A Measure of Countries' Distance to Frontier Based on Comparative Advantage

Ulrich Schetter

CID Research Fellow and Graduate Student
Working Paper No. 135
September 2022

Working Papers
Center for International Development
at Harvard University

A Measure of Countries' Distance to Frontier Based on Comparative Advantage*

Ulrich Schetter

Harvard Kennedy School
Cambridge, MA 02138
ulrich_schetter@hks.harvard.edu

This Version: September 2022

Abstract

This paper presents a structural ranking of countries by their distance to frontier. The ranking is based on comparative advantage. Hence, it reveals information on the productive capabilities of countries that is fundamentally different from GDP per capita. The ranking is centered on the assumption that countries' capabilities across products are similar to those of other countries with comparable distance to frontier. It can be micro-founded using standard trade models. The estimation strategy provides a general, non-parametric approach to uncovering a log-supermodular structure from the data, and I use it to also derive a structural ranking of products by their complexity. The underlying theory provides a flexible micro-foundation for the Economic Complexity Index (Hidalgo and Hausmann, 2009).

Keywords distance to frontier · economic complexity index · gravity model · log-supermodularity · monotonic eigenvector · product complexity · ranking

JEL Classification O11 · O47 · F10 · F14

*I thank Pol Antràs, Jaedo Choi, Davin Chor, Diego Comin, Alje van Dam, Andres Gomez, Matt Grant, Gordon Hanson, Ricardo Hausmann, Will Johnson, Chad Jones, Filipp Levikov, Hannes Malmberg, Marc Muendler, Frank Neffke, Ralph Ossa, Nina Pavčnik, Charly Porcher, Oriol Tejada, David Torun, David Weil and seminar participants at APTS 2021, ESEM 2021, ETSG 2021, VfS 2021, GEOINNO 2022, INFER 2022, MWMM 2022, University of St. Gallen, Harvard Growth Lab, Harvard Intl Lunch, ETH Zurich, and Dartmouth College for valuable comments and suggestions. Financial support from the basic research fund of the University of St. Gallen under grant 1031513 and from the Swiss National Science Foundation under grant IZSEZ0 178724 is gratefully acknowledged. An earlier version of this paper has been circulated with the title 'A Structural Ranking of Economic Complexity'.

1 Introduction

A country's distance to the technological frontier is a core concept in macro-development. It is typically assessed based on a country's GDP per capita. While this provides an important benchmark that entails key information on countries' productive capabilities and welfare, using GDP per capita to measure distance to frontier is not without limitations. For one thing, GDP per capita can be poorly measured and particularly so in developing countries (Deaton and Heston, 2010; Jerven, 2013; Feenstra et al., 2013; Johnson et al., 2013; Subramanian, 2019; Angrist et al., 2021; Martinez, 2022). For another, GDP per capita is not a direct measure of the productive capabilities of countries and it also reflects natural resource rents or whether or not a country is a tax haven, for example. More generally, if we think of convergence as a means of closing gaps in income, we may want to also consider measures that are not based on GDP per capita. In this paper, I propose a theoretically-founded alternative that evaluates a country's distance to the frontier based on its *comparative* advantage.

The basic idea is simple: I start from the assumption that while being close to the frontier may increase productivity across the board, it is more beneficial in some (high complex) products than in other (low complex) products, analogous to a 'technology gap' model of trade (Krugman, 1985). Hence, a country's distance to frontier not only impacts its absolute advantage, but also its comparative advantage. In turn, this allows inferring countries' distance to frontier from the implied pattern of international specialization. To formalize this intuition, I consider a random-utility variant of a multi-product (or industry) Eaton and Kortum (2002)-model similar to Costinot et al. (2012), and assume that fundamental productivities are systematically related to countries' distance to frontier. I then show how countries can be consistently ranked by their distance to frontier based on estimated exporter-product fixed effects from a standard gravity regression. This ranking provides information on countries' productive capabilities that is fundamentally different from their GDP per capita. The ranking is in line with an intuitive notion of distance to frontier, and its main underlying assumption is strongly supported by the data.

In essence, the country ranking measures how similar a country's export basket is to the export basket of frontier economies. Intuitively, if it is true for all countries that their distance to frontier is systematically reflected in their productive capabilities across products then this should in turn imply that a country's capabilities are similar to those of other countries with comparable distance to frontier. My main assumption formalizes this idea. Specifically, let κ_i^s be country i's capability in product s. For every pair of countries i' and i, I consider the similarity of their capabilities across products, $A_{i'i} := \sum_s \kappa_{i'}^s \kappa_i^s$, and assume that this measure of country-country similarity is log-supermodular in countries' distance to frontier. This is a weak assumption and provides a flexible way of introducing a link between a country's distance to frontier and its capabilities across products. It does not involve any functional form assumptions, nor require taking a stance on the type of products that countries close to the frontier are good at producing. I show that this assumption can be micro-founded by assuming that κ_i^s is itself log-supermodular in country capability and, say, product complexity, analogous to Krugman (1985) and Costinot (2009a).

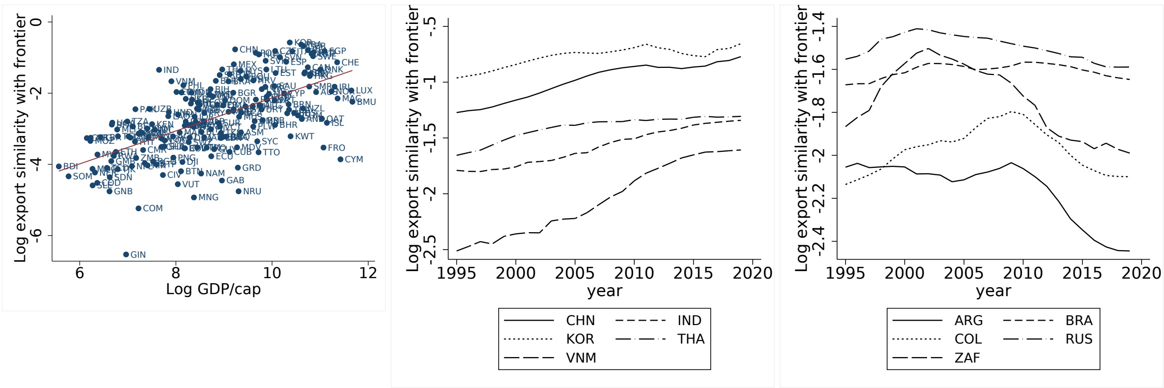
Empirically, deriving the country ranking involves two steps. First, estimating the country-country similarity matrix $\boldsymbol{A}$ with elements $A_{i'i}$ as previously defined. This can be easily achieved using exporter-product fixed effects from a standard gravity regression. Second, ordering matrix $\boldsymbol{A}$ in accordance with the underlying log-supermodularity. Theorem 1 shows how this second step can be achieved using a simple generalized eigenvector. This result is entirely generic and provides a non-parametric approach to ordering 'items'—countries or products in the application here—in accordance with an underlying log-supermodularity based on a matrix of observed 'outcomes'—a country-product matrix of estimated capabilities here. Log-supermodularity is a common assumption in the literature and it is often thought to reflect a complementarity between characteristics that are inherently interesting but difficult to measure. The results presented here provide a simple way of learning about such characteristics without the need of functional form assumptions. Moreover, they provide a flexible micro-foundation for the Economic Complexity Index (ECI, Hidalgo and Hausmann 2009) that accords well with its guiding rationale.

A key premise of this paper is that a country's export mix entails important information about its distance to frontier. Figure 1 suggests that this is indeed the case. It shows countries' export similarity with the frontier—the average of the United States, Japan, and Germany.[1] Export similarity is based on Revealed Comparative Advantages (RCA, Balassa 1965) at the 4-digit HS level, as detailed in the notes of Figure 1. The figure clearly reveals that, on balance, richer countries have an export basket that is more similar to the frontier (panel (a)). Moreover, over time, the export baskets of countries that have been rapidly catching up like China, South Korea, or Vietnam, have become increasingly similar to the frontier (panel (b)). By contrast, the export

[1] I take the average of the US, Japan, and Germany to reduce noise arising from idiosyncratic productivity shocks of the 'frontier' economy and to mitigate concerns of geographic bias. The figures are very similar when using either one of these countries.

Figure 1: Export Similarity with Frontier Basket

(a) Exp. similarity & GDP/cap (b) Evolution of exp. similarity (c) Evolution of exp. similarity



Notes: (a) The figure shows a scatter plot with a country's log GDP per capita in USD on the horizontal axis and a measure of its export similarity with the 'frontier' on the vertical axis: the log cosine-similarity between a country's vector of Revealed Comparative Advantages (RCA, Balassa 1965) across products at the 4-digit HS level and the vector of average RCAs for Germany, Japan, and the United States. The RCA is a country's share in world exports of a product over its share in total world exports (across all products). Data on GDP per capita is from World Bank (2021a). Trade data is from Harvard Growth Lab (2021). All data refer to 2019.
(b), (c) The figure shows for a selection of countries the evolution over time of their export similarity with the frontier.

similarity with the frontier of more stagnant countries like Argentina, Russia, or South Africa remained constant or even declined (panel (c)). I propose a principled way of extracting this information about a country's distance to frontier without the need of ex-ante knowledge about a frontier export basket.

