard, A.B., Redding, S.J. and Schott, P.K., 2010. Multiple-Product Firms and Product Switching. *American Economic Review* , 100(1), pp.70-97.

Bernard, A.B., Redding, S.J. and Schott, P.K., 2011. Multiproduct Firms and Trade Liberalization. *Quarterly Journal of Economics*, 126(3), pp.1271-1318.

Bernhofen, D.M. and Brown, J.C., 2004. A Direct Test of the Theory of Comparative Advantage: The Case of Japan. *Journal of Political Economy*, 112(1), pp.48-67.

Brandt, L., Van Biesebroeck, J. and Zhang, Y., 2012. Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, 97(2), pp.339-351.

Brandt, L., Van Biesebroeck, J., Wang, L. and Zhang, Y., 2017. WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9), pp.2784-2820.

Bulow, J.I. and Pfleiderer, P., 1983. A Note on the Effect of Cost Changes on Prices. *Journal of Political Economy*, 91(1), pp.182-185.

Burstein, A. and Vogel, J., 2017. International Trade, Technology, and the Skill Premium. *Journal of Political Economy*, 125(5), pp.1356-1412.

Chor, D., 2010. Unpacking Sources of Comparative Advantage: A Quantitative Approach. *Journal of International Economics*, 82(2), pp.152-167.

Corcos, G., Del Gatto, M., Mion, G. and Ottaviano, G.I., 2012. Productivity and Firm Selection: Quantifying the 'New' Gains from Trade. *The Economic Journal*, 122(561), pp.754-798.

Costinot, A., Donaldson, D. and Komunjer, I., 2012. What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *Review of Economic Studies*, pp.581-608.

Costinot, A., Donaldson, D. and Smith, C., 2016. Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World. *Journal of Political Economy*, 124(1), pp.205-248.

Deardorff, A.V., 1980. The General Validity of the Law of Comparative Advantage. *Journal of Political Economy*, 88(5), pp.941-957.

Dornbusch, R., Fischer, S. and Samuelson, P.A., 1977. Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review*, 67(5), pp.823-839.

Dornbusch, R., Fischer, S. and Samuelson, P.A., 1980. Heckscher-Ohlin Trade Theory with a Continuum of Goods. *Quarterly Journal of Economics*, 95(2), pp.203-224.

Eaton, J. and Kortum, S., 2002. Technology, Geography, and Trade. *Econometrica*, 70(5), pp.1741-1779.

Eckel, Carsten, and J. Peter Neary. Multi-Product Firms and Flexible Manufacturing in the Global Economy. *Review of Economic Studies*, 77.1 (2010): 188-217.

Feenstra, R. and Ma, H., 2009. Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition, E. Helpman, D. Marin and T. Verdier, eds., *The Organization of Firms in a Global Economy*, Harvard University Press, 173-199.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), The Next Generation of the Penn World Table. *American Economic Review*, 105(10), 3150-3182.

French, S., 2017. Revealed Comparative Advantage: What Is It Good For?. *Journal of International Economics*, 106, pp.83-103.

Gan, L., Hernandez, M.A. and Ma, S., 2016. The Higher Costs of Doing Business in China: Minimum Wages and Firms' Export Behavior. *Journal of International Economics*, 100, pp.81-94.

Gaubert, C. and Itskhoki, O., 2021. Granular Comparative Advantage. *Journal of Political Economy*, 129(3), pp.871-939.

Haberler, G., 1930. Die Theorie der komparativen Kosten und ihre Auswertung für die Begründung des Freihandels. *Weltwirtschaftliches Archiv*, pp.349-370.

Hau, H., Huang, Y. and Wang, G., 2020. Firm Response to Competitive Shocks: Evidence from China's Minimum Wage Policy. *Review of Economic Studies*, 87(6), pp.2639-2671.

Head, K. and Mayer, T., 2014. Gravity Equations: Workhorse, Toolkit, and Cookbook. *In Handbook of International Economics*, (Vol. 4, pp. 131-195). Elsevier.

Head, K., Mayer, T. and Thoenig, M., 2014. Welfare and Trade without Pareto. *American Economic Review*, 104(5), pp.310-316.

Head, K. and Ries, J., 2001. Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade. *American Economic Review*, pp.858-876.

Huang, H., Ju, J. and Yue, V.Z., 2017. Structural Adjustments and International Trade: Theory and Evidence from China. Working Paper.

Iacovone, L. and Javorcik, B.S., 2010. Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries. *The Economic Journal*, 120(544), pp.481-499.

Krugman, P., 1980. Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review* , 70(5), pp.950-959.

Levchenko, A.A. and Zhang, J., 2016. The Evolution of Comparative Advantage: Measurement and Welfare Implications. *Journal of Monetary Economics*, 78, pp.96-111.

Manova, K. and Yu, Z., 2017. Multi-Product Firms and Product Quality. *Journal of International Economics*, 109, pp.116-137.

Mayer, T., Melitz, M.J. and Ottaviano, G.I., 2014. Market Size, Competition, and the Product Mix of Exporters. *American Economic Review*, 104(2), pp.495-536.

Mayer, T., Melitz, M.J. and Ottaviano, G.I., 2021. Product Mix and Firm Productivity Responses to Trade Competition. *Review of Economics and Statistics*, 103(5), pp.874-891.

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

Melitz, M.J. and Ottaviano, G.I., 2008. Market Size, Trade, and Productivity. *The Review of Economic Studies*, 75(1), pp.295-316.

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

Mill, J. 1821. Elements of Political Economy. London: Baldwin, Cradock & Joy.

Mill, J.S., 1871. Principles of Political Economy with Some of their Applications to Social Philosophy. 3rd edition. London. J. W. Parker.

Mrázová, M. and Neary, J.P., 2017. Not So Demanding: Demand Structure and Firm Behavior. *American Economic Review*, 107(12), pp.3835-74.

Ottaviano, G.I. and Suverato, D., 2023. Fantastic Beasts and Where to Find Them, *Bocconi University*, mimeo.

Redding, S. and Weinstein, D.E., 2018. Accounting for Trade Patterns. *manuscript, Princeton University*.

Romalis, J., 2004. Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94(1), pp.67-97.

Silva, J.S. and Tenreyro, S., 2006. The Log of Gravity. *Review of Economics and Statistics*, 88(4), pp.641-658.

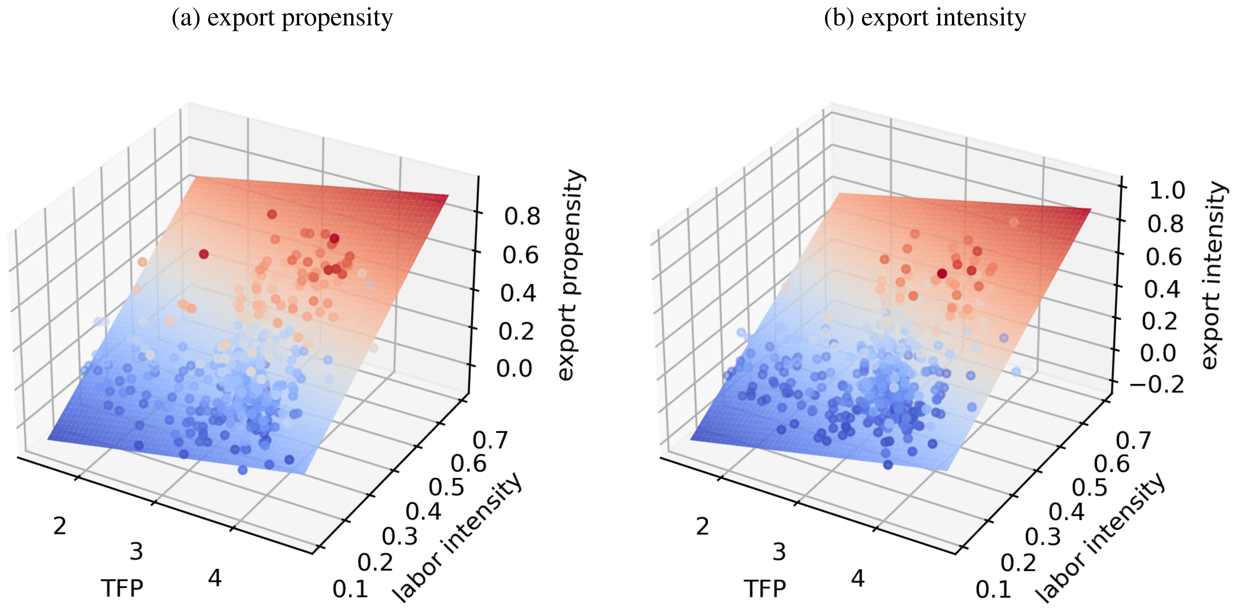
Timmer, M.P., Dietzenbacher, E., Los, B., Stehrer, R. and Vries, G.J., 2015. An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production. *Review of International Economics*, 23(3), pp.575-605.

Yang, D.T., Chen, V.W. and Monarch, R., 2010. Rising Wages: Has China Lost Its Global Labor Advantage?. *Pacific Economic Review*, 15(4), pp.482-504.

Yi, K.M., 2003. Can Vertical Specialization Explain the Growth of World Trade?. *Journal of Political Economy*, 111(1), pp.52-102.

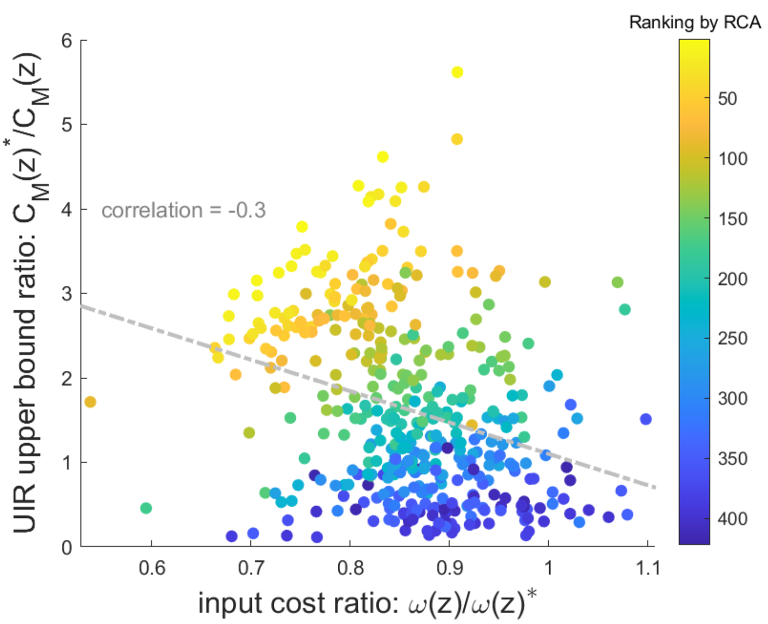
Yu, M., 2015. Processing Trade, Tariff Reductions and Firm Productivity: Evidence from Chinese Firms. *The Economic Journal*, 125(585), pp.943-988.

Figure 1: Export Propensity and Intensity



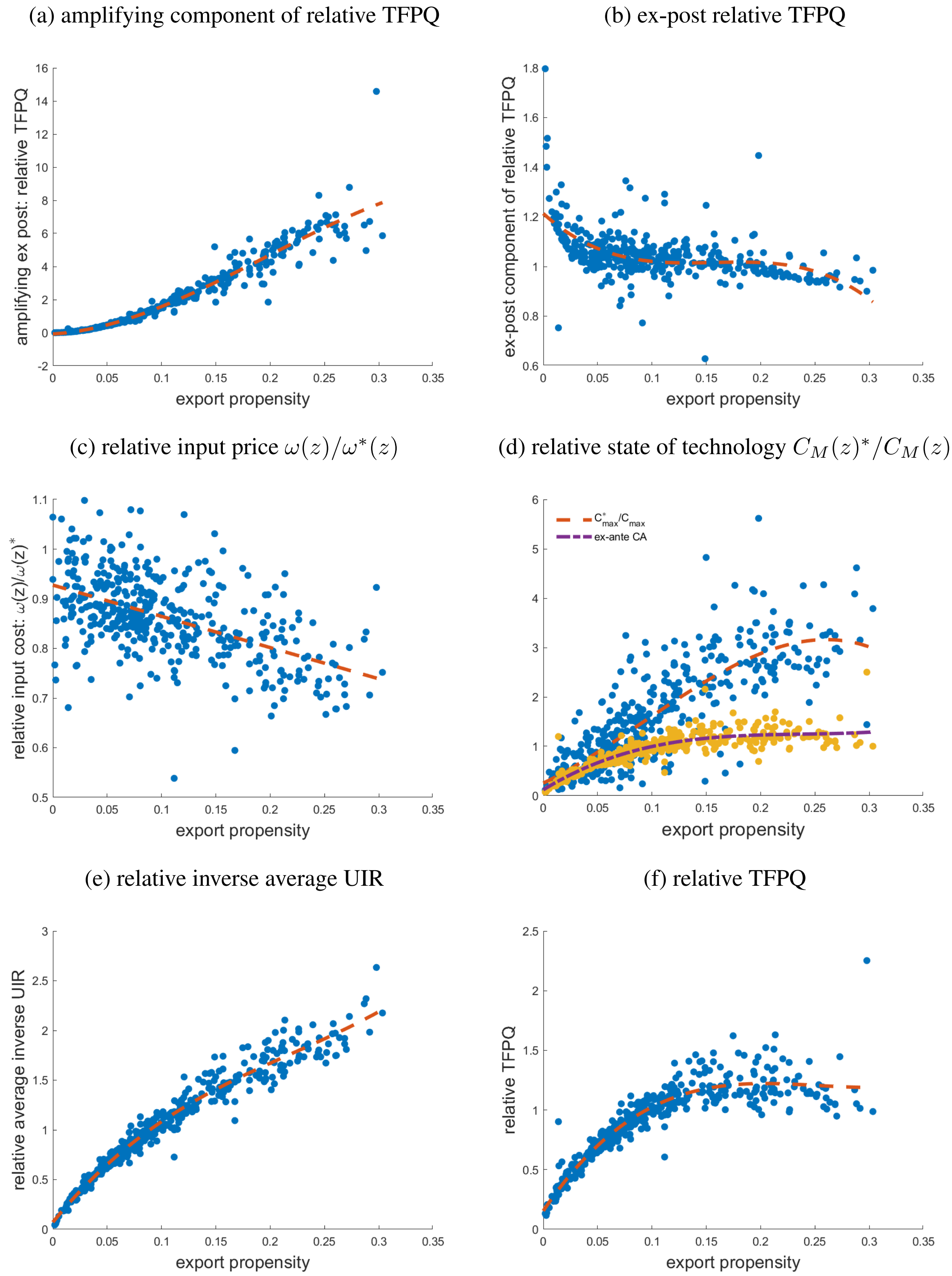
Notes: Figure (a) plots export propensity, the fraction of firms within each industry that are exporters. Figure (b) displays export intensity, the share of exported sales. The surfaces represent the best-fitting linear plane of the scatter plots. Industry TFP and labor intensity are simple averages across firms.

Figure 3: Comparative Advantage and RCA Rankings



Notes: This diagram shows the relative input price ratio ($\omega(z)/\omega^*(z)$, horizontal axis) and the relative state of technology ($C_M(z)^*/C_M(z)$, vertical axis) across industries. The color scheme represents Balassa's RCA index. The dashed line captures the correlation between the input price ratio and technology level.