The paper proceeds as follows. Section 2 introduces the economic environment. The country ranking is consistent with the entire class of models that yield gravity equations in expected terms. To introduce the main assumptions in a transparent way, I therefore build on Costinot et al. (2012) and Head and Mayer (2014) and consider a simple random-utility variant of a multi-product (or industry) Eaton and Kortum (2002)-model. This provides a tractable structure that allows exploiting information at the extensive exporter-product margin, the extensive exporter-importer-product margin, and the intensive margin.

Section 3 presents the main theoretical result of the paper, Theorem 1, and provides Monte Carlo Simulations showing its robustness: The generalized eigenvector in Theorem 1 can be used to rank countries by their distance to frontier even if the underlying country-country similarity matrix $\boldsymbol{A}$ is locally only marginally more log-supermodular than an iid random matrix. The basic intuition is that the eigenvector can exploit the log-supermodularity of pairs of elements at greater distances, i.e., in rows and columns that are further apart.

Section 4 discusses how the country-country similarity matrix $\boldsymbol{A}$ can be estimated following standard steps from the literature. Section 5 presents the country ranking. The ranking is very robust and highly correlated with GDP per capita ($\sim .83$). This suggests that there is a tight connection between a country's absolute advantage and its comparative advantage in 'complex' products, in line with a 'technology gap' model (Krugman, 1985). Intuitively, factors that allow countries to develop a comparative advantage in complex products such as skills, technologies, or institutions should also allow them to increase their productivity across all products. In line with this reasoning, the ranking also correlates highly with proxies for potential sources of comparative advantage in 'complex' products. Importantly, however, deviations from the correlation between my ranking and a country's GDP per capita tend to be meaningful: Natural resource rich countries and tax havens tend to be richer than suggested by their distance to frontier. And countries like China, India, or many countries from Eastern Europe or South-East Asia are closer to the frontier than suggested by their GDP per capita. The ranking also reflects well the recent catch-up of these countries.

These patterns suggest that the country ranking does indeed capture important information about countries' distance to frontier. Ultimately, however, this hinges on the validity of the underlying assumptions. Section 5.2 provides evidence in support of these assumptions. It first shows that the estimated exporter-product fixed effects are log-supermodular in country capability and product complexity as discussed momentarily. It then provides evidence for the log-supermodularity of the estimated country-country similarity matrix using a test from Davis and Dingel (2020).

Section 6.1 discusses the connection to the Economic Complexity Index (ECI, Hidalgo and Hausmann 2009). The ECI starts from a country-product matrix $\boldsymbol{M}$ of binarized Revealed Comparative Advantages (RCA, Balassa (1965)), but is then based on the same eigenvector as in Theorem 1—see Hausmann et al. (2011); Caldarelli et al. (2012); Mealy et al. (2019). My results therefore imply that the ECI (PCI) correctly ranks countries (products) by their complexity if the underlying matrix $\boldsymbol{M}$ is log-supermodular. My country ranking provides a structural alternative that exploits information at both the extensive and the intensive margin, and it directly connects the complexity ranking to primitive parameters of my model. Section 6.2 shows how the same reasoning and estimation can be used to derive a structural ranking of products by their complexity. The product ranking is somewhat less robust, which may not come as a surprise given that it uses exports of 130 countries to rank more than

1200 products. But yet, this ranking may serve as an alternative to proxies that have previously been used in the literature. The ranking suggests that the skill-intensity of production is particularly important for the specialization of countries at different distances to frontier.

Lastly, Section 7 concludes.

Related Literature

In terms of the economic environment, the paper is closest related to two strands of literature. First, I consider a 'gravity' framework and extract 'diff-in-diff' productivities—or unit costs—at the country-product level from a fixed effects gravity regression as in Costinot et al. (2012); Hanson et al. (2015); Levchenko and Zhang (2016); Malmberg (2020). My use of these estimated productivities is then, however, very different. I show how they allow to rank countries by their distance to frontier. Second, a large literature in international trade and related fields assumes log-supermodular productivities (e.g. Sattinger 1975; Krugman 1985; Acemoglu et al. 2007; Levchenko 2007; Nunn 2007; Costinot 2009a; Costinot and Vogel 2010; Cuñat and Melitz 2012; Sampson 2014; Grossman et al. 2017; Gaubert 2018; Davis and Dingel 2020)—see Costinot and Vogel (2015) for a review. I deviate from this literature in two ways. On the one hand, these papers typically start from an underlying log-supermodularity and are then interested in learning about the ensuing pattern of specialization and its implications. My main focus is the reverse: I start from the implied pattern of specialization and seek to exploit this in order to learn about the drivers of the underlying log-supermodularity. In that sense, my paper is closer to Chor (2010) and Shikher (2017) who study drivers of comparative advantage, but do not derive structural country rankings.[2] On the other hand, my main assumption generalizes a log-supermodularity between country capability and, say, product complexity, by directly building on the implied pattern of country-country similarity.

The idea to measure countries' economic capabilities based on trade data was introduced in the seminal contribution by Hausmann et al. (2007) and developed further in

[2]Shikher (2017) considers a singular value decomposition of a matrix of estimated productivities from a gravity regression. Considering 53 countries and 15 industries, he shows that the first singular vectors represent well the data. He then studies how these vectors correlate with country and product characteristics, respectively. My work differs along important dimensions. (i) Shikher (2017) motivates his decomposition with log-supermodular productivities, but does not provide an exact mapping to a one-dimensional ranking. In fact, he does not even provide such a country ranking. (ii) The singular vectors in Shikher (2017) are impacted by both absolute and comparative advantage. (iii) I base my analysis directly on the implied pattern of country-country similarities, which allows generalizing the main assumption.

the Economic Complexity Index (ECI, Hidalgo and Hausmann 2009). The ECI spurred a large and growing literature that provides interpretations of the ECI and proposes alternatives (e.g. Tacchella et al. 2012; Caldarelli et al. 2012; Mealy et al. 2019; Sciarra et al. 2020; Bustos and Yildirim 2020; Yildirim 2021; McNerney et al. 2021; van Dam et al. 2021; Gomez-Lievano and Patterson-Lomba 2021; O'Clery et al. 2021)—see Hidalgo (2021) and Balland et al. (2022) for reviews. I contribute to this literature in two ways. On the one hand, this paper provides a theoretical micro-foundation that shows how the ECI can uncover structural notions of 'complexity' from the data. This micro-foundation is very general, and it can readily be used to motivate the application of the same method to other datasets (e.g. Balland and Rigby 2017; Petralia et al. 2017; Mealy et al. 2019). The underlying assumption of log-supermodularity not only provides a weak formalization of the basic idea that 'complex countries make complex products', but it is also nicely in line with the underlying paradigm of the related literature as discussed in Section 6.1. On the other hand, I introduce a notion of country capability into a workhorse gravity model and show how a structural alternative to the ECI can consistently rank countries by their capability in a world with trade frictions. In spite of the substantial differences in the way the underlying country-country similarity matrices are constructed, the derived rankings are highly correlated. Hence, my work also lends additional support to applications of the ECI and the PCI in the literature (e.g. Hausmann et al. 2011; Poncet and Starosta de Waldemar 2013; Maggioni et al. 2016; Hartmann et al. 2017; Javorcik et al. 2018; Gersbach et al. 2019; Lo Turco and Maggioni 2020) and in numerous policy reports, and it may guide the way for more structural applications of these concepts in future.

In contemporaneous work, Atkin et al. (2021) analyze the dynamic gains from specializing in more complex goods. While their main focus is thus different, as part of their analysis they also propose a theoretically-grounded measure of country capability and product complexity based on trade data. My measure differs along two key dimensions. First, Atkin et al. (2021) start from a linear probability model of the extensive margin while I start from a standard gravity regression, which allows bringing the measure closer to earlier work measuring comparative advantage based on trade data. More substantively, it allows exploiting information at both the intensive and the extensive margin. Second, my approach generalizes a complementarity between country capability and product complexity by directly focusing on the implied pattern of country-country similarity.[3] I find strong support for this assumption in the data.

[3] In fact, applying my Proposition 1 and Theorem 1 to exponentiated exporter-product fixed effects from their linear probability model would allow to correctly rank countries in the set-up of Atkin et al. (2021) as well.

At a more general level, the country ranking contributes to the broader literature that provides summarizing statistics on aggregate economic performance (broadly defined) and makes it comparable across countries and over time. This literature includes papers using survey data (e.g. Ravallion 2003; Deaton 2005) or satellite imagery (e.g. Chen and Nordhaus 2011; Henderson et al. 2012; Pinkovskiy and Sala-i Martin 2016; Baragwanath et al. 2021) to measure economic activity, and the large literature on development accounting that decomposes differences in aggregate income (e.g. Hall and Jones 1999; Caselli 2005; Comin and Hobijn 2010; Hsieh and Klenow 2010; Jones 2016; Hendricks and Schoellman 2022; Rossi 2022). I add to this literature by exploiting different information in a principled way: countries' comparative advantages across a broad spectrum of products. The ranking is thus also conceptually very different from e.g. the Global Competitiveness Index (Sala-i Martin and Artadi, 2004) that aggregates a multitude of proxies for drivers of competitiveness.

Lastly, the eigenproblem that is used here to rank countries and related eigenproblems have extensively been studied in the literature—see Fouss et al. (2016) for an overview. Among others, the eigenvector has been proposed as an approximate solution to the Ncut problem of partitioning a graph into clusters (Shi and Malik, 2000), and as a dimensionality reduction algorithm that 'optimally preserves local neighborhood information in a certain sense' Belkin and Niyogi (2003, p. 1374). To my knowledge, the result in Theorem 1 is new and it provides a general, non-parametric way of ordering a log-supermodular adjacency matrix of a unipartite graph—or, when combined with Proposition 1, a bipartite graph.

2 Economic Model

The country ranking proposed here is consistent with the entire class of models that yield gravity equations in expected terms. To introduce the main assumptions, I therefore consider a multi-product (or industry) trade model with heterogeneous consumers similar to Head and Mayer (2014). This model allows for zeros in international trade and, hence, it allows exploiting information at the intensive margin, the extensive exporter-importer-product margin, and the extensive exporter-product margin.

There are I countries, indexed by $i, j \in \mathcal{I}$, and S products, indexed by $s \in \mathcal{S}$.[4]

[4]The literature on international trade typically refers to the upper-tier level of goods-differentiation as industries (or sectors). In my empirical analysis, I consider products at the 4-digit HS level. I therefore refer to the upper-tier level as products and the lower-tier level as varieties of a given product.

Countries differ in their *capability*, which determines their *distance to frontier*, and I will use these terms interchangeably. To simplify notation, I assume that countries are ranked by their capability such that for every $i, i' \in \mathcal{I} : i < i'$ it holds that country i' has higher capability than country i. At times, I will similarly assume that products differ by their complexity, but this will not be essential for the country ranking. When referring to product complexity, I assume that they are ranked in increasing order of complexity. That is, for every $s, s' \in \mathcal{S} : s < s'$ it holds that product s' is more complex than product s. Importantly, however, I think of these characteristics as being unobservable. In fact, I am ultimately interested in finding a way of ranking countries—and products, for that matter—according to these characteristics. Products are differentiated by country of origin as in an Armington (1969) model, but where households have idiosyncratic preferences over these varieties, as detailed in the next section. Trade is subject to an iceberg trade cost such that $d_{ij}^s \geq 1$ units of a variety of product s have to be shipped from country i for one unit to arrive at destination country j. Trade costs satisfy the triangle inequality and $d_{ii}^s = 1$ for all i and s. There is perfect competition in all markets.

2.1 Households and Preferences

Country j is populated by a finite number L_j of households that inelastically supply one unit of labor. Households receive utility from a two-tier utility function, the upper-tier being Cobb-Douglas with product shares α^s, $\sum_{s \in \mathcal{S}} \alpha^s = 1$, and the lower-tier being linear over varieties with household-specific idiosyncratic preferences. Specifically, if household h in country j consumes $q_{ij}^{h,s}$ units of the variety of product s from country i, this translates into

$$\tilde{q}_{ij}^{h,s} = q_{ij}^{h,s} \varphi_{ij}^{h,s}$$

units of effective consumption. $\varphi_{ij}^{h,s}$ is an idiosyncratic preference shock, drawn from a Fréchet distribution with shape parameter θ and location parameter normalized to 1

$$F(\varphi) = \exp\left(-\varphi^{-\theta}\right).$$

Perfect substitutability of varieties of the same product implies that households consume only the variety with lowest price per unit of effective consumption. Hence,

$$q_{ij}^{h,s} p_{ij}^s = \mathbb{1}\left[\frac{p_{ij}^s}{\varphi_{ij}^{h,s}} \leq \min_{\hat{i} \in \mathcal{I}} \left\{\frac{p_{\hat{i}j}^s}{\varphi_{\hat{i}j}^{h,s}}\right\}\right] \alpha^s w_j$$

where p_{ij}^s is the price of product s from country i in destination j and w_j is the wage in country j. $\mathbb{1}[\cdot]$ denotes the indicator function that takes on value of 1 if the term in squared brackets is true and zero otherwise. Note that different households consume different varieties of a product, due to the idiosyncratic preference shocks.

2.2 Production

Production is constant returns to scale using labor as the only input. The constant productivity of a worker in country i when making product s is denoted by T_i^s. This productivity has two components: Country i's fundamental productivity in product s, $\tilde{T}_i^s > 0$, which captures systematic differences arising from countries' distance to frontier. Productivity is further affected by an idiosyncratic component, ϵ_i^s, that captures other sources of productivity differences at the country-product level. This idiosyncratic component is independently distributed across countries and products with strictly positive support

$$T_i^s = \tilde{T}_i^s \epsilon_i^s, \quad \mathbb{E}[\epsilon_i^s] = 1,$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and where the restriction $\mathbb{E}[\epsilon_i^s] = 1$ is without further loss. ϵ_i^s may imply that e.g. a country far from the frontier has a comparative advantage for a complex product.

Zeros are prevalent in international trade. The random utility model allows for zeros at the exporter-importer-product level. To further accommodate the many zeros at the exporter-product level, i.e., to allow for the fact that countries do not at all make certain products, I introduce a binary random variable z_i^s that takes on value of 1 with probability $\tilde{\rho}_i^s > 0$ and 0 otherwise. This variable indicates whether country i has the capability to make product s.

2.3 Main Assumption

The country ranking is centered on the idea that a country's distance to frontier is systematically reflected in its *fundamental capabilities*, $\tilde{T}_i^s$ and $\tilde{\rho}_i^s$. A general way of implementing this idea is by focusing on the implied pattern of country-country similarities as I will now explain, first by introducing my main assumption and then by considering a micro-foundation.

2.3.1 General Case

Suppose that there is a systematic link between a country's distance to frontier and its fundamental capabilities across different products. If this is true for all countries, it should in turn imply that a country's fundamental capabilities are systematically similar to those of countries with comparable distance to frontier. Assumption 1 formalizes this idea. In particular, let $A_{i'i}$ denote the dot-product between the fundamental capabilities of countries i' and i, i.e.,

$$A_{i'i} := \sum_{s \in \mathcal{S}} \tilde{T}_{i'}^{s} \tilde{\rho}_{i'}^{s} \tilde{T}_{i}^{s} \tilde{\rho}_{i}^{s}. \tag{1}$$

With this notation, the main assumption can be summarized as follows:

Assumption 1
For any quadruple of countries, i, i', k, k' such that $i' > i$ and $k' > k$ it holds that

$$\frac{A_{i'k'}}{A_{i'k}} > \frac{A_{ik'}}{A_{ik}}.$$

$A_{i'i}$ is a measure of similarity between countries i' and i. It assigns high values to pairs of countries that have high fundamental capabilities in similar sets of products. According to Assumption 1, this measure of similarity is log-supermodular in countries' distance to frontier. In words, for each country i, Assumption 1 considers its similarity with some country k' relative to some other country k that is further from the frontier, $\frac{A_{ik'}}{A_{ik}}$. The assumption then requires that this ratio is increasing in the capability of country i. That is, the productive capabilities of countries close to the frontier are—in expectation and in a diff-in-diff sense—similar to those of other countries close to the frontier.

Assumption 1 formalizes a general notion of the idea that a country's distance to frontier is systematically reflected in its fundamental capabilities. It is invariant to multiplicative country shifters, i.e., it is about comparative not absolute advantage. Moreover, it does not entail any functional form assumptions nor require taking a stance on the nature of the products that countries close to the frontier are (relatively) good at producing. Yet, it imposes enough structure on the data to allow inferring country-capabilities. I will show this in Section 3 and discuss a flexible micro-foundation for Assumption 1 first.

2.3.2 Special Case: Micro-foundation Based on Krugman (1985); Costinot (2009a)

Krugman (1985) and Costinot (2009a) provide a flexible way of introducing a systematic link between a country's distance to frontier and its fundamental capabilities across products. In line with these papers, suppose for the purpose of this micro-foundation that products differ in their complexity, and that there is a complementarity between a country's capability and a product's complexity. This complementarity operates at both the extensive and the intensive margin. Formally, let $\tilde{T}_i^s$ and $\tilde{\rho}_i^s$ both be log-supermodular.