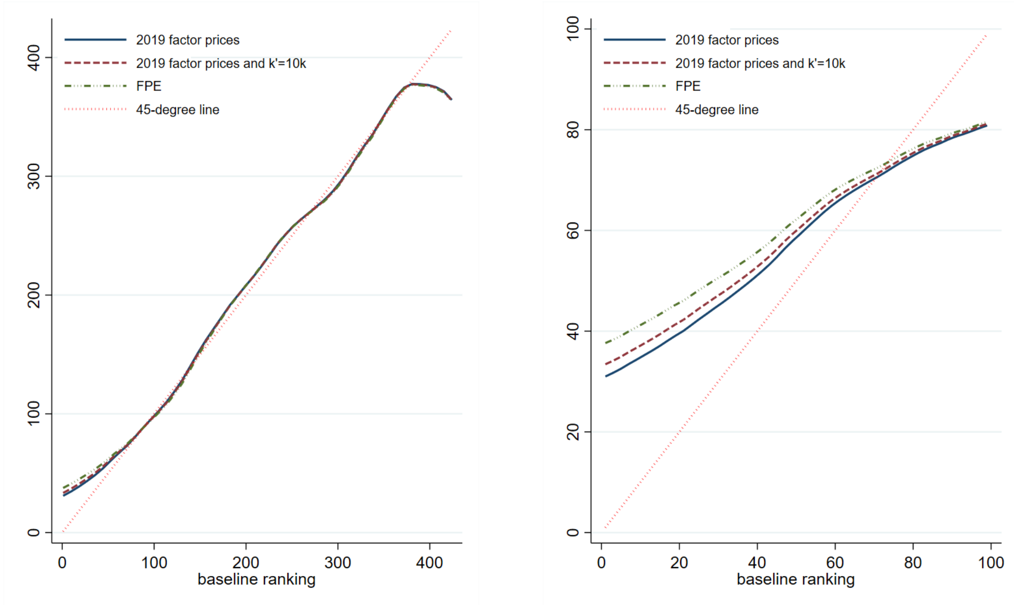
Figure 2: Quantification of Ricardian Comparative Advantage



Notes: These figures plot China's estimated comparative advantage against export propensity across 4-digit CIC industries. The dashed lines represent the best-fitting cubic polynomial lines. Figure (a) shows the ex-post component of the relative TFPQ ($XPA(z)$ in equation 22). Figure (b) displays the endogenous component of relative TFPQ ($XPA(z) \cdot XPD(z)$ in equation 22). Figure (c) displays China's estimated input price relative to RoW. Figure (d) displays China's estimated state of technology relative to RoW (blue dots and long dashed lines), and ex-ante Ricardian CA (orange dots and short dashed lines). Figure (e) plots the estimated relative productivity between China and RoW in terms of relative inverse average UIR, while figure (f) shows China's TFPQ relative to RoW.

Figure 4: Counterfactual Ranking of Industries by Balassa's RCA index

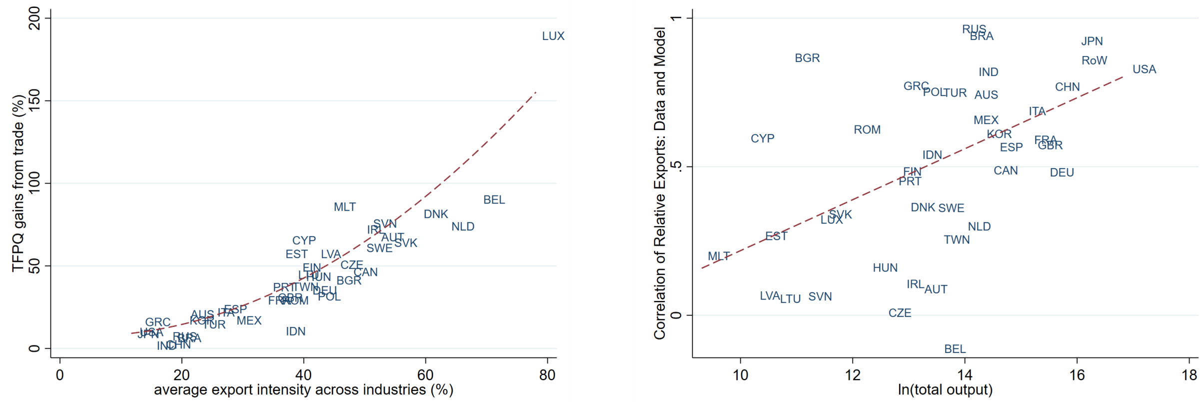
(a) All industries　　(b) Top 100 industries



Notes: The diagram shows local polynomial regression plots comparing the counterfactual ranking of industries based on Balassa's RCA index against their baseline ranking in 2004. Figure (a) represents the analysis of the entire industry sample, while Figure (b) focuses on the top 100 industries. The scenario "2019 factor prices" uses relative factor prices of 2019, while "2019 factor prices and $k' = 10k$" further includes a Pareto shape 10 times as large as the baseline, and "FPE" equalizes factor prices between China and the rest of the world.

Figure 5: Macrodata Results

(a) Relative Productivity Gains from Trade　　(b) Relative Export and Total Output



Notes: Figure (a) shows the relationship between industry export intensity and estimated productivity gains from trade, both averaged across industries for each economy. Figure (b) plots correlations of industry ranks of relative exports between the model and data for each economy against its total output (in logs).

Table 1: Motivating Facts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Fact 1								
	Export propensity				Export intensity			
industry labor intensity	1.289^a	1.220^a	1.234^a	0.933^a	1.344^a	1.316^a	1.343^a	1.033^a
	(0.131)	(0.142)	(0.144)	(0.132)	(0.131)	(0.120)	(0.117)	(0.122)
industry TFP	0.0735^a	0.0544^a	0.0498^a	0.0173	0.0942^a	0.0689^a	0.0603^a	0.0268^b
	(0.0158)	(0.0137)	(0.0134)	(0.0122)	(0.0162)	(0.0148)	(0.0144)	(0.0124)
Ownership controls	N	Y	Y	Y	N	Y	Y	Y
Average firm age	N	N	Y	Y	N	N	Y	Y
Processing trade	N	N	N	Y	N	N	N	Y
R^2	0.322	0.537	0.543	0.665	0.322	0.560	0.577	0.686
No. of observations	422	422	422	422	422	422	422	422
Panel B. Fact 2								
	Multi-product exporter or not				Number of exported products			
industry labor intensity	0.750^a	0.444^a	0.294^a	0.557^a	6.738^a	5.048^a	4.391^a	3.287^a
	(0.0550)	(0.0299)	(0.0202)	(0.0429)	(0.357)	(0.288)	(0.285)	(0.275)
industry TFP	0.0337^a	0.0113^a	0.0126^a	0.0345^a	0.552^a	0.358^a	0.351^a	0.309^a
	(0.0037)	(0.0021)	(0.0020)	(0.0068)	(0.046)	(0.042)	(0.042)	(0.040)
City FE	N	N	Y	Y	N	N	Y	Y
Ownership FE	N	N	Y	Y	N	N	Y	Y
Firm age & size	N	Y	Y	Y	N	Y	Y	Y
Processing trade	N	Y	Y	Y	N	Y	Y	Y
Sample	full	full	full	matched	full	full	full	matched
R^2	0.0347	0.340	0.387	0.0744	-	-	-	-
No. of observations	227067	227067	227067	41824	226530	226530	226530	41824
Panel C. Fact 3								
	Sales of the second relative to core product				Herfindahl-Hirschman index			
industry labor intensity	1.417^a	1.418^a	1.383^a	1.388^a	-0.542^a	-0.545^a	-0.528^a	-0.531^a
	(0.141)	(0.139)	(0.144)	(0.146)	(0.0529)	(0.0513)	(0.0542)	(0.0548)
industry TFP	0.114^a	0.110^a	0.112^a	0.113^a	-0.0395^a	-0.0377^a	-0.0392^a	-0.0396^a
	(0.022)	(0.023)	(0.022)	(0.022)	(0.0068)	(0.0069)	(0.0069)	(0.0068)
City FE	N	N	Y	Y	N	N	Y	Y
Ownership FE	N	N	N	Y	N	N	N	Y
Firm age & size	N	Y	Y	Y	N	Y	Y	Y
Processing trade	N	Y	Y	Y	N	Y	Y	Y
Sample	matched	matched	matched	matched	matched	matched	matched	matched
R^2	-	-	-	-	0.0346	0.0494	0.0707	0.0722
No. of observations	41809	41809	41809	41809	41824	41824	41824	41824

Notes: Industry labor intensity and TFP (in logarithm) are measured by simple averages across firms within each 4-digit CIC industry. The data sample used in *Panel A* is the 4-digit manufacturing industry. The dependent variable is the fraction of firms that are exporters within each industry in columns (1) - (4), and the fraction of sales exported in columns (5) - (8). *In Panel B*, columns (1) - (3) and (5) - (7) use the full sample of firms in the firm survey in 2004. Columns (4) and (8) use the matched sample between the firm survey and customs data. The dependent variable is a dummy that equals one if the exporter exported more than one HS 8-digit product in columns (1) - (4), and is the number of exported products in columns (5) - (8). Columns (1)-(4) are estimated by OLS, and columns (5)-(8) by PPML. The data sample used in *Panel C* is the matched manufacturing exporters for 2004. The dependent variable is the sales ratio of the second relative to the best-selling product for each firm in columns (1) - (4), and the Herfindahl-Hirschman Index of exports across products in columns (5) - (8). Columns (1)-(4) are estimated by PPML, and columns (5)-(8) by OLS. Robust standard errors are reported in parentheses (clustered at the city level in Panels B and C). Significance levels are indicated by a and b at 0.01 and 0.05, respectively.

Table 2: Sufficient Statistics under Different Data Availability

Object of interest \ Data	Micro data	Macro data
Ricardian Comparative Advantage	$\theta(z)$, ρ, k, $\chi(z)$, $[\frac{\omega(z)}{\omega^*(z)}]$	$\theta(z)$, ρ, k, $\frac{L}{L^*}$, $[\frac{\omega(z)}{\omega^*(z)}]$
Balassa's RCA index	$\theta(z)$, ρ, k, $\chi(z)$, $\frac{I}{I^*}$, $[Exp(z)]$	$\theta(z)$, ρ, $\frac{LI}{L^*I^*}$, $[Exp(z)]$

Notes: $\frac{\omega(z)}{\omega^*(z)}$ is needed to quantify ex-ante Ricardian CA. $Exp(z)$ is needed to pin down the level of Balassa's RCA.

Table 3: China's Relative Productivity Gains from Trade

		Relative productivity (China/RoW)			Gains	
		autarky	open economy			
			TFPQ	inverse average UIR	TFPQ	inverse average UIR
	industry weights	(1)	(2)	(3)	(4)	(5)
(a)	none	89.27%	91.20%	104.25%	2.16%	16.78%
(b)	firm number	83.62%	86.29%	97.92%	3.20%	17.11%
(c)	outputs	84.47%	85.67%	97.28%	1.42%	15.16%

Notes: This table presents the estimated productivity of China relative to the Rest of the World (RoW) in 2004. Industry TFPQ, or firms' average quantity-based total factor productivity, is also the inverse of an industry's output-weighted unit input requirement (UIR). Column (1) estimates the relative productivity under autarky. Column (2) estimates the relative TFPQ in the open economy. Column (3) estimates relative productivity measured by the inverse relative (unweighted) average firm UIR. Column (4) is the difference between column (2) and column (1), i.e., (2) minus (1), column (5) is the difference between column (3) from column (1), i.e., (3) minus (1). Row (a) is a simple average across industries. Row (b) uses the number of firms in each industry as the weight and row (c) uses the outputs of each industry as the weight.

Appendix of

Rethinking Revealed Comparative Advantage with Micro and Macro Data

(For Online Publication)

by

Hanwei Huang, and Gianmarco I.P. Ottaviano

A1 Appendix Proofs

A1.1 Proof of Lemma 1

Let's denote the total variable cost of a firm in industry z with unit input requirement c as $T(z,c)$, then $T(z,c) = Q(z,c)c$. Then the TFPQ of industry z is given by:

$$\begin{aligned}
\overline{\Phi}^A(z) &= \frac{\int_0^{C_D^A(z)} Q(z,c)dG(z,c)/G(C_D^A(z))}{\int_0^{C_D^A(z)} T(z,c)dG(z,c)/G(C_D^A(z))} = \frac{\int_0^{C_D^A(z)} \sum_{m=0}^{M(z,c)-1} q(z,v(m,c))dG(z,c)}{\int_0^{C_D^A(z)} \sum_{m=0}^{M(z,c)-1} q(z,v(m,c))c\xi^{-m}dG(z,c)} \\
&= \frac{\sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} q(z,v(m,c))dG(z,c)}{\sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} q(z,v(m,c))c\xi^{-m}dG(z,c)} \\
&= \frac{\frac{L}{C_M(z)^k}(\frac{\gamma}{\lambda(z)})^{1/\delta} \sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} (\frac{p_{max}(z)}{1-\delta} - \frac{\omega(z)c/\xi^m}{1-\delta})^{-1/\delta} kc^{k-1}dc}{\frac{L}{C_M(z)^k}(\frac{\gamma}{\lambda(z)})^{1/\delta} \sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} (\frac{p_{max}(z)}{1-\delta} - \frac{\omega(z)c/\xi^m}{1-\delta})^{-1/\delta} \xi^{-m}kc^{k}dc} \\
&= \frac{\sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} (p_{max}(z) - \omega(z)c/\xi^m)^{-1/\delta} c^{k-1}dc}{\sum_{m=0}^{\infty} \int_0^{\xi^m C_D^A(z)} (p_{max}(z) - \omega(z)c/\xi^m)\xi^{-m} c^{k}dc}
\end{aligned}$$

Using a change of variable $t \equiv \frac{c}{\xi^m C_D^A(z)}$ inside each integration, we have

$$\overline{\Phi}^A(z) = \frac{C_D^A(z)^{k-\frac{1}{\delta}}(\int_0^1 (1-t)^{-\frac{1}{\delta}} t^{k-1}dt)\sum_{m=0}^{\infty}\xi^{mk}}{C_D^A(z)^{k+1-\frac{1}{\delta}}(\int_0^1 (1-t)^{-\frac{1}{\delta}} t^{k}dt)\sum_{m=0}^{\infty}\xi^{mk}} = \frac{1}{C_D^A(z)}\frac{B(k,1-\frac{1}{\delta})}{B(k+1,1-\frac{1}{\delta})}.$$

We note that the Gamma function satisfies $\Gamma(x+1) = x\Gamma(x)$ and the Beta function satisfies $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, therefore, $\frac{B(k,1-1/\delta)}{B(k+1,1-1/\delta)} = \frac{\Gamma(k)\Gamma(1-1/\delta)}{\Gamma(k+1-1/\delta)}\frac{\Gamma(k+2-1/\delta)}{\Gamma(1-1/\delta)\Gamma(k+1)} = \frac{k+1-1/\delta}{k}$, and

$$\overline{\Phi}^A(z) = \frac{1}{C_D^A(z)}\frac{k+1-1/\delta}{k}.$$

As for the unweighted average firm UIR, it is given by

$$\overline{c}^A(z) = [\int_0^{C_D^A(z)} \sum_{m=0}^{M_D(z,c)-1} \xi^{-m}cdG(z,c)]/G(C_D^A(z))$$

$$= C_D^A(z)^{-k}\sum_{m=0}^{\infty}[\int_0^{C_D^A(z)\xi^m} k\xi^{-m}c^k dc = \frac{k}{k+1}\Psi C_D^A(z).$$