Assumption 1′ (Micro-foundation for Assumption 1)
For any pair of countries $i' > i \in \mathcal{I}$ and products $s' > s \in \mathcal{S}$ it holds that
(i) $\frac{\tilde{T}_{i'}^{s'}}{\tilde{T}_{i'}^{s}} > \frac{\tilde{T}_{i}^{s'}}{\tilde{T}_{i}^{s}}$ *(ii)* $\frac{\tilde{\rho}_{i'}^{s'}}{\tilde{\rho}_{i'}^{s}} > \frac{\tilde{\rho}_{i}^{s'}}{\tilde{\rho}_{i}^{s}}$

In words, Assumption 1′ implies that countries close to the frontier are relatively—in a diff-in-diff sense—more likely to be able to make the complex products and, if they can, they tend to be relatively more productive. Because this is true for all countries close to the frontier, Assumption 1′ is sufficient for Assumption 1 to hold:

Proposition 1
Let Assumption 1′ be satisfied. Then, $A_{i'i} := \sum_{s\in\mathcal{S}} \tilde{T}_{i'}^s \tilde{\rho}_{i'}^s \tilde{T}_i^s \tilde{\rho}_i^s$ *satisfies Assumption 1.*

Proof: See Appendix A.1. □

It is worth noting that nothing in Proposition 1 is specific to countries or products, and that the result is solely based on the log-supermodularity of $\tilde{T}_i^s \rho_i^s$. So, for example, instead of considering the similarities between pairs of countries, I could consider the similarities between pairs of products in terms of countries' fundamental capabilities for these products

$$B_{s's} := \sum_{i\in\mathcal{I}} \tilde{T}_i^{s'} \tilde{\rho}_i^{s'} \tilde{T}_i^s \tilde{\rho}_i^s. \tag{2}$$

Proposition 1 implies that $\boldsymbol{B}$ is also log-supermodular. Hence, considering these similarities between pairs of products allows deriving a theoretically grounded ranking of products by their complexity. I will get back to this point in Section 6.2. More generally, Proposition 1 shows how a log-supermodular adjacency matrix of a bipartite graph can be reduced to a log-supermodular adjacency matrix of a unipartite graph, which can then be used to order rows and columns by the underlying log-supermodularity. I will show this in Section 3, and consider equilibrium trade-flows first.

2.4 Equilibrium Trade Flows

Markets are perfectly competitive, i.e., all varieties are offered at their marginal cost and consumers in every country shop around the world for the cheapest supplier of each product, given their idiosyncratic demand-shifters. Let $\mathcal{I}^s$ denote the set of countries that can make product s, i.e., $\mathcal{I}^s$ is the set of countries for which $z_i^s = 1$. With a Fréchet distribution of the idiosyncratic preference shocks, the probability that a consumer in country j buys product s from country $i \in \mathcal{I}^s$ is then given by the following well-known expression (cf. Eaton and Kortum 2002; Costinot et al. 2012)

$$\mu_{ij}^s = \frac{\left(w_i d_{ij}^s\right)^{-\theta} T_i^{s\theta}}{\sum_{\hat{i} \in \mathcal{I}^s} \left(w_{\hat{i}} d_{\hat{i}j}^s\right)^{-\theta} T_{\hat{i}}^{s\theta}}.$$

With Cobb-Douglas utility, the total expenditure of a household on a product is independent of where it is buying the product from.[5] In turn, this implies that conditional on T_i^s and z_i^s, country i's total expected sales of product s to country j are given by

$$\mathbb{E}\left[x_{ij}^s | T_i^s, z_i^s\right] = \mathbb{1}\left[z_i^s = 1\right] \frac{\left(w_i d_{ij}^s\right)^{-\theta} T_i^{s\theta}}{\sum_{\hat{i} \in \mathcal{I}_j^s} \left(w_{\hat{i}} d_{\hat{i}j}^s\right)^{-\theta} T_{\hat{i}}^{s\theta}} \alpha^s L_j w_j. \tag{3}$$

In what follows, I use Assumption 1 and Equation (3) to rank countries by their distance to frontier. Deriving this ranking involves two main steps: First, estimating the country-country similarities $A_{i'i}$. Second, ranking countries based on these similarities. I begin with discussing the latter, which represents the main theoretical contribution of the paper.

3 A Non-Parametric Estimator for Ordering Items According to an Underlying Log-supermodularity

In this section, I show how a matrix can be ordered in accordance with an underlying log-supermodularity by solving a generalized eigenproblem. I begin with introducing some definitions before introducing the main result.

3.1 Definitions

The main theoretical result is centered on strictly log-supermodular matrices, which are defined as follows:

Definition 1 (Log-supermodular matrix)
A positive matrix $\boldsymbol{M}$ is strictly log-supermodular if for every pair of rows, $r' > r$, and columns, $c' > c$ it holds that

$$\frac{M_{r'c'}}{M_{r'c}} > \frac{M_{rc'}}{M_{rc}}. \tag{4}$$

[5] With Fréchet preference shocks, this is more generally true in expectation due to the well-known result that the distribution of prices per unit of effective consumption conditional on sourcing from a country is the same, irrespective of the source country (Eaton and Kortum, 2002).

Definition 1 may most easily be understood by means of a simple example. According to this definition, a matrix $\boldsymbol{M}$ is log-supermodular if for every quadruple of elements (a, b, c, d) in the intersections of any pairs of rows and columns

$$\begin{pmatrix} & \vdots & & \vdots & \\ \cdots & a & \cdots & b & \cdots \\ & \vdots & & \vdots & \\ \cdots & c & \cdots & d & \cdots \\ & \vdots & & \vdots & \end{pmatrix}$$

it holds that

$$a \cdot d > b \cdot c.$$

This definition of log-supermodularity is a global property of a matrix. It is satisfied if and only if all 2 by 2 blocks of $\boldsymbol{M}$ are log-supermodular, where, recall, a block of matrix $\boldsymbol{M}$ is defined as follows:

Definition 2 (Block of matrix)
A block of matrix $\boldsymbol{M}$ is a submatrix formed by the elements in the intersection of contiguous rows and columns of $\boldsymbol{M}$.

Lastly, to show how a matrix can be arranged in accordance with the underlying log-supermodularity, it will be convenient to introduce the notion of a monotonic vector:

Definition 3 (Monotonic vector)
A vector $\boldsymbol{v}$ is (strictly) monotonic if its elements are in either (strictly) increasing or (strictly) decreasing order.

3.2 Main Theoretical Result

Suppose for the purpose of this section that it is possible to observe the country-country matrix $\boldsymbol{A}$ as defined in Equation (1). Then, by Assumption 1, the task of ranking countries by their distance to frontier boils down to ordering rows and columns of matrix $\boldsymbol{A}$ in accordance with the underlying log-supermodularity. Theorem 1 shows that this can be accomplished solving a generalized eigenproblem.

Theorem 1
Let $\boldsymbol{A}$ be an $I \times I$ positive and symmetric matrix. Let $\boldsymbol{D}$ be the $I \times I$ diagonal matrix with element D_{ii} equal to the sum of the i^{th} row of $\boldsymbol{A}$, and let $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$ be the Laplacian matrix *of $\boldsymbol{A}$. If $\boldsymbol{A}$ is strictly log-supermodular, then the eigenvector corresponding to the second smallest eigenvalue of*

$$\boldsymbol{L}\boldsymbol{y} = \lambda \boldsymbol{D}\boldsymbol{y} \tag{5}$$

is strictly monotonic.

Proof: See Appendix A.2. □

Theorem 1 provides a general, non-parametric way of ordering rows and columns of a matrix in accordance with an underlying log-supermodular structure. In particular, the second eigenvector of (5) is only monotonic if matrix $\boldsymbol{A}$ is ordered. This is typically not the case, and instead we might have a matrix $\boldsymbol{A}$ that can be made log-supermodular by appropriate permutations of rows and columns. Any such permutation results in the exact same permutation of the elements of the second eigenvector of (5). Hence, rearranging rows and columns of $\boldsymbol{A}$ such that the second eigenvector of (5) is monotonic allows to order matrix $\boldsymbol{A}$ according to the underlying log-supermodular structure.

3.3 Discussion of Theorem 1

The eigenvector corresponding to the second smallest eigenvalue of eigenproblem (5)—which I will henceforth simply refer to as the second eigenvector of (5)—, and eigenproblem (5) more broadly, have extensively been studied as a means of analyzing properties of graphs or networks. Among others, the second eigenvector has been shown to deliver an approximate solution for the 'N-cut' graph partitioning or clustering problem (Shi and Malik, 2000). It has also been proposed as a way of embedding a graph in a lower-dimensional space such that *local* neighborhood information is preserved (Hall, 1970; Belkin and Niyogi, 2003; von Luxburg, 2007). In line with this literature, the country-country similarity matrix $\boldsymbol{A}$ can be seen as an adjacency matrix of a weighted undirected graph that I seek to reduce to a single dimension. In the case considered here, however, I start from a precise notion of what this dimension is and how it affects similarities both locally and globally, i.e., I seek to uncover an underlying log-supermodular structure. To my knowledge, the insight that the second eigenvector of (5) achieves precisely that is new. It does so by exploiting both local and global information, a point I will get back to in Section 3.4.[6]

To get some intuition for why the second eigenvector of (5) is well suited for our purposes, note that it solves the following constrained minimization problem (e.g.

[6] In network analysis, it is often advised to consider similarity matrices that emphasize connections with neighboring nodes, e.g. by considering 'k-nearest neighbors' or 'ϵ-neighborhoods' (Belkin and Niyogi, 2003; von Luxburg, 2007). As opposed to that, I consider a similarity matrix that connects all pairs of countries.

Chung 1997; Shi and Malik 2000; Belkin and Niyogi 2003)

$$\begin{aligned} \arg\min\ & \boldsymbol{y^T L y} \\ s.t.\ & \boldsymbol{y^T D y} = 1 \\ & \boldsymbol{y^T D 1} = 0, \end{aligned} \tag{6}$$

where $\mathbf{1}$ denotes a vector of ones. The objective in this minimization problem can be rewritten as

$$\boldsymbol{y^T L y} = \frac{1}{2}\sum_{ij}(y_i - y_j)^2 A_{ij}, \tag{7}$$

while the constraints essentially rule out trivial solutions.[7] This suggests that the minimization problem tends to assign similar values y_i and y_j to similar countries, i.e., to pairs of countries with large values A_{ij}. By Assumption 1 these tend to be countries with comparable levels of distance to frontier.