A1.2 Proof of Lemma 2

Home country's export propensity in industry z is

$$\chi(z) = (\frac{C_X(z)}{C_D(z)})^k,$$

where $C_X(z)$ is the unit factor requirement cutoff for export. The model predicts that exports from the home country to the foreign in industry z is

$$\begin{aligned}
Exp(z) &= N_E(z)\int_0^{C_X(z)} r_X(z,c)dG(z,c) = N_E(z)\int_0^{C_X(z)}\sum_{m=0}^{M_X(z,c)-1} r_X(z,v(m,c))dG(z,c) \\
&= N_E(z)\sum_{m=0}^{\infty}\int_0^{\xi^m C_X(z)} r_X(z,v(m,c))dG(z,c) \\
&= N_E(z)\sum_{m=0}^{\infty}\int_0^{\xi^m C_X(z)} (\frac{-\delta p^*_{max}(z)+\tau\omega(z)c/\xi^m}{1-\delta})L^*\left(\frac{\gamma}{\lambda^*(z)}\right)^{\frac{1}{\delta}}\left(\frac{p^*_{max}(z)-\tau\omega(z)c/\xi^m}{1-\delta}\right)^{-\frac{1}{\delta}} dG(z,c) \\
&= \frac{N_E(z)L^*\left(\frac{\gamma}{\lambda^*(z)}\right)^{\frac{1}{\delta}}}{(1-\delta)^{1-\frac{1}{\delta}}}\sum_{m=0}^{\infty}\int_0^{\xi^m C_X(z)} (-\delta p^*_{max}(z)+\tau\omega(z)c/\xi^m)\,(p^*_{max}(z)-\tau\omega(z)c/\xi^m)^{-\frac{1}{\delta}}\,dG(z,c) \\
&= N_E(z)\frac{kL^*(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}\tau\omega(z)C_X(z)^{1+k}}{(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta\int_0^1(1-t)^{-1/\delta}t^{k-1}dt+\int_0^1(1-t)^{-1/\delta}t^k dt\}\sum_{m=0}^{\infty}\xi^{mk} \\
&= N_E(z)\frac{kL^*(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}\tau\omega(z)C_X(z)^{1+k}}{(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta B(k,1-\frac{1}{\delta})+B(k+1,1-\frac{1}{\delta})\}\Psi,
\end{aligned}$$

where we have used $p^*_{\text{max}}(z) = \frac{\alpha}{\lambda^*(z)}$ and $p^*_{max}(z) = \tau\omega(z)C_X(z)$, and a change of variable $t \equiv \frac{c}{\xi^m C_X(z)}$ in the second but last equality. In the last equality, $\Psi = \sum_{m=0}^{\infty}\xi^{mk} = (1-\xi^k)^{-1}$ and the Beta function is defined as $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$. Similarly, domestic sales are given by

$$S_D(z) = N_E(z)\frac{kL(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}\omega(z)C_D(z)^{1+k}}{(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta B(k,1-\frac{1}{\delta})+B(k+1,1-\frac{1}{\delta})\}\Psi.$$

Therefore, export intensity is given by

$$\theta(z) \equiv \frac{Exp(z)}{Exp(z)+S_D(z)} = \frac{L^*\tau C_X(z)^{1+k}}{L^*\tau C_X(z)^{1+k}+LC_D(z)^{1+k}} = \frac{\frac{L^*}{L}\tau\chi(z)^{\frac{k+1}{k}}}{1+\frac{L^*}{L}\tau\chi(z)^{\frac{k+1}{k}}}.$$

We note that $\frac{\partial\theta(z)}{\partial\chi(z)} = \frac{(\frac{L^*}{L}\tau)^2\frac{k+1}{k}\chi(z)^{\frac{k+2}{k}}}{(1+\frac{L^*}{L}\tau\chi(z)^{\frac{k+1}{k}})^2} > 0$, therefore, $\frac{\partial\theta(z)}{\partial z} = \frac{\partial\theta(z)}{\partial\chi(z)}\frac{\partial\chi(z)}{\partial z} \geq 0 \Leftrightarrow \frac{\partial\chi(z)}{\partial z} \geq 0$.

A1.3 Proof of Proposition 2

We notice that the export propensity can be rewritten as

$$\chi(z) = (\frac{C_X(z)}{C_D(z)})^k = (\frac{\omega(z)^* C_D^*(z)}{\tau\omega(z)C_D(z)})^k. \tag{E1}$$

Then $\frac{\partial\chi(z)}{\partial z} \geq 0$ implies $\frac{\partial(\frac{\omega(z)^* C_D^*(z)}{\tau\omega(z)C_D(z)})}{\partial z} \geq 0$. Since $k > 0$, it also implies $\frac{\partial\left(\frac{\omega(z)C_D(z)}{\omega(z)^* C_D^*(z)}\right)^{k+1}}{\partial z} \leq 0$. From equation (18) which determines $C_D(z)$ and $C_D^*(z)$, we know that

$$\left(\frac{\omega(z)C_D(z)}{\omega^*(z)C_D^*(z)}\right)^{k+1} = \frac{L^*}{L}\frac{\omega(z)^{k+1}C_M(z)^k - \rho\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{*k+1}C_M^*(z)^k - \rho\omega(z)^{k+1}C_M(z)^k} = \frac{L^*}{L}\frac{1-\rho\frac{\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}}{\frac{\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k} - \rho}.$$

If we denote $f(z) = \frac{\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}$, we have $\left(\frac{\omega(z)C_D(z)}{\omega^*(z)C_D^*(z)}\right)^{k+1} = \frac{L^*}{L}\frac{1-\rho f(z)}{f(z)-\rho}$. Then

$$\frac{\partial\left(\frac{\omega(z)C_D(z)}{\omega(z)^* C_D^*(z)}\right)^{k+1}}{\partial z} = \frac{L^*}{L}\frac{(\rho^2-1)f'}{(f(z)-\rho)^2} \leq 0, \tag{E2}$$

implies $f' = \frac{\partial f(z)}{\partial z} \geq 0$ given that $\rho \leq 1$. Therefore, we have

$$\frac{\partial\frac{\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}}{\partial z} \geq 0, \tag{E3}$$

which is the result (a).

If we denote $A(z) = \frac{\omega(z)^{*k+1}}{\omega(z)^{k+1}}$ and $B(z) = \frac{C_M^*(z)^k}{C_M(z)^k}$, we find that

$$\frac{\partial\frac{\omega(z)^{*k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}}{\partial z} \geq 0 \Leftrightarrow \frac{\partial A(z)B(z)}{\partial z} \geq 0 \Leftrightarrow \frac{\partial A(z)}{\partial z}B(z) + A(z)\frac{\partial B(z)}{\partial z} \geq 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{\partial \ln A(z)}{\partial \ln z} + \frac{\partial \ln B(z)}{\partial \ln z} \geq 0 \Leftrightarrow (k+1)\frac{\partial \ln\frac{\omega(z)^*}{\omega(z)}}{\partial \ln z} + k\frac{\partial \ln\frac{C_M^*(z)}{C_M(z)}}{\partial \ln z} \geq 0.$$

Define $\varepsilon_\omega = \frac{\partial \ln(\frac{\omega^*(z)}{\omega(z)})}{\partial \ln(z)}$ and $\varepsilon_c = \frac{\partial \ln(\frac{C_M^*(z)}{C_M(z)})}{\partial \ln(z)}$, we immediately have result (b) that

$$\varepsilon_\omega + \frac{k}{k+1}\varepsilon_c \geq 0.$$

A1.4 Proof of Proposition 3

Let us denote the export product scope of a firm with core competency c in industry z as $M_X(z,c)$. It satisfies $M_X(z,c) = \max\{m|\tau\omega(z)v(m,c) \leq \omega^*(z)C_D^*(z)\} + 1$. Since $v(m,c) = \xi^{-m}c$ and

$\xi \in (0,1)$, we have

$$M_X(z,c) = \max\{m | \ln \tau + \ln c + m \ln(\frac{1}{\xi}) \leq \ln \frac{\omega^*(z)}{\omega(z)} C_D^*(z)\} + 1.$$

To match stylized fact 2 that firms' export product scope rises with z, i.e., $\frac{\partial M_X(z,c)}{\partial z} \geq 0$, we need

$$\frac{\partial \frac{\omega^*(z)}{\omega(z)} C_D^*(z)}{\partial z} \geq 0. \tag{E4}$$

Since $C_X(z) = \frac{\omega^*(z) C_D^*(z)}{\tau \omega(z)}$, inequality (E4) also implies that $\frac{\partial C_X(z)}{\partial z} \geq 0$, which equivalent to $\frac{\partial \ln(C_X(z))}{\partial \ln(z)} \geq 0$ or $\varepsilon_{C_X} \geq 0$.

A1.5 Proof of Proposition 4

We consider export revenue for a product m, which is given by

$$r_X(z, v(m,c)) = L^* \left(\frac{\gamma}{\lambda^*(z)}\right)^{\frac{1}{\delta}} \left(\frac{p^*_{max}(z) - \tau\omega(z)v(m,c)}{1-\delta}\right)^{-\frac{1}{\delta}} \frac{-\delta p^*_{max}(z) + \tau\omega(z)v(m,c)}{1-\delta},$$

while $p^*_{max}(z) = \omega^*(z) C_D^*(z)$. For any two products, m and m', such that $m < m'$, we have

$$\begin{aligned}
\frac{r_X(z, v(m,c))}{r_X(z, v(m',c))} &= \left(\frac{p^*_{max}(z) - \tau\omega(z)v(m,c)}{p^*_{max}(z) - \tau\omega(z)v(m',c)}\right)^{-1/\delta} \frac{-\delta p^*_{max}(z) + \tau\omega(z)v(m,c)}{-\delta p^*_{max}(z) + \tau\omega(z)v(m',c)} \\
&= \left(\frac{\frac{p^*_{max}(z)}{\omega(z)} - \tau v(m,c)}{\frac{p^*_{max}(z)}{\omega(z)} - \tau v(m',c)}\right)^{-1/\delta} \frac{\frac{p^*_{max}(z)}{\omega(z)} - \frac{1}{\delta}\tau v(m,c)}{\frac{p^*_{max}(z)}{\omega(z)} - \frac{1}{\delta}\tau v(m',c)}.
\end{aligned}$$

If we denote $P = \frac{p^*_{max}(z)}{\omega(z)}$, $A = \tau v(m,c)$ and $B = \tau v(m',c)$, then $A < B$ since $m < m'$. The expression above can be simplified as

$$\frac{r_X(z, v(m,c))}{r_X(z, v(m',c))} = \left(\frac{P-A}{P-B}\right)^{-1/\delta} \frac{P - \frac{1}{\delta}A}{P - \frac{1}{\delta}B}.$$

Then we have

$$\begin{aligned}
\frac{\partial \ln\left(\frac{r_X(z,v(m,c))}{r_X(z,v(m',c))}\right)}{\partial P} &= (-1/\delta)\left(\frac{1}{P-A} - \frac{1}{P-B}\right) + \frac{1}{P - \frac{1}{\delta}A} - \frac{1}{P - \frac{1}{\delta}B} \\
&= (-1/\delta)(A-B)\left(1 - \frac{1}{\delta}\right) \frac{(A+B)P - (1+\frac{1}{\delta})AB}{(P-A)(P-B)(P-\frac{1}{\delta}A)(P-\frac{1}{\delta}B)}
\end{aligned}$$

Given that $\delta < 0$ and $A < B$, as long as,

$$(A+B)P - (1+\frac{1}{\delta})AB \geq 0, \tag{E5}$$

we have $\frac{\partial \ln(\frac{r_X(z,v(m,c))}{r_X(z,v(m',c))})}{\partial P} \leq 0$ and

$$\frac{\partial \ln(\frac{r_X(z,v(m,c))}{r_X(z,v(m',c))})}{\partial z} = \frac{\partial \ln(\frac{r_X(z,v(m,c))}{r_X(z,v(m',c))})}{\partial P}\frac{\partial P}{\partial z} \leq 0, \tag{E6}$$

given that $\frac{\partial P}{\partial z} = \frac{\partial \frac{\omega^*(z)C_D^*(z)}{\omega(z)}}{\partial z} \geq 0$, as required to match stylized fact 2.

We now examine condition (E5), which implies that

$$\begin{aligned}
(\tau v(m,c) + \tau v(m',c))\frac{p^*_{max}(z)}{\omega(z)} &\geq (1+\frac{1}{\delta})\tau v(m,c)\tau v(m',c) \Leftrightarrow \\
(\xi^{-m}c + \xi^{-m'}c)\frac{p^*_{max}(z)}{\omega(z)} &\geq (1+\frac{1}{\delta})\tau\xi^{-m-m'}c^2 \Leftrightarrow \\
(\xi^m + \xi^{m'})\frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)c} &\geq 1+\frac{1}{\delta}.
\end{aligned}$$

Given that $c \leq C_D(z)$, the inequality above is true as long as

$$(\xi^m + \xi^{m'})\frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)C_D(z)} \geq 1+\frac{1}{\delta}. \tag{E7}$$

Since $\chi(z) = (\frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)C_D(z)})^k$, condition (E7) is equivalent to

$$(\xi^m + \xi^{m'})\chi(z)^{1/k} \geq 1+\frac{1}{\delta}.$$

To match stylized fact 3 that $\frac{r_X(z,v(0,c))}{r_X(z,v(1,c))}$ decreases with z for $m=0$ and $m'=1$, we have

$$(1+\xi)\chi(z)^{1/k} \geq 1+\frac{1}{\delta}.$$

A1.6 Proof of Lemma 3

In the open economy, the TFPQ of industry z in Home is:

$$\begin{aligned}
\overline{\Phi}(z) &= \frac{\int_0^{C_D(z)} Q_D(z,c)dG(z,c)/G(C_D(z)) + \int_0^{C_X(z)} Q_X(z,c)dG(z,c)/G(C_D(z))}{\int_0^{C_D(z)} T_D(z,c)dG(z,c)/G(C_D(z)) + \int_0^{C_X(z)} T_X(z,c)dG(z,c)/G(C_D(z))} \\
&= \frac{\int_0^{C_D(z)} \sum_{m=0}^{M_D(z,c)} q(z,v(m,c))dG(z,c) + \int_0^{C_X(z)} \sum_{m=0}^{M_X(z,c)} q_X(z,v(m,c))dG(z,c)}{\int_0^{C_D(z)} \sum_{m=0}^{M_D(z,c)} q(z,v(m,c))c\xi^{-m}dG(z,c) + \int_0^{C_X(z)} \sum_{m=0}^{M_X(z,c)} q_X(z,v(m,c))\tau c\xi^{-m}dG(z,c)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{m=0}^{\infty}\{L\int_0^{\xi^m C_D(z)}(1-\frac{c}{\xi^m C_D(z)})^{-\frac{1}{\delta}}dG(z,c)+L^*\int_0^{\xi^m C_X(z)}(1-\frac{c}{\xi^m C_X(z)})^{-\frac{1}{\delta}}dG(z,c)\}}{\sum_{m=0}^{\infty}\{\int_0^{\xi^m C_D(z)}L(1-\frac{c}{\xi^m C_D(z)})^{-\frac{1}{\delta}}\frac{c}{\xi^m}dG(z,c)+\int_0^{\xi^m C_X(z)}L^*(1-\frac{c}{\xi^m C_X(z)})^{-\frac{1}{\delta}}\tau\frac{c}{\xi^m}dG(z,c)\}} \\
&= \frac{LC_D(z)^k\Psi+L^*C_X(z)^k\Psi}{LC_D(z)^{k+1}\Psi+\tau L^*C_X(z)^{k+1}\Psi}\frac{B(k,1-1/\delta)}{B(k+1,1-1/\delta)} \\
&= \frac{LC_D(z)^k+\rho L^*(\frac{\omega^*(z)}{\omega(z)})^k C_D^*(z)^k}{LC_D(z)^{k+1}+\rho L^*(\frac{\omega(z)^*}{\omega(z)})^{k+1}C_D^*(z)^{k+1}}\frac{B(k,1-1/\delta)}{B(k+1,1-1/\delta)}. \qquad \text{(E8)}
\end{aligned}
$$