To see the potential use of Theorem 1 for other purposes, recall from Proposition 1 that a log-supermodular matrix $\boldsymbol{A}$ can be derived from an underlying log-supermodular bipartite structure.[8] Log-supermodularity is a common assumption in the literature—see literature review. Broadly speaking, it formalizes the idea that a complementarity between, say, a country and a product characteristic gives rise to a chain of comparative advantages. If so, Theorem 1 provides a simple way of learning about the underlying characteristics.

It is well known that in equilibrium Assumption 1′ implies that countries closer to the frontier specialize in complex products. Hence, when equipped with a good measure for product complexity, it would in principle be possible to learn about country capability by simply observing which countries specialize in the complex products. In fact, this idea is not new. It has, for example, been used in ecology at least since the 1950s and is known there as direct gradient analysis—see Whittaker (1967) for an early review. It has also been used in economics by Costinot (2009b) to come up with a measure of 'revealed institutional quality' of countries. The problem with this approach for our purpose here is that any such country ranking would be intimately related to the proxy for product complexity used, and we do not necessarily have good such proxies.

[7] The first constraint rules out solutions where all values of y are zero or arbitrarily close to zero. The second constraint rules out solutions where the entries in y are different from zero but all the same.

[8] A log-supermodular country-country matrix can also be derived from a 'pure ladder economy' (Atkin et al., 2021) or 'nested' pattern of specialization (Hausmann and Hidalgo, 2011; Bustos et al., 2012; Schetter, 2020). That is, it is consistent with set-ups where—in the case considered here—specialization occurs only at the extensive margin, and countries further from the frontier make subsets of the products that countries closer to the frontier are making. Further details are provided in Supplementary Material S3.

Proposition 1 and Theorem 1 show how instead the implied pattern of country-country similarity can be used to learn about country capabilities.[9] This has the additional advantage of relying only on the strictly more general Assumption 1.

Theorem 1 assumes a perfectly log-supermodular matrix, which is unlikely to occur in practical applications. Proposition 1 and, more generally, the construction of the country-country similarity matrix alleviates this concern as it is based on the dot-product of fundamental capabilities over possibly many products, which should reduce noise that may exist at the country-product level. Indeed, the economic model of Section 2 allows for idiosyncratic sources of comparative advantage at both the extensive and the intensive margin. Still, an ensuing question is whether the result in Theorem 1 is robust to deviations from the perfectly log-supermodular structure of matrix $\boldsymbol{A}$, i.e., whether it holds up in situations where Condition (4) is not satisfied everywhere. I turn to this issue next.

3.4 Robustness of Theorem 1: A Monte Carlo Simulation

To evaluate the robustness of Theorem 1 to deviations from perfectly log-supermodular matrices $\boldsymbol{A}$, I perform a Monte Carlo study that involves $80k$ randomly drawn 100×100 matrices—$10k$ for each column in Table 1. Each simulated matrix $\boldsymbol{A}$ starts from a randomly drawn symmetric matrix $\tilde{\boldsymbol{A}}$ that is supermodular. Details on the construction of this matrix are provided in Supplementary Material S2. For the purposes of the discussion here it suffices to note that all 2 by 2 blocks of $\tilde{\boldsymbol{A}}$ are supermodular by a margin that is randomly drawn from a uniform distribution on $[0, 1]$. That is, for every quadruple of elements in a 2 by 2 block of $\tilde{\boldsymbol{A}}$ it holds

$$\tilde{A}_{i'k'} + \tilde{A}_{ik} = \tilde{A}_{ik'} + \tilde{A}_{i'k} + u \qquad (i' > i, k' > k),$$

where u is iid from a uniform distribution with support $[0, 1]$.

In a second step, another symmetric random matrix is added to matrix $\tilde{\boldsymbol{A}}$. Elements of this matrix are drawn from a uniform distribution with lower bound 0 and upper bound ranging from 0 to 500 as specified in the column-headers of Table 1. Finally, the matrix is exponentiated element-wise to get the simulated matrix $\boldsymbol{A}$.

[9] Theorem 1 uses an eigenvector to derive the ranking and, hence, it is defined up to sign only. At a more general level, this is the case because the ranking is based on a similarity matrix, and this similarity does not entail information on the sign of the ranking. In practice, it implies that some additional information is needed to determine the sign of the ranking, i.e., to determine which countries should be ranked atop in the application here. This is arguably less of a concern in practical applications, where the underlying theory readily lends itself to a strong prior regarding the direction of the ranking. In case of the country ranking here, for example, the direction of the ranking can be determined by requiring that industrialized countries be ranked high.

Table 1: Robustness of Monotonicity of Eigenvector

	Upper bound of uniform distribution							
	0.0	0.3	1.0	3.0	10.0	50.0	100.0	500.0
Avg rank correlation	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
Avg share rows/columns correct	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.416
Share of iterations all correct	1.000	1.000	1.000	1.000	1.000	0.997	0.592	0.000
Avg share LSM	1.000	0.930	0.774	0.608	0.533	0.507	0.503	0.501

This table shows summarizing statistics for the $80k$ randomly generated 100×100 matrices—$10k$ for each column. Random matrices have been generated as described in the main text and further detailed in Appendix S2. The random shocks to the supermodular matrix have been drawn from a uniform distribution with support $[0, a]$ where the upper bound a is as specified in the header of the respective column. 'Avg rank correlation' is the average correlation between the ranking implied by the second eigenvector and the 'true' ranking of the random matrix—i.e., a vector with elements $[1, 2, ..., 100]$. The average is taken across the $10k$ random matrices in the respective column. 'Avg share rows/columns correct' is the average share of rows / columns that are ranked exactly correctly by the second eigenvector. 'Share of iterations all correct' is the share of matrices for which the second eigenvector ranks all rows / columns correctly. 'Avg share LSM' is the average share of all 2 by 2 blocks of $\boldsymbol{A}$ that are log-supermodular.

Table 1 reports four average statistics over the $10k$ random matrices in the respective column. First, the rank correlation between the second eigenvector and the 'true' ranking implied by the log-supermodularity of the unshocked matrix ('Avg rank correlation'). Second, the share of all rows/columns that the second eigenvector ranks exactly correctly ('Avg share rows correct'). Third, an indicator whether the second eigenvector ranks all rows/columns exactly correctly ('Share of iterations all correct'). The table finally presents a measure of the importance of the random shocks: the share of all 2 by 2 blocks of the random matrix that are log-supermodular ('Avg share LSM').

The first column in Table 1 shows the benchmark with no random shocks. By construction, all 2 by 2 blocks are log-supermodular and, hence, all random matrices $\boldsymbol{A}$ in this column are log-supermodular. As predicted by Theorem 1, the second eigenvector of (5) always ranks all rows (and columns) correctly.

The remaining columns of Table 1 introduce the random shocks, increasing their variance as we move to the right. Note that in the rightmost columns these shocks are large compared to the margin with which 2 by 2 blocks in matrix $\boldsymbol{A}$ are log-supermodular. Consider, for example, the third to last column. In this column, the random shocks are drawn from a uniform distribution with support $[0, 50]$, implying that—on average across the $10k$ iterations—just over 50% of all 2 by 2 blocks of matrix $\boldsymbol{A}$ are log-supermodular. To put this into perspective, note that for a purely random matrix the expected value of this share is 50%. Nonetheless, the second eigenvector almost always ranks all rows and columns correctly. While this is no longer the case when increasing further the variance of the random shocks, the rank correlation between the second eigenvector and the 'true' ranking is still extremely high.

How is that possible? To see this, note that 'Avg share LSM' limits attention to 2 by 2 blocks of matrix $\boldsymbol{A}$. The fact that these blocks are only marginally more often log-supermodular when compared to an iid matrix does not imply that the same is true for elements in pairs of rows and columns that are further apart. Indeed, quadruples of elements in rows and columns $i, i+10$ and $k, k+10$ are log-supermodular in $\sim 57\%$ of the cases, and quadruples of elements in rows and columns $i, i+30$ and $k, k+30$ in more than 90% of the cases, even when random shocks are drawn from a uniform distribution with support $[0, 500]$. It is this structure of log-supermodularity at greater distances that the second eigenvector can successfully exploit.[10]

In summary, these simulations suggest that the ordering of positive and symmetric log-supermodular matrices is very robust to adding noise. Further details and additional results are provided in Supplementary Material S2.

4 Ranking Countries by their Distance to Frontier

This section discusses how Theorem 1 can be used to rank countries by their distance to frontier. It begins with discussing how matrix $\boldsymbol{A}$ as defined in Equation (4) can be estimated using a two-step estimator. It then discusses the data.

4.1 Two-step Estimator for $\boldsymbol{A}$

In the model of Section 2, conditional on T_i^s and z_i^s, expected tradeflows of product s from i to j, x_{ij}^s, follow a gravity equation. This allows estimating T_i^s using a standard fixed effects regression

$$x_{ij}^s = \exp\left(\delta_{ij} + \delta_j^s + \delta_i^s\right) + \tilde{v}_{ij}^s. \tag{8}$$

In the baseline specification, (8) is estimated using Poisson Pseudo Maximum Likelihood (PPML), which allows for zero trade flows at the bilateral-product level (Wooldridge,

[10] In line with this reasoning, I show in Supplementary Material S2 that smaller matrices, i.e., matrices with less information at greater distances to exploit, are less robust to adding noise. Hence, indeed, for a (noisy) log-supermodular matrix, the eigenvector corresponding to the second smallest eigenvalue of (5) preserves global and not just local neighborhood information, as suggested in Belkin and Niyogi (2003).

1999a,b; Silva and Tenreyro, 2006). From this regression, the estimated T_i^s are derived by exponentiating the exporter-product fixed effects

$$\hat{T}_i^{s,PPML} = \exp(\hat{\delta}_i^{s,PPML}). \tag{9}$$

This allows estimating T_i^s up to a normalizing factor by country and product (Costinot et al., 2012). The normalization does not matter for the asymptotic ability to rank countries based on the estimated matrix $\hat{\boldsymbol{A}}$. It is further discussed in Section 4.2.

As robustness checks, I also estimate T_i^s using (i) OLS for the fixed effects gravity regression, $\hat{T}_i^{s,OLS}$, and (ii) Revealed Comparative Advantages (RCA, Balassa 1965), $\hat{T}_i^{s,RCA}$.[11] Supplementary Material S5 shows that the country ranking is robust to using either of them, with—ceteris paribus—rank correlations between the implied country rankings of around 0.99 across a broad set of robustness checks.

The fixed effects regressions yield estimates of $T_i^s = \tilde{T}_i^s \epsilon_i^s$. The idiosyncratic term ϵ_i^s in these productivities may imply that e.g. a country far from the frontier has a high productivity for a complex product. To rank countries by their distance to frontier, it is therefore necessary to exploit the structure imposed on trade flows by the fundamental productivities, $\tilde{T}_i^s$, and the probabilities of positive exports, $\tilde{\rho}_i^s$. $\boldsymbol{A}$ is based on these fundamental capabilities. This matrix can be estimated using the sample analog. In particular, element $A_{ii'}$ satisfies

$$A_{ii'} := \frac{1}{S} \sum_{s \in \mathcal{S}} \tilde{T}_i^s \rho_i^s \tilde{T}_{i'}^s \rho_{i'}^s = \frac{1}{S} \sum_{s \in \mathcal{S}} \mathbb{E}\left[\tilde{T}_i^s \epsilon_i^s z_i^s \tilde{T}_{i'}^s \epsilon_{i'}^s z_{i'}^s \right].$$

Hence, it can be estimated by

$$\hat{A}_{ii'} = \frac{1}{S} \sum_{s \in \mathcal{S}} z_i^s \hat{T}_i^s z_{i'}^s \hat{T}_{i'}^s,$$

where z_i^s is a binary variable that takes on value of one if country i is exporting product s at all and zero otherwise. Applying Kolmogorov's strong law of large numbers (Sen and Singer, 1993, Theorem 2.3.10), it follows that $\hat{A}_{ii'}$ is a consistent estimator of $A_{ii'}$ given that the first step regression is consistent.[12]

[11] The ranking based on RCAs is consistent with e.g. a Ricardo-Roy model—see Costinot and Vogel (2015, R-R Rybczynski theorem).

[12] This follows from rewriting $\hat{A}_{ii'}$ as

$$\hat{A}_{ii'} = \frac{1}{S} \sum_{s \in \mathcal{S}} [z_i^s z_{i'}^s T_i^s T_{i'}^s] + \frac{1}{S} \sum_{s \in \mathcal{S}} \left[z_i^s z_{i'}^s \left(\hat{T}_i^s \hat{T}_{i'}^s - T_i^s T_{i'}^s \right) \right]$$

and from the Assumption that z_i^s and T_i^s are independently distributed across i and s.

4.2 Data and Estimation

To derive the country rankings, I use data on bilateral trade flows at the product level as provided by the Harvard Growth Lab (2021). This data covers more than 200 countries and is available for the years 1995-2019 at the 4-digit HS0 classification level (1239 products). In the baseline specification, I use data for year 2019. From this data, all importers and exporters are excluded that are not part of the list of 130 countries included in the country rankings available on http://www.atlas.cid.harvard.edu.[13] To reduce noise, export values of less than USD 1000 at the bilateral-product level are then set to 0, and countries' exports of a given product are dropped if they are not shipped to at least 3 destinations in that year (after dropping export values of less than USD 1000). Robustness checks with regards to the year and the data cleaning thresholds are provided in Supplementary Material S5.

The fixed effects gravity regressions are run using the Stata *reghdfe* (OLS, Correia 2017) and *ppmlhdfe* command (PPML, Correia et al. 2019a,b), respectively. As discussed above, this allows estimating the T_i^s up to a scaling factor for each country and product. In the baseline specification, the estimated exporter-product fixed effects are normalized such that for every country i and every product s it holds

$$\sum_{\hat{s}\in\mathcal{S}_i} \delta_i^{\hat{s}} = 0 \tag{10a}$$

$$\sum_{\hat{i}\in\mathcal{I}^s} \delta_{\hat{i}}^{s} = 0, \tag{10b}$$

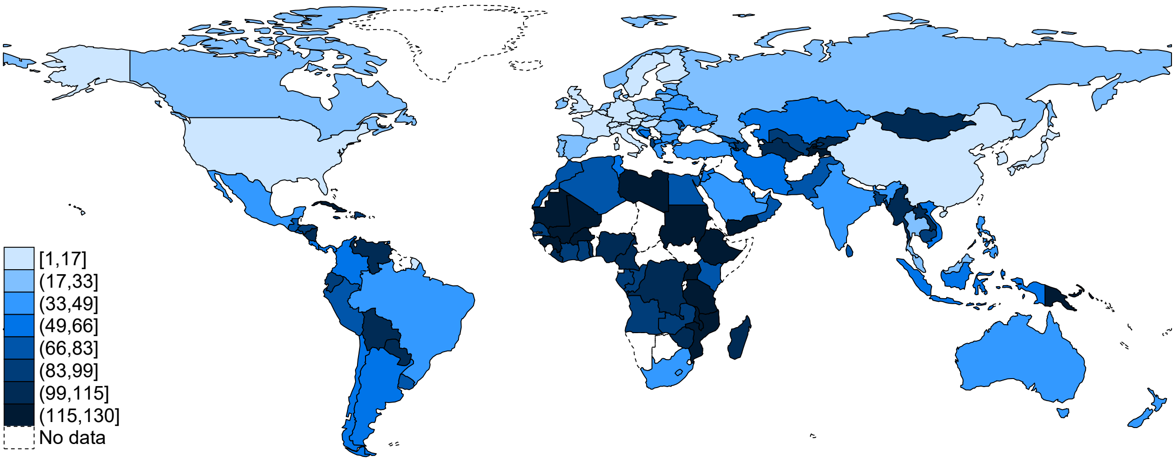
where $\mathcal{I}^s$ denotes the set of countries that are exporting product s and $\mathcal{S}_i$ denotes the set of products that country i exports. This normalization does not matter asymptotically. (10) balances countries and products and it avoids that normalized productivities scale with the random T_i^s in one reference country and product.[14] The normalized exporter-product fixed effects are then exponentiated and taken to the power .5, because the country-country similarity matrix is based on a quadratic form.[15]

[13] The list of countries may be seen from Table S1.1. To come up with this list, the Growth Lab at Harvard University starts from the list of countries included in the UN Comtrade database and then eliminates countries with at least one of the following: (i) population of less than 1m; (ii) average exports over the preceding three years of less than USD 1bn; (iii) unsatisfactory data quality due to e.g. failure of disclosure or war. I thank Sebastian Bustos for sharing the list of countries. The importers and exporters not included in the list of 130 countries have been aggregated to a rest-of-the-world when estimating the exporter-product fixed effects, but they have been dropped before computing the country ranking.

[14] Robustness checks with regards to the normalization are provided in Supplementary Material S5.

[15] Taking the square-root does again not impact the log-supermodularity of $\hat{T}_i^s$. It does imply that two countries that share two products and both have productivity T in both products have the same similarity as two countries that also share two products and both have productivity $T+\delta$

Figure 2: Distance to Frontier Across the World



Notes: The figure shows a color-coded map of the world based on the country ranking for 2019.

The estimated $\hat{T}_i^s$ are highly skewed to the right. To avoid that the country rankings are heavily influenced by outliers, $\hat{T}_i^s$ is winsorized at the top. In the baseline specification, all values above the 95$^{\text{th}}$ percentile are winsorized, treating missing $\hat{T}_i^s$ as zeros. The country ranking is robust to different choices for this threshold—see Supplementary Material S5.

Lastly, to determine the order of the ranking, Japan is chosen to be ranked atop, which implies that industrialized countries are ranked high.