Let's denote $\eta_D(z)=\frac{LC_D(z)^{k+1}}{LC_D(z)^{k+1}+\rho L^*(\frac{\omega(z)^*}{\omega(z)})^{k+1}C_D^*(z)^{k+1}}$ and $\eta_X(z)=\frac{\rho L^*(\frac{\omega^*(z)}{\omega(z)})^{k+1}C_D^*(z)^{k+1}}{LC_D(z)^{k+1}+\rho L^*(\frac{\omega(z)^*}{\omega(z)})^{k+1}C_D^*(z)^{k+1}}$, which are the cost shares for output sold in the domestic and foreign markets, respectively. Then using $\omega^*(z)C_D^*(z)=\tau\omega(z)C_X(z)$, we find that the open economy TFPQ, $\overline{\Phi}(z)$, can be rewritten as

$$
\begin{aligned}
\overline{\Phi}(z) &= \frac{LC_D(z)^k+\rho L^*(\frac{\omega^*(z)}{\omega(z)})^k C_D^*(z)^k}{LC_D(z)^{k+1}+\rho L^*(\frac{\omega(z)^*}{\omega(z)})^{k+1}C_D^*(z)^{k+1}}\frac{B(k,1-1/\delta)}{B(k+1,1-1/\delta)} \\
&= \left(\eta_D(z)\frac{1}{C_D(z)}+\eta_X(z)\frac{1}{\tau C_X(z)}\right)\frac{k+1-1/\delta}{k}.
\end{aligned}
$$

A1.7 Proof of Proposition 5

Substitute the free entry condition (17) into TFPQ of equation (E8), we find that

$$
\overline{\Phi}(z)=\frac{LC_D(z)^k+\rho L^*(\frac{\omega^*(z)}{\omega(z)})^k C_D^*(z)^k}{\frac{f_E}{\beta}(\frac{\alpha}{\gamma})^{1/\delta}\frac{C_M(z)^k}{\Psi}}\frac{B(k,1-1/\delta)}{B(k+1,1-1/\delta)} \qquad \text{(E9)}
$$

Therefore, the relative TFPQ in the open economy between the two countries is

$$
\begin{aligned}
\frac{\overline{\Phi}(z)}{\overline{\Phi}^*(z)} &= \frac{LC_D(z)^k+\rho L^*(\frac{\omega^*(z)}{\omega(z)})^k C_D^*(z)^k}{L^*C_D^*(z)^k+\rho L(\frac{\omega(z)}{\omega^*(z)})^k C_D(z)^k}\frac{C_M^*(z)^k}{C_M(z)^k} \qquad \text{(E10)} \\
&= \frac{LC_M^*(z)^k}{L^*C_M(z)^k}\frac{1+\frac{L^*}{L}\chi(z)}{\rho^{-1}\chi(z)+\rho\frac{L}{L^*}}(\frac{\omega^*(z)}{\omega(z)})^k
\end{aligned}
$$

From equation (E18) in the next proof, we find that

$$
\frac{\omega^*(z)^k}{\omega(z)^k}=\left[\frac{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L}{L^*}+\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}\frac{C_M(z)^k}{C_M^*(z)^k}\right]^{\frac{k}{k+1}}=\left[\frac{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{1+\frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}\frac{L^*}{L}\frac{C_M(z)^k}{C_M^*(z)^k}\right]^{\frac{k}{k+1}},
$$

substituting it back to equation (E10), we find that

$$\frac{\overline{\Phi}(z)}{\overline{\Phi}^*(z)} = \frac{LC_M^*(z)^k}{L^*C_M(z)^k}\frac{1+\frac{L^*}{L}\chi(z)}{\rho^{-1}\chi(z)+\rho\frac{L}{L^*}}\left[\frac{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{1+\frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}\frac{L^*}{L}\frac{C_M(z)^k}{C_M^*(z)^k}\right]^{\frac{k}{k+1}} \tag{E11}$$

$$= \underbrace{\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}}_{ex-ante}\underbrace{\left[\frac{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L^*}{L}\rho+\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}\right]^{\frac{k}{k+1}}}_{ex-post\ component\ 1}\underbrace{\frac{\rho\chi(z)^{-1}+\frac{L^*}{L}\rho}{\rho^{-1}\chi(z)+\frac{L}{L^*}\rho}}_{ex-post\ component\ 2},$$

noting that we have used

$$\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = \frac{C_D^{A*}(z)}{C_D^A(z)} = (\frac{L}{L^*}\frac{C_M^*(z)^k}{C_M(z)^k})^{1/(k+1)},$$

which are straightforward implications of equations (8), (9) and Lemma 1.

It is easy to see that the first ex-post component increases with $\chi(z)$. So, this component amplifies relative TFPQ. In contrast, the second component decreases with $\chi(z)$ and dampens it.

A1.8 Proof of Proposition 6

First, from Lemma 1, it is easy to know that

$$\overline{c}(z) = \int_0^{C_D(z)}\sum_{m=0}^{M_D(z)-1}\nu(m,c)dG(z,c)/G(C_D(z)) = \frac{k}{k+1}\Psi C_D(z), \tag{E12}$$

and

$$\overline{c}^*(z) = \int_0^{C_D^*(z)}\sum_{m=0}^{M_D^*(z)-1}\nu(m,c)dG(z,c)/G(C_D^*(z)) = \frac{k}{k+1}\Psi C_D^*(z). \tag{E13}$$

Hence, we have

$$\frac{\overline{c}(z)^{-1}}{\overline{c}^*(z)^{-1}} = \frac{C_D^*(z)}{C_D(z)}. \tag{E14}$$

Second, according to equation (18), we have

$$\frac{C_D(z)^{1+k}}{C_D^*(z)^{1+k}} = \frac{C_M(z)^k-\rho\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}C_M^*(z)^k}{C_M^*(z)^k-\rho\left(\frac{\omega(z)}{\omega(z)^*}\right)^{k+1}C_M(z)^k}\frac{L^*}{L} \tag{E15}$$

$$= \frac{C_M(z)^k}{C_M^*(z)^k}\frac{L^*}{L}\frac{1-\rho\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\frac{C_M^*(z)^k}{C_M(z)^k}}{1-\rho\left(\frac{\omega(z)}{\omega^*(z)}\right)^{k+1}\frac{C_M(z)^k}{C_M^*(z)^k}}. \tag{E16}$$

Third, from equations (8) and (9), we know that

$$\frac{C_D^A(z)}{C_D^{A*}(z)} = (\frac{L^*}{L}\frac{C_M(z)^k}{C_M^*(z)^k})^{1/(k+1)}.$$

Therefore, equation (E16) can be rewritten as

$$\frac{C_D(z)}{C_D^*(z)} = \frac{C_D^A(z)}{C_D^{A*}(z)}\left(\frac{1-\rho\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\frac{C_M^*(z)^k}{C_M(z)^k}}{1-\rho\left(\frac{\omega(z)}{\omega^*(z)}\right)^{k+1}\frac{C_M(z)^k}{C_M^*(z)^k}}\right)^{1/(k+1)}. \tag{E17}$$

Combing $\chi(z) = (\frac{C_X(z)}{C_D(z)})^k = (\frac{\omega(z)^* C_D^*(z)}{\tau\omega(z) C_D(z)})^k$ and equation (18), we find that

$$\frac{\omega^*(z)^{k+1}}{\omega(z)^{k+1}}\frac{C_M^*(z)^k}{C_M(z)^k} = \frac{\frac{L}{L^*}\rho + \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L}{L^*} + \rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}, \tag{E18}$$

substituting it back to equation (E17), we find that

$$\frac{C_D(z)}{C_D^*(z)} = \frac{C_D^A(z)}{C_D^{A*}(z)}\left(\frac{1-\rho\frac{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L}{L^*}+\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}}{1-\rho\frac{\frac{L}{L^*}+\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L}{L^*}\rho+\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}}\right)^{1/(k+1)} = \frac{C_D^A(z)}{C_D^{A*}(z)}\left(\frac{1+\frac{L}{L^*}\rho\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}{1+\frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}\right)^{\frac{1}{k+1}}.$$

Therefore, the relative productivity between home and foreign in the open economy is

$$\frac{\bar{c}(z)^{-1}}{\bar{c}^*(z)^{-1}} = \frac{C_D^*(z)}{C_D(z)} = \underbrace{\frac{C_D^{A*}(z)}{C_D^A(z)}}_{ex\ ante}\underbrace{\left(\frac{1+\frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{1+\frac{L}{L^*}\rho\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}\right)^{\frac{1}{k+1}}}_{ex\ post}, \tag{E19}$$

of which the *ex post* component increases with $\chi(z)$, given that $\frac{k+1}{k} > 1$.

A1.9 Proof of Proposition 7

According to the proof of Lemma 2, the export intensity is

$$\theta(z) = \frac{\frac{L^*}{L}\tau\chi(z)^{\frac{k+1}{k}}}{1+\frac{L^*}{L}\tau\chi(z)^{\frac{k+1}{k}}}, \tag{E20}$$

from which we can infer that

$$\frac{L}{L^*} = \frac{1-\theta(z)}{\theta(z)}\chi(z)^{\frac{k+1}{k}}\rho^{-1/k}. \tag{E21}$$

Substitute equation (E21) into equation (22), we find

$$\begin{aligned} XPA(z) &= \left[\frac{\frac{L}{L^*}\rho + \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L^*}{L}\rho + \left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}\right]^{\frac{k}{k+1}} = \left[\frac{\chi(z)^{\frac{k+1}{k}}\cdot\frac{1-\theta(z)}{\theta(z)}\rho^{\frac{k-1}{k}} + \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\chi(z)^{-\frac{k+1}{k}}\cdot\frac{\theta(z)}{1-\theta(z)}\rho^{\frac{k+1}{k}} + \left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}\right]^{\frac{k}{k+1}}, \\ &= \left(\frac{\chi(z)}{\rho}\right)^2\left[(1-\theta(z))\,(1+\rho^2\frac{1-\theta(z)}{\theta(z)})\right]^{\frac{k}{k+1}}. \end{aligned}$$

$$\begin{aligned} XPD(z) &= \frac{\rho\chi(z)^{-1} + \frac{L^*}{L}\rho}{\rho^{-1}\chi(z) + \frac{L}{L^*}\rho} = \frac{\rho\chi(z)^{-1} + \rho^{\frac{k+1}{k}}\chi(z)^{-\frac{k+1}{k}}\frac{\theta(z)}{1-\theta(z)}}{\rho^{-1}\chi(z) + \chi(z)^{\frac{k+1}{k}}\frac{1-\theta(z)}{\theta(z)}\rho^{\frac{k-1}{k}}}. \\ &= \frac{\frac{\rho}{\chi(z)} + (\frac{\rho}{\chi(z)})^{\frac{k+1}{k}}\frac{\theta(z)}{1-\theta(z)}}{\frac{\chi(z)}{\rho} + \rho^2(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}\frac{1-\theta(z)}{\theta(z)}} \end{aligned}$$

Therefore, we can quantify the *ex-post* components of Ricardian CA as long as we have data on export propensity $\chi(z)$ and intensity $\theta(z)$, trade freeness ρ and trade elasticity k.

As for the *ex-ante* component, from equation (E18), we have,

$$\frac{C_M^*(z)^k}{C_M(z)^k} = \frac{\omega(z)^{k+1}}{\omega(z)^{*k+1}}\frac{\frac{L}{L^*}\rho + (\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}}{\frac{L}{L^*} + \chi(z)^{\frac{k+1}{k}}\rho^{-1/k}}, \tag{E22}$$

Substituting (E21) into (E22), we find that

$$\frac{C_M^*(z)^k}{C_M(z)^k} = \frac{\omega(z)^{k+1}}{\omega(z)^{*k+1}}\frac{\frac{1-\theta(z)}{\theta(z)}\chi(z)^{\frac{k+1}{k}}\rho + \chi(z)^{\frac{k+1}{k}}\rho^{-1}}{\frac{1-\theta(z)}{\theta(z)}\chi(z)^{\frac{k+1}{k}} + \chi(z)^{\frac{k+1}{k}}} = \frac{\omega(z)^{k+1}}{\omega(z)^{*k+1}}\left(\rho(1-\theta(z)) + \rho^{-1}\theta(z)\right), \tag{E23}$$

therefore,

$$\frac{C_M^*(z)}{C_M(z)} = \left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{k+1}{k}}\left[\rho\,(1-\theta(z)) + \rho^{-1}\theta(z)\right]^{\frac{1}{k}}. \tag{E24}$$

From the proof in A1.7, we know that $\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = (\frac{L}{L^*}\frac{C_M^*(z)^k}{C_M(z)^k})^{1/(k+1)}$, combining it with equation (E21), we have

$$\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} = \left(\frac{C_M^*(z)}{C_M(z)}\right)^{\frac{k}{k+1}}\left(\frac{1-\theta(z)}{\theta(z)}\right)^{\frac{1}{k+1}}\chi(z)^{\frac{1}{k}}\rho^{-\frac{1}{k(k+1)}}. \tag{E25}$$

A1.10 Proof of Corollary 1

Substituting equation (E21) into equation (22) to get rid of $\chi(z)$, we find

$$
\begin{aligned}
XPA(z) &= \left[\frac{\frac{L}{L^*}\rho + \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\frac{L^*}{L}\rho + \left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}}\right]^{\frac{k}{k+1}} = \left[\left(\frac{L}{L^*}\right)^2 \frac{\theta(z)}{\rho}(\rho + \rho^{-1}\frac{\theta(z)}{1-\theta(z)})\right]^{\frac{k}{k+1}}, \\
&= \left[\left(\frac{L}{L^*}\right)^2 \left(\theta(z) + \frac{(\theta(z)/\rho)^2}{1-\theta(z)}\right)\right]^{\frac{k}{k+1}},
\end{aligned}
$$

$$
XPD(z) = \frac{\rho\chi(z)^{-1} + \frac{L^*}{L}\rho}{\rho^{-1}\chi(z) + \frac{L}{L^*}\rho} = \frac{\left(\frac{\theta(z)}{1-\theta(z)}\frac{L}{\rho L^*}\right)^{-\frac{k}{k+1}} + \frac{L^*}{L}\rho}{\left(\frac{\theta(z)}{1-\theta(z)}\frac{L}{\rho L^*}\right)^{\frac{k}{k+1}} + \frac{L}{L^*}\rho}.
$$

Substituting (E24) and (E21) into (E25) and getting rid of $\chi(z)$, we find that

$$
\frac{\overline{\Phi}^{A}(z)}{\overline{\Phi}^{A*}(z)} = \frac{\omega(z)}{\omega^*(z)}\left[\rho\left(1-\theta(z)\right) + \rho^{-1}\theta(z)\right]^{\frac{1}{k+1}}\left(\frac{L}{L^*}\right)^{\frac{1}{k+1}}.
$$