5 Country Ranking

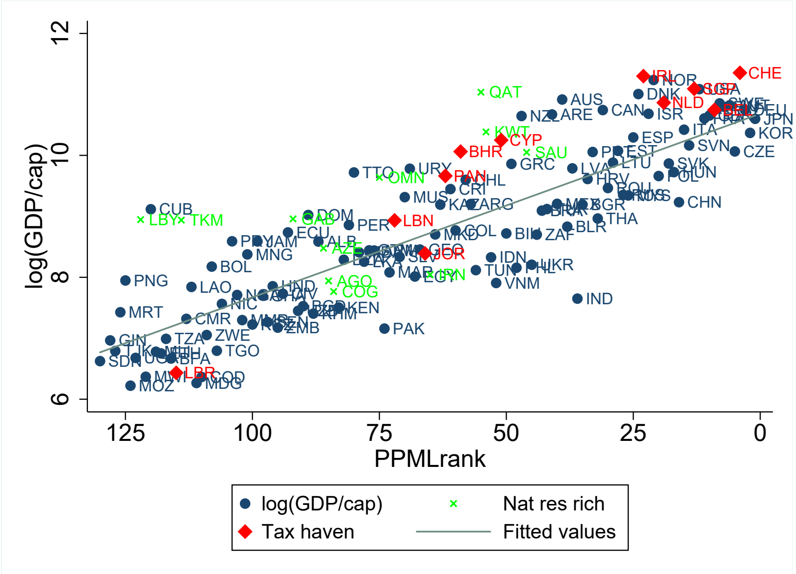
This section outlines the country ranking. It discusses its relation to intuitive notions of distance to frontier, and provides evidence in support of the main underlying assumption.

5.1 Overview

Figure 2 summarizes the baseline country ranking—the detailed ranking is provided in Supplementary Material S1. As may be seen from this figure, the ranking is in line with common perceptions of countries' distance to frontier: It ranks the United States, leading economies in Western Europe, as well as Japan, South Korea, and also China high, while many countries in Latin America, Central Asia, and Sub-Saharan Africa are lagging behind.

in one product and $T - \delta$ in the other product. A robustness check is provided in Supplementary Material S5.

Figure 3: Distance to Frontier and Income



Notes: The figure shows a scatter plot for 2019 with a country's rank of distance to frontier on the horizontal axis and its log GDP/cap in current USD on the vertical axis. Countries with a natural resource rent of over 15% of GDP in 2019 according to World Bank (2021a) are marked by green crosses. Countries that are tax havens according to Tørsløv et al. (2018) are marked by red diamonds. The line shows a linear fit for all countries that are neither tax havens nor natural resource rich.

To more systematically assess the quality of the ranking, Figure 3 locates each country in a scatter plot with its rank by its distance to frontier on the horizontal axis and its log GDP/cap on the vertical axis. The figure reveals a tight relationship with a rank correlation of .83. This suggests that there is a close connection between a country's absolute advantage—as reflected in its GDP/cap—and its comparative advantage across products—as considered here—, in line with a 'technology gap' model (Krugman, 1985). Intuitively, factors that allow countries to develop a comparative advantage in 'complex' products such as skills, technologies, or institutions should also allow them to increase their productivity across all products. In line with this reasoning, the country ranking correlates strongly with country characteristics that we might think of as helping countries to develop a comparative advantage in 'complex' products—see Table 2. Moreover, with the exception of the Human Capital Index, changes in the country ranking are associated with changes in these country characteristics, as shown in Supplementary Material S4.2.

While there is a tight relationship between the country ranking and income, this relationship is not perfect. It is therefore interesting to note that deviations from this relationship tend to be meaningful as may be seen from Figure 3: Emerging economies in East Asia such as Vietnam, Thailand, or China, for example, have export baskets that are closer to the frontier than what their levels of income suggest. The same

Table 2: Distance to Frontier and Sources of Comparative Advantage in Complex Products

	PPML	log(RD)	GE	log(HC)	log(K/L)	log(TFP)
PPML	1.00	0.80	0.86	0.77	0.83	0.54
log(RD)		1.00	0.67	0.68	0.60	0.30
GE			1.00	0.75	0.84	0.64
log(HC)				1.00	0.76	0.49
log(K/L)					1.00	0.64
log(TFP)						1.00

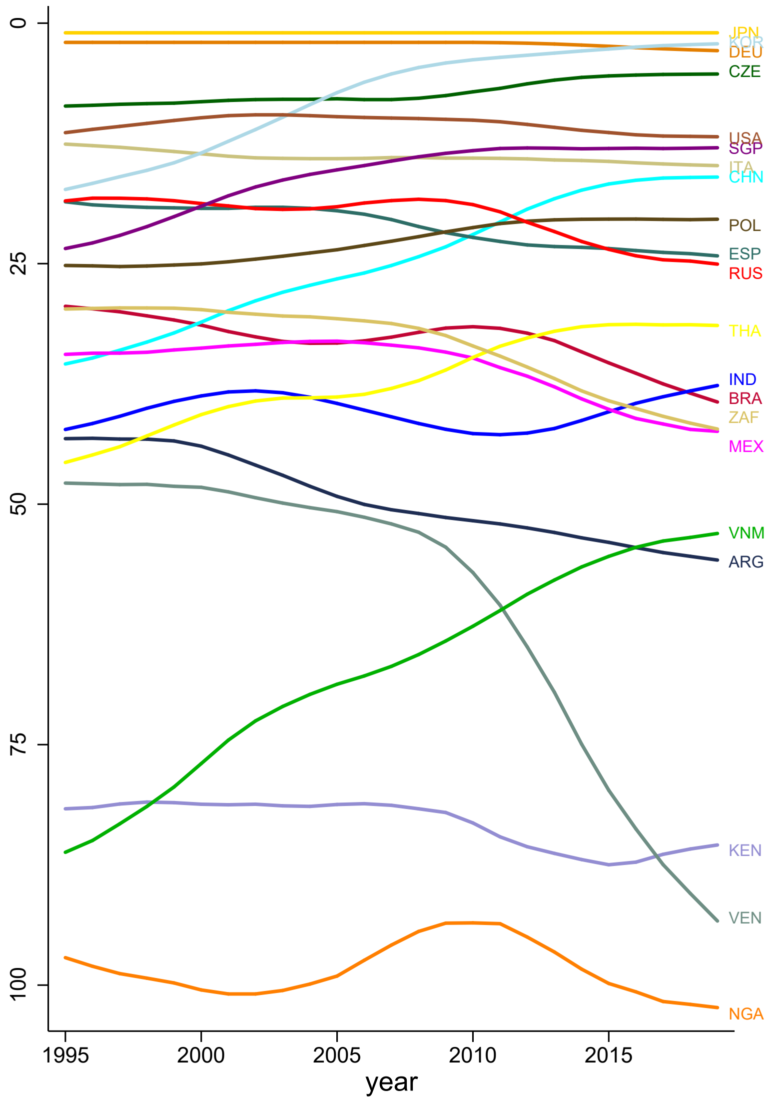
This table shows rank correlations between the ranking of distance to frontier and different country characteristics. log(RD) is the 19-year centered moving average of a country's log share of GDP invested in R&D from World Bank (2021a). GE is 'government effectiveness' from the Worldwide Governance Indicators (WGI) (World Bank, 2021b). log(HC) is the log Human Capital Index, log(K/L) is the log capital stock per capita in current PPPs, and log(TFP) is log total factor productivity relative to the US in current PPPs, all from the Penn World Table 10.0 (Feenstra et al., 2015). All data is for year 2019. Correlations with other measures of country characteristics are provided in Supplementary Material S4.2.

is true for many Eastern European countries such as Poland, Hungary, or the Czech Republic. On the contrary, tax havens—marked with a red diamond in Figure 3—as well as natural resource rich countries—marked with a green cross in Figure 3—tend to be richer given their distance to frontier. This suggests that the ranking indeed entails valuable information about countries' capabilities that is orthogonal to GDP. In fact, the ranking is related to the Economic Complexity Index (Hidalgo and Hausmann, 2009), which has been shown to be a strong predictor of future growth controlling for current income. I will get back to this point in Section 6.1.

The ranking also reflects well common priors about economic convergence in the past as may be seen from Figure 4. This figure traces rankings over time for a selection of countries. The rank-changes for all countries, as well as the same figures for the time interval 1965-2019 using SITC trade data are provided in Supplementary Material S1.2. The figure clearly shows that East Asian countries like South Korea, China, but also Vietnam, or Thailand succeeded in moving closer to the frontier. The same is true for the two-largest EU-10 economies, Poland and the Czech Republic, albeit starting from a higher level. At the same time, leading economies from Southern Europe, Latin America, or Sub-Saharian Africa stagnated or slightly fell behind, while the ranking of Venezuela collapsed over the last years.

In short, these discussions suggest that the country ranking indeed captures important aspects of countries' distance to frontier. Moreover, it is highly robust to the choice of the first-step estimator, the set of countries in the sample, the industry classification used, and to the various data cleaning steps involved—see Supplementary Material S5.

Figure 4: Country Rankings over Time



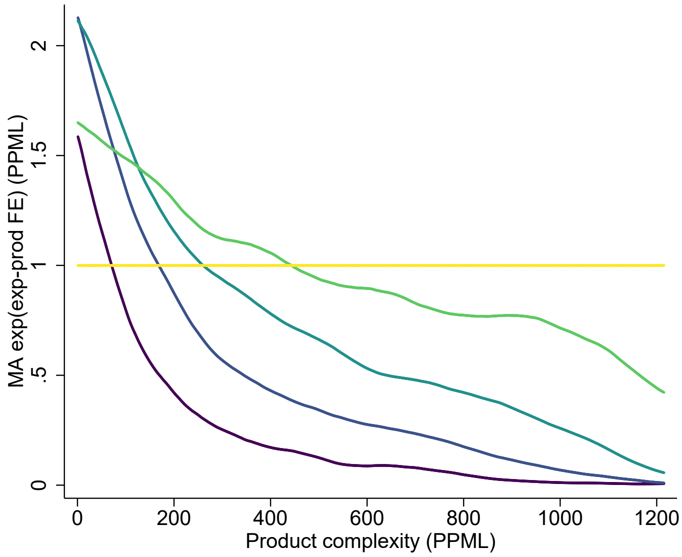
Notes: The figure shows for a selection of countries their 11-year weighted moving average ranking from 1995 to 2019. Weights: $\{1/36, 2/36, ..., 6/36, 5/36, ..., 1/36\}$.

Ultimately, however, the ranking hinges on the validity of the underlying assumptions. I turn to this issue next.

5.2 Test of Main Assumption

The main assumption underlying the country ranking is strongly supported by the data. Figure 5 shows that productivities are log-supermodular, with countries closer to the frontier having a relatively—in a diff-in-diff sense—higher productivity in more complex products. This figure orders products by their complexity from the least to the most complex. Further details on the measure for product complexity are provided in Section 6.2. For now, I note that it is consistent with the micro-foundation of

Figure 5: Country Specialization Across Products



Notes: The figure shows line plots summarizing the specialization across products for 5 bins of countries. Products are ordered by their complexity from least to most complex, where the complexity metric from Section 6.2 has been used. Lines are color-coded by country capability with brighter / more yellow indicating higher capability. Each line has been constructed as follows: (i) for each country, a weighted 201-product centered moving average of its estimated productivities has been computed using the following weights: $\{1/(101)^2, 2/(101)^2, ..., 101/(101)^2, 100/(101)^2, ..., 1/(101)^2\}$. (ii) These centered moving averages have been averaged over all countries in the respective bin. (iii) For each country-bin and product, the resulting average has been normalized by the corresponding average for the country bin with highest capability.

Section 2.3.2. Figure 5 further groups countries into 5 bins from highest (yellow line) to lowest capability (dark blue line)—a plot with separate lines for all countries is provided in Supplementary Material S4.3. For each country-bin-product, Figure 5 then shows the productivity averaged over all countries in a bin and over a moving range of products, with the resulting average normalized to 1 for the country bin with highest capability and all products—see the notes of Figure 5 for further details. This figure clearly reveals a strongly log-supermodular pattern.

While a complementarity between country capability and product complexity provides a micro-foundation for the country ranking, this ranking ultimately hinges on the more general Assumption 1. That is, it hinges on the assumption that countries close to the frontier have—in a diff-in-diff sense—an export basket that is similar to that of other countries close to the frontier. Table 3 provides a formal test of this assumption following Davis and Dingel (2020). Specifically, this table starts from the estimated country-country similarity matrix $\hat{\boldsymbol{A}}$, re-ordered according to the country ranking.

[16] I also computed p-values for a null model where M_{is} has been randomly sampled from a log-normal distribution, but then the same steps (ii) and (iii) have been pursued. This results in a distribution for the null model that is even closer to .5. Details are available upon request.

Table 3: Test of Log-supermodularity of Country-Country Similarity Matrix

Bins	# comparisons	Success rate (p-value)
130	35149920	0.741 (0.000)
65	2162160	0.809 (0.000)
44	446985	0.852 (0.000)
26	52650	0.894 (0.000)
13	3003	0.956 (0.000)
6	105	1.000 (0.000)
4	15	1.000 (0.000)

This table shows results from a test of log-supermodularity following Davis and Dingel (2020) of the estimated matrix $\hat{\boldsymbol{A}}$ for the year 2019. Column 'bins' refers to the number of bins in which the 130 countries have bin grouped and aggregated. Column '# comparisons' specifies the number of inequality tests performed in the respective row. Column 'Success rate' specifies the share of these inequality-tests that satisfied the condition of log-supermodularity. p-values are based on 2000 iterations of a null-model, where in each iteration (i) matrix $\boldsymbol{M}$ with element $M_{is} = \hat{T}_i^s z_i^s$ has been randomly reshuffled s.t. row and column sums were preserved;[16](ii) the random matrix has been cleaned and rearranged based on the same eigen-procedure used for the data matrix; (iii) the same test of LSM has been applied as for the data matrix.

For each quadruple of elements in the intersections of pairs of rows and columns, it then asks whether these elements are log-supermodular. Taking into account the symmetry of $\hat{\boldsymbol{A}}$ and ignoring quadruples with the same pair of countries in the rows and columns, this yields $\frac{[I(I-1)/2]^2}{2} - I(I-1)/4$ inequality tests, where $I = 130$ is the number of countries in the sample.[17] Table 3 shows the share of these inequality tests that were passed. It compares this share to a random null-model as detailed in the notes of Table 3, and repeats the same test, aggregating countries into increasingly large bins. As may be seen from Table 3, these tests strongly indicate that the estimated country-country similarity matrix is indeed log-supermodular.

6 Discussion and Additional Results

This section provides further discussions, first on the connection between the ranking here and the Economic Complexity Index (ECI, Hidalgo and Hausmann 2009), and then on ranking products by their 'complexity'.

[17]Quadruples of elements in the intersections of the same pair of rows and columns are log-supermodular by construction.

6.1 A Micro-foundation for the Economic Complexity Index

Proposition 1 and Theorem 1 provide a flexible micro-foundation for the Economic Complexity Index (ECI). The ECI was originally introduced as an iterative algorithm that starts from a binarized matrix of Revealed Comparative Advantages (RCA, Balassa 1965) and seeks to measure the economic complexity of countries that is revealed through the products they make. The ECI is widely used in the literature and in policy, and it has spurred an entire strand of literature that explores different ways of measuring the complexity of countries and products, and of applying a similar logic to related fields—see literature review. Yet, while the ECI is based on a very intuitive narrative, it has so far lacked a rigorous theoretical underpinning that starts from a structural notion of complexity and shows that the ECI can uncover this from the data. The results presented here provide such a micro-foundation.

In particular, the iterative algorithm of the ECI converges to the same eigenvector as the one studied in Section 3, but based on a particular choice for the underlying country-country similarity matrix derived from binarized RCAs (Hausmann et al., 2011; Caldarelli et al., 2012; Mealy et al., 2019).[18] Theorem 1 therefore immediately implies that the ECI correctly ranks countries if the underlying similarity matrix is log-supermodular. This not only is a parsimonious, non-parametric condition, but it is also in line with the guiding rationale of the ECI. It is consistent with a world where complex countries specialize in complex products, but also a 'pure ladder' (Atkin et al., 2021) or 'nested' economy, where countries further from the frontier make subsets of the products that countries closer to the frontier are making (Hausmann and Hidalgo, 2011; Bustos et al., 2012; Schetter, 2020)—see Supplementary Material S3 for further details. Moreover, a log-supermodular country-country matrix is consistent with Hausmann and Hidalgo (2011), who provide a probabilistic model where production requires

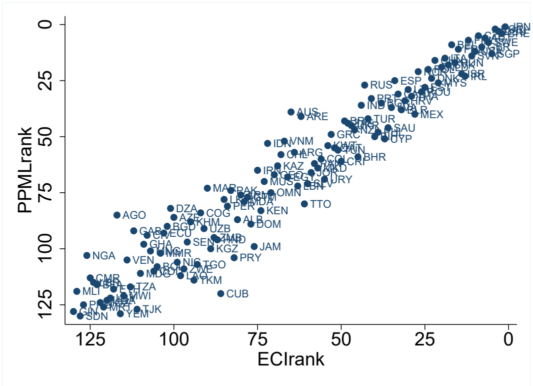
[18] Specifically, the ECI is the eigenvector corresponding to the second smallest eigenvalue of (Hausmann et al., 2011; Caldarelli et al., 2012; Mealy et al., 2019)

$$[\boldsymbol{D_{ECI}} - \boldsymbol{A_{ECI}}]\,\boldsymbol{y} = \lambda \boldsymbol{D_{ECI}}\boldsymbol{y}.$$

$\boldsymbol{A_{ECI}}$ is a country-country matrix that specifies for each pair of countries the number of products that they both export with $RCA \geq 1$, with each product weighted by the inverse of its 'ubiquity'. As before, $\boldsymbol{D_{ECI}}$ is the diagonal matrix with diagonal entries equal to the respective row sum of $\boldsymbol{A_{ECI}}$. Formally, let $\boldsymbol{M_{ECI}}$ denote the $I \times S$ binary country-product matrix with entry $M_{ECI,is} = 1$ if country i has an RCA of at least 1 in product s and $M_{ECI,is} = 0$ otherwise. Further, let $\boldsymbol{U_{ECI}}$ be the $S \times S$ diagonal matrix with entry $U_{ECI,ss}$ equal to the ubiquity of product s, i.e., $U_{ECI,ss}$ is the sum of the sth column of matrix $\boldsymbol{M_{ECI}}$. Then, the country-country similarity matrix used for the ECI is defined as

$$\boldsymbol{A_{ECI}