A1.11 Proof of Proposition 8

Using the expression for exports $Exp(z)$ in the proof of Lemma 2, we have

$$
\begin{aligned}
\frac{Exp(z)}{Exp^*(z)} &= \frac{N_E(z)}{N_E^*(z)}\frac{L^*}{L}\frac{\left(\frac{C_X(z)}{C_M}\right)^k}{\left(\frac{C_X^*(z)}{C_M^*}\right)^k}\frac{\tau\omega(z)C_X(z)}{\tau\omega^*(z)C_X^*(z)} \\
&= \frac{N_E(z)}{N_E^*(z)}\frac{L^*}{L}\frac{\left(\frac{C_X(z)}{C_M}\right)^k}{\left(\frac{C_X^*(z)}{C_M^*}\right)^k}\frac{\tau\omega(z)C_X(z)}{\tau\omega^*(z)C_X^*(z)}.
\end{aligned}
$$

Using the cutoff relation

$$
\tau\omega(z)C_X(z) = \omega^*(z)C_D^*(z),
$$

we can further rewrite the above equation as

$$
\frac{Exp(z)}{Exp^*(z)} = \frac{N_E(z)}{N_E^*(z)}\frac{L^*}{L}\frac{\left(\frac{C_X(z)}{C_M}\right)^k}{\left(\frac{C_X^*(z)}{C_M^*}\right)^k}\frac{\omega^*(z)}{\omega(z)}\frac{C_D^*(z)}{C_D(z)}
$$

$$
\begin{aligned}
&= \frac{N_E(z)}{N_E^*(z)} \frac{L^*}{L} \left(\frac{C_X(z)}{C_X^*(z)}\right)^k \left(\frac{C_M^*}{C_M}\right)^k \frac{\omega^*(z)}{\omega(z)} \frac{C_D^*(z)}{C_D(z)} \qquad \text{(E26)} \\
&= \frac{N_E(z)}{N_E^*(z)} \frac{L^*}{L} \left(\frac{\frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)}}{\frac{\omega(z)C_D(z)}{\tau\omega^*(z)}}\right)^k \left(\frac{C_M^*}{C_M}\right)^k \frac{\omega^*(z)}{\omega(z)} \frac{C_D^*(z)}{C_D(z)} \\
&= \frac{N_E(z)}{N_E^*(z)} \frac{L^*}{L} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{2k+1} \left(\frac{C_M^*}{C_M}\right)^k \left(\frac{C_D^*(z)}{C_D(z)}\right)^{k+1}.
\end{aligned}
$$

We next solve for $\frac{N_E(z)}{N_E^*(z)}$. We note that the number of sellers $N(z)$ and $N^*(z)$ are given by:

$$
\begin{aligned}
N(z) &= N_E(z)(\frac{C_D(z)}{C_M(z)})^k + N_E^*(z)(\frac{C_X^*(z)}{C_M^*(z)})^k, \\
N^*(z) &= N_E^*(z)(\frac{C_D^*(z)}{C_M^*(z)})^k + N_E(z)(\frac{C_X(z)}{C_M(z)})^k.
\end{aligned}
$$

using Cramer's rule, we find that

$$
N_E(z) = \frac{N(z)\left(\frac{C_D^*(z)}{C_M^*(z)}\right)^k - N^*(z)\rho\left(\frac{\omega(z)C_D(z)}{\omega^*(z)C_M^*(z)}\right)^k}{(1-\rho^2)\left(\frac{C_D(z)}{C_M(z)}\frac{C_D^*(z)}{C_M^*(z)}\right)^k}
$$

$$
N_E^*(z) = \frac{N^*(z)\left(\frac{C_D(z)}{C_M(z)}\right)^k - N(z)\rho\left(\frac{\omega^*(z)C_D^*(z)}{\omega(z)C_M(z)}\right)^k}{(1-\rho^2)\left(\frac{C_D(z)}{C_M(z)}\frac{C_D^*(z)}{C_M^*(z)}\right)^k}.
$$

hence,

$$
\frac{N_E(z)}{N_E^*(z)} = \frac{\frac{N(z)}{N^*(z)}\left(\frac{C_D^*(z)}{C_M^*(z)}\right)^k - \rho\left(\frac{\omega(z)C_D(z)}{\omega^*(z)C_M^*(z)}\right)^k}{\left(\frac{C_D(z)}{C_M(z)}\right)^k - \frac{N(z)}{N^*(z)}\rho\left(\frac{\omega^*(z)C_D^*(z)}{\omega(z)C_M(z)}\right)^k}. \qquad \text{(E27)}
$$

We still need to find the relative number of sellers $\frac{N(z)}{N^*(z)}$. Similar to Mayer et al. (2014), since firms' core competencies are Pareto-distributed, all varieties will share the same inverse Pareto distribution as $H(c) = \sum_{m=0}^{\infty} G(\xi^m c) = \Psi G(c)$, where $\Psi = (1-\xi^k)^{-1}$. The measure of varieties consumed is proportional to the measure of sellers. Denote the measure of domestic and foreign varieties consumed as $M(z)$ and $M^*(z)$, we have $M(z) = H(C_D(z))N_E(z) + H(C_X^*(z))N_E^*(z) = \Psi N(z)$, and $M^*(z) = H(C_D^*(z))N_E^*(z) + H(C_X^*(z))N_E(z) = \Psi N^*(z)$. Therefore, we have

$$
\frac{N(z)}{N^*(z)} = \frac{M(z)}{M^*(z)}. \qquad \text{(E28)}
$$

To find the relative measure of consumed varieties, we multiply both sides of the equation (2) by $q_i^c(z)$, and integrate across all consumed varieties:

$$\begin{aligned}\int_0^{M(z)} p_i(z)q_i^c(z)di &= \int_0^{M(z)} \frac{\alpha}{\lambda(z)}q_i^c(z)di - \int_0^{M(z)} \frac{\gamma}{\lambda(z)}q_i^c(z)^{1-\delta}di \\ b(z)I &= \frac{\alpha}{\lambda(z)}\int_0^{M(z)} q_i^c(z)di - \frac{\gamma}{\lambda(z)}\int_0^{M(z)} q_i^c(z)^{1-\delta}di.\end{aligned}$$

Plug q_i^c from equation (3) and $p_i(z)$ from equation (4) into the equation above, we find that

$$b(z)I = \frac{\alpha}{\lambda(z)}M(z)\left(\frac{\gamma}{\alpha}\right)^{\frac{1}{\delta}}(1-\delta)^{\frac{1}{\delta}}kB(k,1-\frac{1}{\delta}) - \frac{\gamma}{\lambda(z)}M(z)\left(\frac{\gamma}{\alpha}\right)^{\frac{1-\delta}{\delta}}(1-\delta)^{\frac{1-\delta}{\delta}}kB(k,2-\frac{1}{\delta}),$$

which implies that

$$M(z) = \frac{b(z)I}{\omega(z)C_D(z)(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}k\left((1-\delta)^{\frac{1}{\delta}}B(k,1-\frac{1}{\delta}) - (1-\delta)^{\frac{1-\delta}{\delta}}B(k,2-\frac{1}{\delta})\right)}.$$

Similarly, in the foreign country, we have

$$M^*(z) = \frac{b(z)I^*}{\omega^*(z)C_D^*(z)(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}k\left((1-\delta)^{\frac{1}{\delta}}B(k,1-\frac{1}{\delta}) - (1-\delta)^{\frac{1-\delta}{\delta}}B(k,2-\frac{1}{\delta})\right)},$$

and therefore

$$\frac{N(z)}{N^*(z)} = \frac{M(z)}{M^*(z)} = \frac{I}{I^*}\frac{\omega(z)^*C_D^*(z)}{\omega(z)C_D(z)} = \frac{I}{I^*}(\frac{\chi(z)}{\rho})^{1/k}. \tag{E29}$$

Plug the equation above into equation (E27), we get

$$\frac{N_E(z)}{N_E^*(z)} = \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}}\left(\frac{\omega(z)C_M(z)}{\omega^*(z)C_M^*(z)}\right)^k.$$

Substituting the equation above back to (E26), we find

$$\begin{aligned}\frac{Exp(z)}{Exp^*(z)} &= \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{-k}\left(\frac{C_M^*}{C_M}\right)^{-k}\frac{L^*}{L}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{2k+1}\left(\frac{C_M^*}{C_M}\right)^k\left(\frac{C_D^*(z)}{C_D(z)}\right)^{k+1} \\ &= \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}}\frac{L^*}{L}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\left(\frac{C_D^*(z)}{C_D(z)}\right)^{k+1}\end{aligned}$$

Using equation (E19) to substitute $\frac{C_D^*(z)}{C_D(z)}$, we have

$$\frac{Exp(z)}{Exp^*(z)} = \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}} \frac{L^*}{L} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} \left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)} \left(\frac{1 + \frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{1 + \frac{L}{L^*}\rho\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}} \right)^{\frac{1}{k+1}} \right)^{k+1}$$

$$= \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}} \frac{1 + \frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{1 + \frac{L}{L^*}\rho\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}}} \frac{L^*}{L} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} \left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\right)^{k+1}$$

$$= \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}} \left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}} \frac{1 + \frac{L^*}{L}\rho\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}} + \frac{L}{L^*}\rho} \frac{L^*}{L} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} \left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\right)^{k+1}.$$

From equation (E21), we find that

$$\frac{L^*}{L} = \frac{\theta(z)}{1-\theta(z)}\rho^{\frac{1}{k}}\chi(z)^{-\frac{k+1}{k}} = \frac{\theta(z)}{1-\theta(z)}\rho^{-1+\frac{k+1}{k}}\chi(z)^{-\frac{k+1}{k}} = \frac{\theta(z)}{1-\theta(z)}\rho^{-1}\left(\frac{\chi(z)}{\rho}\right)^{-\frac{k+1}{k}},$$

substituting it into the above equation for relative exports $\frac{Exp(z)}{Exp^*(z)}$, we find that

$$\frac{Exp(z)}{Exp^*(z)} = \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1 - \frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}} \frac{\rho^{-1} + \frac{L^*}{L}\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}} + \frac{L}{L^*}\rho} \frac{\theta(z)}{1-\theta(z)} \left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} \left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\right)^{k+1}. \quad \text{(E30)}$$

Hence, we have established result (a). For result (b), from equation (25), we have

$$\left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\frac{\omega^*(z)}{\omega(z)}\right)^{k+1} = \left[\rho\left(1-\theta(z)\right) + \rho^{-1}\theta(z)\right]\frac{L}{L^*}. \quad \text{(E31)}$$

If we substitute the equation above into equation (E30) and substitutes L/L^* using (E21), then relative exports $\frac{Exp(z)}{Exp^*(z)}$ would just be a function of $\chi(z)$, $\theta(z)$, k, ρ and I/I^*. And if $Exp(z)$ is known, then we can find $Exp^*(z)$ and get result (c) as:

$$RCA_B(z) = \frac{\frac{Exp(z)}{Exp^*(z)}}{1 + \frac{Exp(z)}{Exp^*(z)}} \frac{\int_0^1 Exp(z)dz + \int_0^1 Exp^*(z)dz}{\int_0^1 Exp(z)}.$$

A1.12 Proof of Corollary 2

From equation (E21), we know that

$$(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} = \frac{\theta(z)}{1-\theta(z)}\frac{L}{L^*}\rho^{-1}. \tag{E32}$$

Therefore,

$$\frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1-\frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}} = \frac{\frac{I}{I^*}\chi(z)^{\frac{k+1}{k}}\rho^{-\frac{k+1}{k}} - \rho}{1-\frac{I}{I^*}\chi(z)^{\frac{k+1}{k}}\rho^{-\frac{1}{k}}} = \frac{(\frac{IL}{I^*L^*}\rho^{-1}+\rho)\theta(z)-\rho}{1-(1+\frac{IL}{I^*L^*})\theta(z)}.$$

In addition, using equations (E21) and (E31), we find that

$$\begin{aligned}
&\frac{\rho^{-1}+\frac{L^*}{L}\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}+\frac{L}{L^*}\rho}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{k+1}\left(\frac{\overline{\Phi}^A(z)}{\overline{\Phi}^{A*}(z)}\right)^{k+1} \\
&= \frac{\rho^{-1}+\frac{L^*}{L}\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}}{\left(\frac{\chi(z)}{\rho}\right)^{\frac{k+1}{k}}+\frac{L}{L^*}\rho}\left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right]\frac{L}{L^*} \\
&= \rho^{-1}\frac{\frac{L}{L^*}+\frac{\theta(z)}{1-\theta(z)}\frac{L}{L^*}}{\frac{\theta(z)}{1-\theta(z)}\frac{L}{L^*}\rho^{-1}+\frac{L}{L^*}\rho}\left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right] \\
&= \rho^{-1}\frac{1}{\theta(z)\rho^{-1}+\rho(1-\theta(z))}\left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right] = \rho^{-1}.
\end{aligned}$$

Therefore, we have

$$\frac{Exp(z)}{Exp^*(z)} = \frac{(\frac{IL}{I^*L^*}\rho^{-1}+\rho)\theta(z)-\rho}{1-(1+\frac{IL}{I^*L^*})\theta(z)}\frac{\theta(z)}{1-\theta(z)}\frac{1}{\rho}. \tag{E33}$$

A1.13 Proof of Corollary 3

As the previous proof has shown, relative export satisfies

$$\frac{Exp(z)}{Exp^*(z)} = \frac{\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} - \rho}{1-\frac{I}{I^*}\rho(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}}\frac{\theta(z)}{1-\theta(z)}\frac{1}{\rho}. \tag{E34}$$

It is easy to see that $\frac{\theta(z)}{1-\theta(z)}\frac{1}{\rho}$ increases with $\theta(z)$. Next, we show that the first term in the equation above increases with $\theta(z)$. From equation (E30), given the restriction that $N_E(z)/N_E^*(z) > 0$, it is straightforward to verify that it requires $\rho < \frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}} < \frac{1}{\rho}$. Then, it is easy to see that the first term increases with $\frac{I}{I^*}(\frac{\chi(z)}{\rho})^{\frac{k+1}{k}}$, hence $\chi(z)$, and hence $\theta(z)$ given Lemma 2. Therefore, we have

$$\frac{\partial \frac{Exp(z)}{Exp^*(z)}}{\partial \theta(z)} > 0.$$

Since Balassa's RCA index $RCA_B(z)$ increases with relative export $\frac{Exp(z)}{Exp^*(z)}$, we have

$$\frac{\partial RCA_B(z)}{\partial \theta(z)} = \frac{\partial RCA_B(z)}{\partial \frac{Exp(z)}{Exp^*(z)}} \frac{\partial \frac{Exp(z)}{Exp^*(z)}}{\partial \theta(z)} > 0. \tag{E35}$$

Finally, according to Proposition 2 (a), $\chi(z)$ increases with $\frac{\omega^*(z)^{k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}$. Given Lemma 2, $\theta(z)$ should also increases with $\frac{\omega^*(z)^{k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}$. Therefore, $RCA_B(z)$ should also increase with $\frac{\omega^*(z)^{k+1}C_M^*(z)^k}{\omega(z)^{k+1}C_M(z)^k}$ given that $\frac{\partial RCA_B(z)}{\partial \theta(z)} > 0$.

A2 The Optimization Problem

Consumers' Problem

Suppose there are a number of L_i consumers in country i. They face the following utility maximization problem

$$\max U = \int_0^1 b(z) \ln \left(\alpha \int_0^{N(z)} q_i^c(z) di - \frac{\gamma}{1-\delta} \int_0^{N(z)} q_i^c(z)^{1-\delta} di \right) dz,$$

which is maximized by subjecting to

$$\int_0^1 \int_0^{N(z)} p_i(z) q_i^c(z) di dz = I,$$

while I is the consumer income and $\alpha > 0$ and $\delta < 0$. Consider the following Lagrangian

$$\mathcal{L} = U + \lambda(I - \int_0^1 \int_{i\in\Omega(z)} p_i(z) q_i^c(z) di dz).$$

Denote $u(z) = \alpha \int_0^{N(z)} q_i^c(z) di - \frac{\gamma}{1-\delta} \int_0^{N(z)} q_i^c(z)^{1-\delta} di$, the first order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q_i^c(z)} &= b(z) \frac{1}{u(z)} (\alpha - \gamma q_i^c(z)^{-\delta}) - \lambda p_i(z) = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= I - \int_0^1 \int_{i\in\Omega(z)} p_i(z) q_i^c(z) di dz = 0, \end{aligned}$$

which implies

$$\frac{b(z)}{u(z)}(\alpha - \gamma q_i^c(z)^{-\delta}) = \lambda p_i(z)$$

$$\alpha - \gamma q_i^c(z)^{-\delta} = \frac{\lambda u(z)}{b(z)} p_i(z).$$

Define $\lambda(z) = \frac{\lambda u(z)}{b(z)}$, it can be rewritten as

$$p_i(z) = \frac{\alpha}{\lambda(z)} - \frac{\gamma}{\lambda(z)} q_i^c(z)^{-\delta}. \tag{E36}$$

At the choke price $p_{max}(z) = \frac{\alpha}{\lambda(z)}$, we have $q_i^c(z) = 0$. For a market with L consumers, the Marshallian market demand for a variety i is

$$q_i(z) = L\left(\frac{\gamma}{\lambda(z)}\right)^{\frac{1}{\delta}} (p_{max}(z) - p_i(z))^{-\frac{1}{\delta}}.$$

Firms' Problem

Suppose the price of a unit factor bundle in sector z is $\omega(z)$ and a firm with unit factor requirement c has marginal cost of $\omega(z)c$. Then profit maximization

$$\max_{p(z,c)} \pi(z,c) = (p(z,c) - \omega(z)c)q(z) = (p(z,c) - \omega(z)c)L\left(\frac{\gamma}{\lambda(z)}\right)^{\frac{1}{\delta}} (p_{max}(z) - p(z,c))^{-\frac{1}{\delta}}$$

gives

$$p(z,c) = -\frac{\delta}{1-\delta} p_{max}(z) + \frac{1}{1-\delta}\omega(z)c,$$

$$\mu(z,c) = p(z,c) - c = -\frac{\delta}{1-\delta} p_{max}(z) + \frac{\delta}{1-\delta}\omega(z)c,$$

$$q(z,c) = L\left(\frac{\gamma}{\lambda(z)}\right)^{\frac{1}{\delta}} \left(\frac{1}{1-\delta} p_{max}(z) - \frac{1}{1-\delta}\omega(z)c\right)^{-\frac{1}{\delta}},$$

$$\pi(z,c) = -\delta L\left(\frac{\gamma}{\lambda(z)}\right)^{\frac{1}{\delta}} \left(\frac{1}{1-\delta} p_{max}(z) - \frac{1}{1-\delta}\omega(z)c\right)^{1-\frac{1}{\delta}}.$$

A3 The Head-Ries Index

From the proof in A1.2, we know that

$$Exp(z) = N_E(z)\frac{kL^*(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}\tau\omega(z)C_X(z)^{1+k}}{(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta B(k, 1-\frac{1}{\delta}) + B(k+1, 1-\frac{1}{\delta})\}\Psi,$$

as $C_X(z) = \frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)C_D(z)}$, we have

$$\begin{aligned} Exp(z) &= N_E(z)\frac{kL^*(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}\tau\omega(z)(\frac{\omega^*(z)C_D^*(z)}{\tau\omega(z)C_D(z)})^{1+k}}{(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta B(k,1-\frac{1}{\delta})+B(k+1,1-\frac{1}{\delta})\}\Psi, \\ &= \rho N_E(z)\frac{kL^*(\frac{\gamma}{\alpha})^{\frac{1}{\delta}}(\omega^*(z)C_D^*(z))^{1+k}}{\omega(z)^kC_D(z)^{1+k}(1-\delta)^{1-\frac{1}{\delta}}C_M^k(z)}\{-\delta B(k,1-\frac{1}{\delta})+B(k+1,1-\frac{1}{\delta})\}\Psi, \end{aligned}$$

therefore,

$$\frac{Exp(z)}{S_D(z)} = \rho\frac{L^*}{L}(\frac{\omega^*(z)C_D^*(z)}{\omega(z)C_D(z)})^{k+1},$$

while $S_D(z)$ is domestic sales. So we have $\frac{Exp(z)}{S_D(z)}\frac{Exp^*(z)}{S_D^*(z)} = \rho^2$, or

$$\rho = \sqrt{\frac{EXP(z)}{S_D(z)}\frac{EXP^*(z)}{S_D^*(z)}},$$

which is the Head-Ries Index.

A4 The CES Model

This appendix proves the existence of the dampening components of relative TFPQ and derives sufficient statistics for Ricardian comparative advantage with CES demand. The utility function is the same as in the main text after imposing $\alpha = 0$ and $\gamma < 0$ with elasticity of substitution between varieties given by $\sigma = 1/\delta > 1$. Technology is also the same as in the main text except that, in order to have endogenous selection in the domestic and export markets, one has to add fixed input requirements for outputs sold in the domestic and export markets as in Melitz (2003). For simplicity, we abstract from multi-product firms ($\xi \longrightarrow 0$) as we have shown that their explicit consideration is immaterial for relative TFPQ and the sufficient statistics we are after, and assume a common fixed input requirement f for outputs sold in the two markets.

Utility maximization gives the standard CES demand function for a variety produced by a firm with UIR c

$$q(z,c) = Q(z)\left(\frac{p(z,c)}{P(z)}\right)^{-\sigma}, \tag{E37}$$

where

$$Q(z) = \left(\int_{i\in\Omega_z} q(z,i)^{\frac{\sigma-1}{\sigma}}di\right)^{\frac{\sigma}{\sigma-1}} \text{ and } P(z) = \left(\int_{i\in\Omega_z} p(z,i)^{1-\sigma}di\right)^{\frac{1}{1-\sigma}}$$

are the CES composite consumption good and its exact price index, such that $P(z)Q(z) = b(z)R$, with R denoting Home aggregate revenues as well as income.

From profit maximization we obtain that the domestic price $p_D(z,c)$ and the export price $p_D(z,c)$ for a variety produced by a firm with UIR c satisfy

$$p_X(z,c) = \tau p_D(z,c) = \frac{\sigma}{\sigma-1}\tau c\omega(z), \tag{E38}$$

which implies corresponding firm revenues

$$r_D(z,c) = b(z)R\left(\frac{\rho P(z)}{c\varpi(z)}\right)^{\sigma-1}, \tag{E39}$$

$$r_X(z,c) = \tau^{1-\sigma}\left(\frac{P(z)^*}{P(z)}\right)^{\sigma-1}\frac{R^*}{R}r_D(z,c), \tag{E40}$$

and profits

$$\pi_D(z,c) = \frac{r_D(z,c)}{\sigma} - f\omega(z),\ \pi_X(z,c) = \frac{r_X(z,c)}{\sigma} - f\omega(z), \tag{E41}$$

where Foreign variables are starred.

The firm decides to produce for its domestic market if domestic revenue at least covers the domestic fixed cost, and $c = c_D(z)$ such that $\pi_D(z,c) \geq 0$ holds with equality defines the domestic UIR cutoff. Similarly, it exports if its export revenue at least covers its export fixed cost, and $c = c_X(z)$ such that $\pi_X(z,c) \geq 0$ holds with equality defines the export UIR cutoff. If $c \leq c_D(z)$, the firm's total profit $\pi(z,c)$ is then given by:

$$\pi(z,c) = \pi_D(z,c) + \max\{0, \pi_X(z,c)\}. \tag{E42}$$

With free entry, expected profits match the sunk entry cost:

$$\int_0^{c_D(z)} \pi_D(z,c)dG(z,c) + \int_0^{c_X(z)} \pi_X(z,c)dG(z,c) = f_E\omega(z). \tag{E43}$$

Combining the free entry condition (E43) with the zero-profit cutoff conditions, we find that $c_D(z)$ and $c_X(z)$ satisfy

$$f\int_0^{c_D(z)}\left[\left(\frac{c}{c_D(z)}\right)^{1-\sigma} - 1\right]dG(z,c) + f\int_0^{c_X(z)}\left[\left(\frac{c}{c_X(z)}\right)^{1-\sigma} - 1\right]dG(z,c) = f_E. \tag{E44}$$

In equilibrium industry z's revenues $R(z)$ equal the sum of domestic and foreign expenditures on its varieties:

$$R(z) = M(z)\left(\frac{p_D(\widehat{c}_D(z))}{P(z)}\right)^{1-\sigma} b(z)R + \chi(z)M(z)\left(\frac{p_X(\widehat{c}_X(z))}{P^*(z)}\right)^{1-\sigma} b(z)R^*, \tag{E45}$$

where

$$\widehat{c}_t(z) = \left[\frac{1}{G(c_t(z))} \int_0^{c_t(z)} c^{1-\sigma} dG(z,c) \right]^{\frac{1}{1-\sigma}} , \; t \in \{D, X\}$$

defines the industry's output-weighted average UIR for varieties sold in the domestic (D) and the export (X) markets respectively, $M(z)$ and $\chi(z)M(z)$ are the corresponding numbers of varieties with export propensity $\chi(z) = G(z, c_X(z))/G(z, c_D(z))$. Analogous expressions apply to Foreign.

Focusing on incomplete specialization, after defining trade freeness $\rho = \tau^{-k}$ and making the Pareto assumption $G(z,c) = (c/c_M(z))^k$ with $c \in [0, c_M(z)]$ and $k > \sigma - 1$, we can state:

Proposition 9. ***(Relative TFPQ Decomposition with CES)*** *The relative TFPQ of Home industry z can be decomposed as:*

$$\frac{TFPQ(z)}{TFPQ^*(z)} = \underbrace{\frac{c_M^*(z)}{c_M(z)}}_{ex-ante} \underbrace{\frac{1 + \rho^{\frac{1}{k}} \chi(z)^{\frac{k-1}{k}}}{1 + \rho^{\frac{2k-1}{k}} \chi(z)^{-\frac{k-1}{k}}}}_{ex-post \; amplifying \; (xpa(z))} \underbrace{\left(\frac{1 + \rho^2 \chi(z)^{-1}}{1 + \chi(z)} \right)^{\frac{k-1}{k}}}_{ex-post \; dampening \; (xpd(z))} .$$

Proof. Under autarky we have

$$TFPQ^A(z) = \frac{\int_0^{C_D^A(z)} q(z,c) dG(z,c)}{\int_0^{C_D^A(z)} q(z,c) c dG(z,c)} = \frac{k - \sigma + 1}{k - \sigma} \frac{1}{C_D^A(z)},$$

whereas with trade we have

$$\begin{aligned} TFPQ &= \frac{\int_0^{C_D(z)} q_D(z,c) dG(z,c) + \int_0^{C_X(z)} q_X(z,c) dG(z,c)}{\int_0^{C_D(z)} q_D(z,c) c dG(z,c) + \int_0^{C_X(z)} q_X(z,c) \tau c dG(z,c)} \\ &= \frac{k - \sigma + 1}{k - \sigma} \frac{RP(z)^{\sigma-1} c_D(z)^{k-\sigma} + R^* P(z)^{*(\sigma-1)} \tau^{-\sigma} c_X(z)^{k-\sigma}}{RP(z)^{\sigma-1} c_D(z)^{k-\sigma+1} + R^* P(z)^{*(\sigma-1)} \tau^{1-\sigma} c_X(z)^{k-\sigma+1}} \end{aligned} \tag{E46}$$

Zero-profit cutoff conditions imply

$$\begin{aligned} RP(z)^{\sigma-1} &= \frac{\sigma f \omega(z)^\sigma}{b(z) \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} c_D(z)^{1-\sigma}}, \\ R^* P^*(z)^{\sigma-1} &= \frac{\sigma f \omega(z)^\sigma}{b(z) \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1} \tau^{1-\sigma} c_X(z)^{1-\sigma}}. \end{aligned}$$

Substituting these expressions for $RP(z)^{\sigma-1}$ and $R^* P^*(z)^{\sigma-1}$ in equation (E46) gives

$$TFPQ = \frac{k - \sigma + 1}{k - \sigma} \frac{1}{C_D(z)} \frac{1 + \tau^{-1} \chi(z)^{\frac{k-1}{k}}}{1 + \chi(z)}$$

with $\chi(z) = (c_X(z)/c_D(z))^k$, which can be divided by the analogous expression for Foreign to obtain

$$\frac{TFPQ(z)}{TFPQ^*(z)} = \frac{C_D^*(z)}{C_D(z)} \frac{1+\chi^*(z)}{1+\chi(z)} \frac{1+\tau^{-1}\chi(z)^{\frac{k-1}{k}}}{1+\tau^{-1}\chi^*(z)^{\frac{k-1}{k}}}. \tag{E47}$$

Solving the integrals in the free entry condition (E44) yields

$$f\frac{\sigma-1}{k+1-\sigma}\left(\frac{c_D(z)}{c_M(z)}\right)^k + f\frac{\sigma-1}{k+1-\sigma}\left(\frac{c_X(z)}{c_M(z)}\right)^k = f_E, \tag{E48}$$

which can be divided by the analogous expression for Foreign to obtain

$$\left(\frac{c_D^*(z)}{c_D(z)}\right)^k = \left(\frac{c_M^*(z)}{c_M(z)}\right)^k \frac{1+\chi(z)}{1+\chi^*(z)}. \tag{E49}$$

Finally, substituting equation (E49) into (E47) gives

$$\frac{TFPQ(z)}{TFPQ^*(z)} = \frac{c_M^*(z)}{c_M(z)} \frac{1+\tau^{-1}\chi(z)^{\frac{k-1}{k}}}{1+\tau^{-1}\chi^*(z)^{\frac{k-1}{k}}} \left(\frac{1+\chi^*(z)}{1+\chi(z)}\right)^{\frac{k-1}{k}}.$$

so that the proof is completed by noting that we have $\chi^*(z) = \chi(z)^{-1}\tau^{-2k}$ by the cutoff conditions and $\tau = \rho^{-\frac{1}{k}}$ by definition. □

Note that with CES the ex-ante component of relative TFPQ is not endogeous as in the VES case, and is equal to the relative state of technology. Turning to sufficient statistics, we can thus state the following:

Proposition 10. ***(Sufficient Statistics with CES)*** *The Pareto shape parameter k, demand elasticity σ, trade freeness ρ, relative unit input price $\omega(z)/\omega^*(z)$, and export propensity $\chi(z)$ or export intensity $\theta(z)$ are sufficient statistics for the ex-ante component of Home industry z's relative TFPQ as the following equivalent expressions hold:*

$$\frac{c_M^*(z)}{c_M(z)} = \left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{\sigma}{\sigma-1}} \left[\frac{\rho+\rho^{-1}\chi(z)}{1+\chi(z)}\right]^{\frac{1}{k}}$$

and

$$\frac{c_M^*(z)}{c_M(z)} = \left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{\sigma}{\sigma-1}} \left[\rho^{-1}\theta(z)+\rho(1-\theta(z))\right]^{\frac{1}{k}}.$$

Proof. First, consider the export intensity of Home industry z:

$$\theta(z) = \frac{\int_0^{c_X(z)} r_X(z,c)dG(z,c)}{\int_0^{c_D(z)} r_D(z,c)dG(z,c) + \int_0^{c_X(z)} r_X(z,c)dG(z,c)},$$

As industry exports per unit measure of firms are given by

$$\int_0^{c_X(z)} r_X(z,c)dG(z,c) = \sigma f \omega(z) \frac{k}{k-\sigma+1} \left(\frac{c_X(z)}{c_M(z)} \right)^k,$$

while industry domestic sales per unit measure of firms are given by

$$\int_0^{c_D(z)} r_D(z,c)dG(z,c) = \sigma f \omega(z) \frac{k}{k-\sigma+1} \left(\frac{c_D(z)}{c_M(z)} \right)^k,$$

export intensity evaluates to

$$\theta(z) = \frac{\chi(z)}{1+\chi(z)}. \tag{E50}$$

Second, consider the industry's export propensity:

$$\chi(z) = \left(\frac{c_X(z)}{c_D(z)} \right)^k.$$

Home export cutoff $c_X(z)$ satisfies the zero cutoff profit condition for Home exporters

$$r_X(c_X(z)) = b(z)R^* \left(\frac{\rho P^*(z)}{\tau c_X(z)\omega(z)} \right)^{\sigma-1} = \sigma f \omega(z), \tag{E51}$$

while Foreign domestic cutoff $c_D^*(z)$ satisfies the analogous condition for Foreign producers

$$r_D(c_D^*(z)) = b(z)R^* \left(\frac{\rho P^*(z)}{c_D^*(z)\omega^*(z)} \right)^{\sigma-1} = \sigma f \omega^*(z). \tag{E52}$$

Taking the ratio of (E51) to (E52) and solving for $c_D^*(z)$ gives

$$c_D^*(z) = \tau \left(\frac{\omega(z)}{\omega^*(z)} \right)^{\frac{\sigma}{\sigma-1}} c_X(z), \tag{E53}$$

which allows us to rewrite Home export propensity as

$$\chi(z) = \tau^{-k} \left(\frac{\omega^*(z)}{\omega(z)} \right)^{\frac{k\sigma}{\sigma-1}} \left(\frac{c_D^*(z)}{c_D(z)} \right)^k. \tag{E54}$$

As shown in the proof of Proposition 9, the free entry condition implies (E49). Substituting this expression into equation (E54) yields

$$(1+\chi^*(z))\chi(z) = (1+\chi(z)) \left[\tau^{-1} \left(\frac{\omega^*(z)}{\omega(z)} \right)^{\frac{\sigma}{\sigma-1}} \frac{c_M^*(z)}{c_M(z)} \right]^k,$$

which can be restated as

$$\chi(z)=\frac{\left[\frac{c_M^*(z)}{c_M(z)}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{\frac{\sigma}{\sigma-1}}\right]^k-\tau^{-k}}{\tau^k-\left[\frac{c_M^*(z)}{c_M(z)}\left(\frac{\omega^*(z)}{\omega(z)}\right)^{\frac{\sigma}{\sigma-1}}\right]^k} \tag{E55}$$

given that $\chi^*(z)\chi(z)=\tau^{-2k}$ holds. Solving for $c_M^*(z)/c_M(z)$ and using the definition $\rho=\tau^{-k}$ delivers

$$\frac{c_M^*(z)}{c_M(z)}=\left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{\sigma}{\sigma-1}}\left[\frac{\rho^{-1}\chi(z)+\rho}{1+\chi(z)}\right]^{\frac{1}{k}},$$

or equivalently

$$\frac{c_M^*(z)}{c_M(z)}=\left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{\sigma}{\sigma-1}}\left[\rho(1-\theta(z))+\rho^{-1}\theta(z)\right]^{\frac{1}{k}}, \tag{E56}$$

given that (E50) also holds. □

Two comments are in order. Firstly, as for the ex-post components of Home industry z's relative TFPQ, their sufficients statistics exclude $\omega(z)/\omega^*(z)$ as shown by their definitions (see $xpa(z)$ and $xpd(z)$ in Proposition 9). Moreover, by (E50), they can equivalently include either $\chi(z)$ or $\theta(z)$ as in the case of the ex-ante component in Proposition 10. Secondly, one can compare (E56) with the corresponding VES expression from the main text:

$$\frac{C_M^*(z)}{C_M(z)}=\left(\frac{\omega(z)}{\omega^*(z)}\right)^{\frac{k+1}{k}}\left[\rho\left(1-\theta(z)\right)+\rho^{-1}\theta(z)\right]^{\frac{1}{k}}.$$

Taking the log of the ratio of the two expressions gives:

$$\ln\left(\frac{\frac{C_M^*(z)}{C_M(z)}}{\frac{c_M^*(z)}{c_M(z)}}\right)=-\frac{k-\sigma+1}{k\left(\sigma-1\right)}\ln\left(\frac{\omega(z)}{\omega^*(z)}\right),$$

where the slope of the log-linear relation is negative due to $k>\sigma-1$, as required in the CES model. Accordingly, in industries where Home faces a lower [larger] unit input price than Foreign ($\omega(z)/\omega^*(z)<[>]1$), CES demand implies that Home has a smaller [larger] exogenous technological advantage than VES does ($C_M^*(z)/C_M(z)>[<]c_M^*(z)/c_M(z)$). Moreover, the gap is larger the larger the absolute value of $\ln\left(\omega(z)/\omega^*(z)\right)$ and the ratio $(k-\sigma+1)/\left[k\left(\sigma-1\right)\right]$. This last ratio is an increasing concave function of k, and a decreasing convex function of σ. Hence, if demand is VES as argued in the main text, then the CES sufficient statistics give a biased estimate of the relative state of technology, with the direction of the bias determined by whether $\omega(z)/\omega^*(z)$ is larger or smaller than one.

A5 Estimation Appendix

Table T1: Estimation of the Degree of Product Differentiation

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	unit value					
choke price	0.757^a	0.757^a	0.757^a	0.757^a	0.711^a	0.711^a
	(0.0833)	(0.0832)	(0.0832)	(0.0833)	(0.101)	(0.101)
ln (distance)		1069.8^a	1071.8^a			
		(120.9)	(121.3)			
common language		-3324.8^a	-3317.1^a			
		(433.2)	(432.2)			
contiguity		-4664.5^a	-4668.3^a			
		(735.1)	(736.2)			
product rank			3.520^b	4.639^b	-1.407	-2.045
			(1.687)	(1.844)	(2.888)	(2.833)
Destination Fixed Effect	N	N	N	Y	Y	Y
Firm Fixed Effect	N	N	N	N	Y	Y
Product category Fixed Effect	N	N	N	N	N	Y
R^2	0.794	0.795	0.795	0.795	0.836	0.836
# of observations	785481	781982	781982	785481	785481	785481
estimated δ	-3.114^b	-3.118^b	-3.118^b	-3.111^b	-2.461^b	-2.462^b
	(1.408)	(1.411)	(1.411)	(1.407)	(1.205)	(1.206)

Notes: This table estimates exporters' pricing equation using matched 2004 Chinese firm and custom data by ordinary least squares. The dependent variable is the exporters' unit value (value/quantity) at HS 8-digit product and market level. The coefficient of choke price is $\frac{-\delta}{1-\delta}$ (see equation 4), from which we can get the point estimate of δ in the second but last row. The choke price is the maximum unit value among all Chinese exporters at the product (HS 8-digit) and destination level. The gravity variables ("distance", "common language," and "contiguity") are from CEPII (Conte et al., 2022). The variable "product rank" ranks exporters' products by global export sales. The product category fixed effects group products at HS 1-digit level. The numbers in the parentheses are robust standard errors clustered at the firm level. The standard error of δ is estimated using the Delta method: if $\widehat{\phi} \sim N(\phi, \sigma^2)$, then $\widehat{\delta} \sim N\left(\phi/(\phi-1), \sigma^2/(\phi-1)^4\right)$. Significance levels are indicated by a and b at 0.01 and 0.05, respectively.

Table T2: Estimated Pareto Shape

	(1)	(2)	(3)
Dependent variable	ln(CDF of 1/TFP)		
ln (1/TFP)	0.887[a]	1.429[a]	1.310[a]
	(0.00174)	(0.00121)	(0.0320)
industry FE	N	Y	Y
R^2	0.535	0.862	-
No. of observations	225386	225386	225386

Notes: This table estimates the shape parameters of the inverse Pareto distribution, k, with a log-log estimation of the cumulative distribution function (CDF) of the inverse of the estimated TFP. Columns (1) and (2) are estimated by OLS. Column (3) uses the M-estimators of Rousseeuw and Yohai (1984), which is implemented in Stata through the command "sregress." a indicates a significance level of 0.01.

Table T3: Parameters for Quantification Using Microdata

Parameters	Definition	Value
ρ	trade freeness $\rho = \tau^{-k}$	0.0754
w/w^*	the relative wage rate of China (w) and RoW (w^*)	0.367
r/r^*	the relative interest rate of China (r) and RoW (r^*)	1.264
I/I^*	the relative individual income China (I) and RoW (I^*)	0.488

Notes: This table shows the estimated parameters for the quantification by sufficient statistics. All parameters are estimated for the year 2004. Trade freeness ρ is estimated using the Head and Ries (2001) method and the World Input-Output Database for manufacturing sectors. Wage rates, interest rates, and incomes are estimated using Penn World Table 10.0 data. The wage rate is estimated as real GDP multiplied by labor share and divided by total employment. Interest rate is measured by the real internal rate of return. The wage rate for the RoW is an employment-weighted average, and the interest rate is a capital-stock-weighted average across all countries in the sample except China. GDP is measured by output-side real GDP at current PPPs, and capital is measured by capital stock at current PPPs. Income is estimated as GDP per capita. GDP per capita for RoW is population-weighted.

Table T4: Decomposition of the Open-economy Relative Productivity

	Relative inverse UIR		Relative TFPQ			
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	ex-ante	ex-post	ex-ante	ex-post	amplifying	dampening
$\ln(relative\ inverse\ UIR)$	0.753^a	0.247^a				
	(0.0174)	(0.0174)				
$\ln(relative\ TFPQ)$			1.128^a	-0.128^a	3.403^a	-3.530^a
			(0.0106)	(0.0106)	(0.0458)	(0.0502)
R^2	0.911	0.524	0.981	0.392	0.925	0.924
No. of observations	424	424	424	424	424	424

Notes: This table illustrates the decomposition of Ricardian comparative advantage into ex-ante and ex-post components. Columns (1) and (2) decompose Ricardian CA measured by relative inverse average UIR by regressing the ex-ante and ex-post component of inverse average UIR on inverse average UIR (in logs). Columns (3) and (4) decompose Ricardian CA measured by relative TFPQ by regressing the ex-ante and ex-post component of TFPQ on TFPQ (in logs). Columns (5) and (6) further decompose the ex-post component of relative TFPQ into amplifying and dampening components. The numbers in the parentheses are robust standard errors, and "a" indicates a significance level of 0.01.

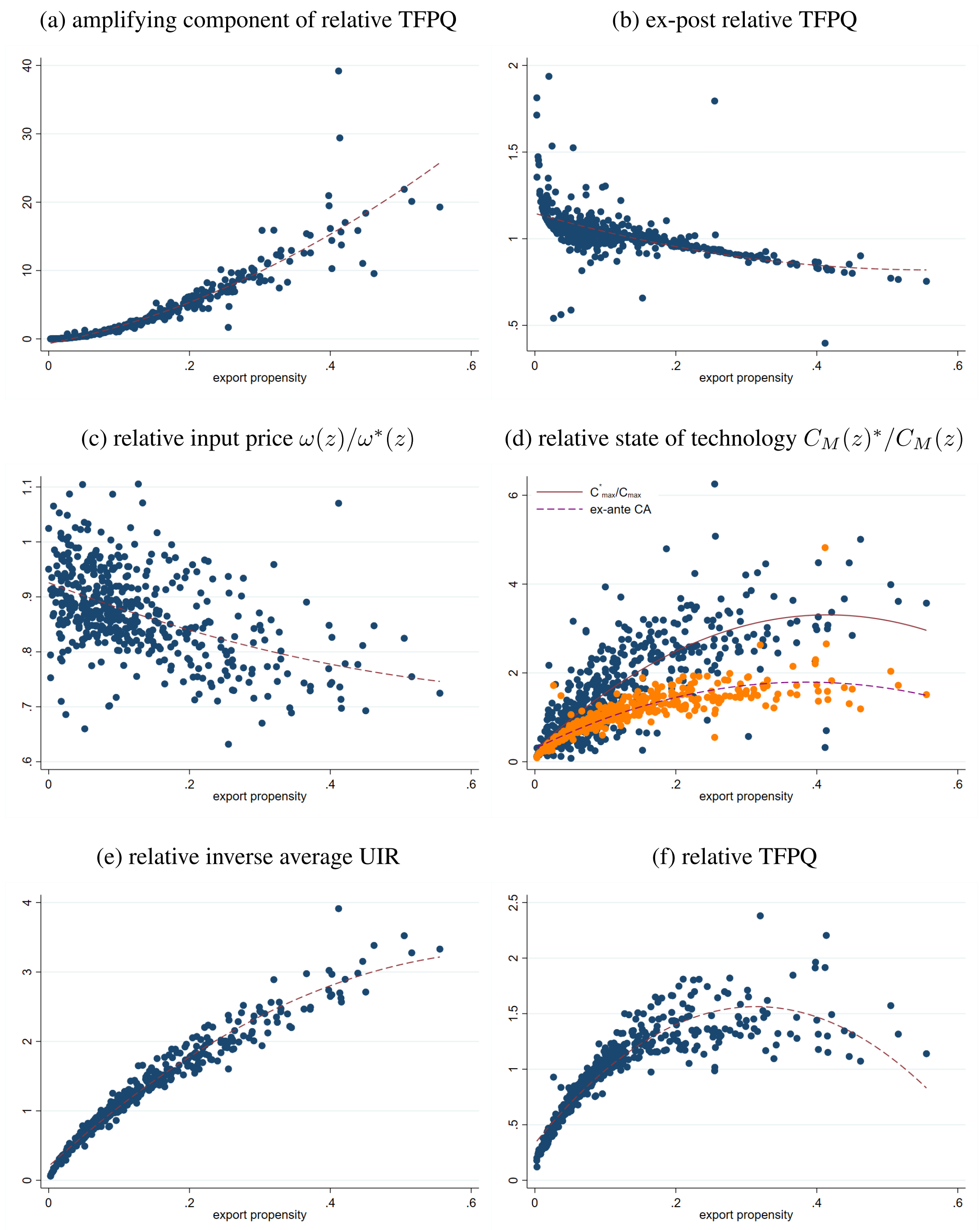
Table T5: Revealed Comparative Advantage and ex-ante Ricardian Comparative of Industries

2004 Baseline (1)	Factor Price Equalization (2)	2019 relative wage and interest rate (3)
	Panel (a): Top 10 industries	
Lacquerware handicraft manufacturing	Container manufacturing [3]	Container manufacturing [3]
Camera and equipment manufacturing	Electronic computer machine manufacturing [39]	Electronic computer machine manufacturing [39]
Container manufacturing	Camera and equipment manufacturing [2]	Camera and equipment manufacturing [2]
Copy and offset printing equipment manufacturing	Copy and offset printing equipment manufacturing [4]	Copy and offset printing equipment manufacturing [4]
Manufacture of garden, art and ceramic products	Manufacturing of slideshow and projection equipment [36]	Manufacture of equipment for open amusement [13]
Manufacture of natural plant fiber weaving crafts	Manufacture of equipment for open amusement [13]	Electronic computer peripheral equipment manufacturing [20]
Toy manufacturing	Electronic computer peripheral equipment manufacturing [20]	Manufacturing of slideshow and projection equipment [36]
Manufacturing of other arts and crafts	Integrated circuit manufacturing [18]	Integrated circuit manufacturing [18]
Manufacturing of flower painting crafts	Pneumatic and power tool manufacturing [23]	Manufacture of natural plant fiber weaving crafts [6]
Hat making	Manufacture of natural plant fiber weaving crafts [6]	Pneumatic and power tool manufacturing [23]
	Panel (b): Bottom 10 industries	
Copper smelting	Lime and gypsum manufacturing [391]	Lime and gypsum manufacturing [391]
Liquor manufacturing	Agriculture & forestry equipment production & repair [380]	Agriculture & forestry equipment production & repair [380]
Manufacturing of other cement products	Manufacture of waterproof building materials [384]	Geological equipment manufacturing [351]
Manufacture of concrete structural components	Cement products manufacturing [389]	Manufacture of waterproof building materials [384]
Manufacture of railway equipment & accessories	Geological equipment manufacturing [351]	Cement products manufacturing [389]
Book, newspaper, journal printing	Manufacturing of bottled (canned) drinking water [383]	Manufacture of clay bricks and building blocks [394]
Coking	Potash manufacturing [382]	Railway equipment manufacturing & repair [390]
Phosphate fertilizer manufacturing	Manufacture of clay bricks and building blocks [394]	Manufacturing of bottled (canned) drinking water [383]
Processing of other tobacco products	Railway equipment manufacturing & repair [390]	Potash manufacturing [382]
Traffic signs & management devices manufacturing	Auto mechanic [387]	Auto mechanic [387]

Notes: This table displays the top and bottom ten industries in 2004 from three different rankings based on the 4-digit Chinese Industry Classification (CIC). The ranking in column (1) is based on the estimated Balassa's RCA index using the sufficient statistics approach. The Pareto shape k, trade freeness ρ, and relative factor input price $\frac{w}{w^*}$ and $\frac{r}{r^*}$ are shown in Table T3, and export propensity, intensity, and volume are the observed value for each industry. In column (2), the ranking of industries is based on the estimated *ex-ante* Ricardian comparative advantage, which is also the ranking of industries with Factor Price Equalization, i.e., factor input prices are the same in China and the rest of the world. In column (3), we used relative wage and interest rates estimated for 2019, $w/w^* = 0.695$ and $r/r^* = 0.953$. The numbers in the square brackets of columns (2) and (3) are the baseline ranking of the industries (there are 424 industries in total).

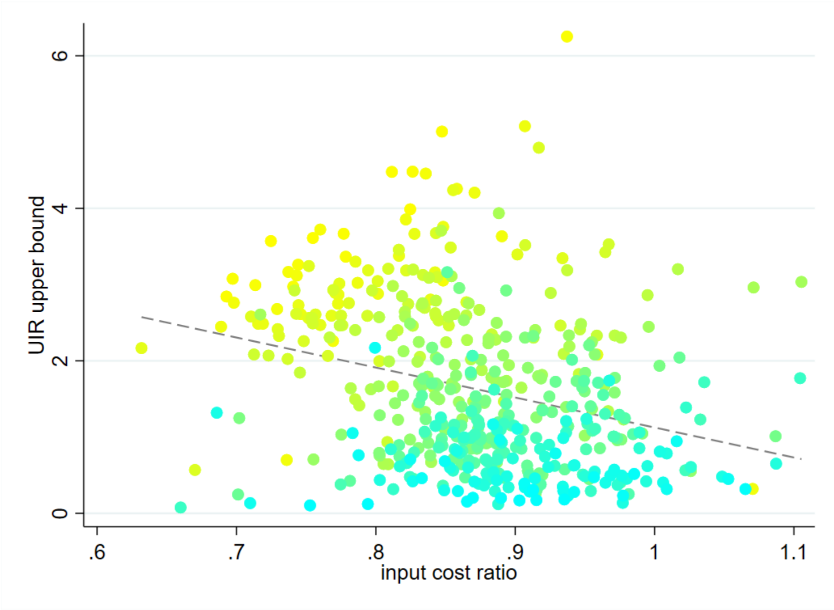
A6 Additional Robustness Figures

Figure F1: Quantification of Ricardian Comparative Advantage: 2004 Census Result



Notes: These figures replicate the Figure 2 using export intensity and export propensity measured from the 2004 Chinese Industrial Census while keeping other parameters the same as Figure 2.

Figure F2: Comparative Advantage and RCA Rankings: 2004 Census Result



Notes: This figure regenerates Figure 3 using export propensity and intensity across 4-digit industries observed in the 2004 Chinese Industrial Census. The horizontal axis represents the input price ratio, which is derived from a Cobb-Douglas combination of the relative wage (w/w^*) and relative interest rate (r/r^*) using the formula $\frac{\omega(z)}{\omega^*(z)} = (\frac{w}{w^*})^z(\frac{r}{r^*})^{1-z}$, where z indicates the average labor intensity of each industry. The vertical axis represents $C_M(z)^*/C_M(z)$, which is estimated through the sufficient statistics approach described in Proposition 7. The dashed line on the graph represents the best-fitting linear relationship between $\omega(z)/\omega^*(z)$ and $C_M(z)^*/C_M(z)$. The color scheme represents Balassa's RCA index (warmer colors represent higher rankings by Balassa's RCA).

A7 Additional Robustness Tables

Table B1: Export Propensity and Export Intensity 2000-2006

Dependent	Export Propensity				Export Intensity			
variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
industry labor intensity	0.794^{a}	1.049^{a}	0.735^{a}	0.729^{a}	0.726^{a}	1.036^{a}	0.693^{a}	0.690^{a}
	(0.0621)	(0.0541)	(0.0501)	(0.0504)	(0.0642)	(0.0503)	(0.0455)	(0.0457)
industry TFP	0.0887^{a}	0.0501^{a}	0.0176^{a}	0.0189^{a}	0.104^{a}	0.0588^{a}	0.0222^{a}	0.0227^{a}
	(0.00609)	(0.00532)	(0.00460)	(0.00470)	(0.00626)	(0.00567)	(0.00465)	(0.00471)
SOE share		-0.439^{a}	-0.288^{a}	-0.275^{a}		-0.527^{a}	-0.338^{a}	-0.337^{a}
		(0.0340)	(0.0306)	(0.0326)		(0.0305)	(0.0265)	(0.0275)
FIE share		1.047^{a}	-0.146^{a}	-0.157^{a}		1.180^{a}	-0.150^{b}	-0.146^{b}
		(0.0503)	(0.0504)	(0.0502)		(0.0561)	(0.0658)	(0.0659)
processing firm share			1.087^{a}	1.101^{a}			1.194^{a}	1.194^{a}
			(0.0332)	(0.0344)			(0.0376)	(0.0381)
ln(average firm age)			-0.00101	0.00277			-0.0162^{b}	-0.0154^{b}
			(0.00604)	(0.00625)			(0.00743)	(0.00779)
Year FE	N	N	N	Y	N	N	N	Y
R^2	0.202	0.462	0.636	0.637	0.180	0.477	0.661	0.661
# of obs	2951	2951	2951	2951	2951	2951	2951	2951

Notes: The sample is the manufacturing firm industries at 4-digit Chinese industry classification from 2000-2006. Industry labor intensity and TFP (in logarithm) are measured by averaging across firms within each 4-digit CIC industry each year. The dependent variable in columns (1) - (4) is the fraction of firms that are exporters within each industry. The dependent variable in columns (5) - (8) is the fraction of sales exported. Robust standard errors are reported in parentheses. Significance levels are indicated by a, and b at 0.01 and 0.05, respectively.

Table B2: Export Product Scope and Product Skewness 2000-2006

	Stylized fact 2: export product scope				Stylized fact 3: export product skewness				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	multi product exporter or not		number of exported product		Herfindal Hirschman Index	second/core export sales	third/core export sales	third/second export sales	core product share in export
industry labor intensity	0.279^a (0.0208)	0.429^a (0.0339)	3.926^a (0.218)	4.427^a (0.253)	-0.474^a (0.0493)	1.202^a (0.130)	2.286^a (0.234)	1.069^a (0.115)	-0.514^a (0.0565)
industry TFP	0.0140^a (0.00205)	0.0236^a (0.00268)	0.373^a (0.0338)	0.403^a (0.0333)	-0.0513^a (0.00569)	0.147^a (0.0165)	0.296^a (0.0301)	0.125^a (0.0143)	-0.0551^a (0.00619)
ln(firm age)	0.00564^a (0.00108)	0.00882^a (0.00119)	0.0847^a (0.0170)	0.109^a (0.0160)	-0.0104^a (0.00275)	0.0272^a (0.00879)	0.0450^a (0.0146)	0.0252^a (0.00693)	-0.0102^a (0.00305)
processing firm	0.565^a (0.00844)	0.519^a (0.00813)	1.683^a (0.0475)	1.379^a (0.0421)	-0.0503^a (0.00475)	0.138^a (0.0126)	0.228^a (0.0247)	0.0804^a (0.0134)	-0.0535^a (0.00545)
ln(sales)	0.0286^a (0.00182)	0.0357^a (0.00228)	0.310^a (0.0135)	0.324^a (0.0151)	-0.0132^a (0.00149)	0.0357^a (0.00424)	0.0447^a (0.00773)	0.0180^a (0.00368)	-0.0129^a (0.00165)
City FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ownership FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sample	full	no unmatched	full	no unmatched	matched	matched	matched	matched	matched
R^2	0.428	0.457	-	-	0.0687	-	-	-	-
# of obs	1269850	1098177	1269237	1097569	228576	228541	228453	178495	228576

Notes: Industry labor intensity and TFP (in logarithm) are measured by averaging across firms within each 4-digit CIC industry each year. In columns (1) to (2), the dependent variable is represented by a dummy variable that takes the value of one if the exporter has exported more than one HS 8-digit product. The dependent variable counts the number of exported products in columns (3) to (4). In column (5), the dependent variable is measured as the Herfindahl-Hirschman Index (HHI) of exports across products. Columns (6) and (7) have dependent variables representing the sales ratio of the second and third best-selling products relative to the best-selling product for each firm, respectively. In column (8), the dependent variable is the sales ratio of the third best-selling product relative to the second best-selling product for each firm. Finally, in column (9), the dependent variable is the share of the best-selling product in the firm's total exports. Columns (1), (2), and (5) are estimated by ordinary least squares, and the remaining columns are estimated by Poisson pseudo maximum likelihood. Columns (1) and (3) use the full sample of firms in the firm survey from 2000-2006 ("full"). Columns (2) and (4) drop firms reporting positive exports in ASIF but cannot be matched with the customs data ("no unmatched"). The remaining columns use the matched sample between the firm survey and customs data ("matched"). We report robust standard errors clustered at the city level in parentheses. The significance levels are denoted by a and b at 0.01 and 0.05, respectively.

Table B3: Estimated Degree of Product Differentiation

	(1)	(2)	(3)	(4)
Dependent variable	unit value			
Estimated δ	-3.107[a]	-3.107^a	-3.076[b]	-3.070[a]
	(1.126)	(1.126)	(1.349)	(1.020)
Gravity Controls	N	Y	Y	Y
Firm Controls	N	N	Y	Y
Product Controls	N	N	N	Y
R^2	0.794	0.794	0.798	0.794
No. of observations	785481	781982	780105	780105

Notes: This table estimates the degree of product differentiation, δ, by non-linear least squares using the pricing equation (4) for exporters using matched firm and custom data in 2004. The dependent variable is the unit value (value/quantity) of an HS 8-digit product exported to a destination market. The gravity controls include the distance, dummies of common language and contiguity, and dummies for each continent. The firm controls include age, industry, province, ownership, wage and interest rates, and an indicator of engaging in processing trade. The product controls include the global rank of a product in a firm's total exports and dummies of HS 1-digit categories. The numbers in the parentheses are robust standard errors. The initial guess of δ in column (1) is -1. The remaining columns use the estimated δ from column (1) as the initial guess. Significance levels are indicated by a and b at 0.01 and 0.05 significance levels, respectively.

Appendix References

Conte, M., P. Cotterlaz and T. Mayer (2022), "The CEPII Gravity Database". CEPII Working Paper N°2022-05, July 2022.

Rousseeuw, P. J., and V. J. Yohai. 1984. Robust Regression by Means of S-Estimators. *In Robust and Nonlinear Time Series Analysis*, ed. J. Franke, W. Härdle, and R. D. Martin, 256–276. New York: Springer.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1963	Natalie Chen Dennis Novy Carlo Perroni Horng Chern Wong	Urban-biased structural change
1962	Robin Kaiji Gong Yao Amber Li Kalina Manova Stephen Teng Sun	Tickets to the global market: First US patent awards and Chinese firm exports
1961	Antonin Bergeaud Arthur Guillouzouic	Proximity of firms to scientific production
1960	Philippe Aghion Antonin Bergeaud Timothee Gigout Matthlieu Lequien Marc Melitz	Exporting ideas: Knowledge flows from expanding trade in goods
1959	Gabriel M. Ahlfeldt Nathaniel Baum-Snow Remi Jedwab	The skyscraper revolution: Global economic development and land savings
1958	David Hémous Simon Lepot Thomas Sampson Julian Schärer	Trade, innovation and optimal patent protection
1957	Maria Cotofan Konstantinos Matakos	Adapting or compounding? The effects of recurring labour shocks on stated and revealed preferences for redistribution
1956	Javad Shamsi	Understanding multi-layered sanctions: A firm-level analysis
1955	Bridget Kauma Giordano Mion	Regional productivity differences in the UK and France: From the micro to the macro

1954	Tim Obermeier	Individual welfare analysis: A tale of consumption, time use and preference heterogeneity
1953	Kirill Borusyak Xavier Jaravel	The distributional effects of trade: Theory and evidence from the United States
1952	Ariela Caglio Sebastien Laffitte Donato Masciandaro Gianmarco Ottaviano	Has financial fair play changed European football?
1951	Eugenie Dugoua Todd D. Gerarden	Induced innovation, inventors and the energy transition
1950	Christian Krekel Johannes Rode Alexander Roth	Do wind turbines have adverse health impacts?
1949	Réka Juhász Claudia Steinwender	Industrial policy and the great divergence
1948	Virginia Minni	Making the invisible hand visible: Managers and the allocation of workers to jobs
1947	Eugenie Dugoua	Induced innovation and international environmental agreements: Evidence from the ozone regime
1946	Marco Bertoni Gabriel Heller-Sahlgren Olmo Silva	Free to improve? The impact of free school attendance in England
1945	Matthias Mertens Bernardo Mottironi	Do larger firms exert more market power? Markups and markdowns along the size of distribution
1944	Ria Ivandić Anne Sophie Lassen	Gender gaps from labor market shocks

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE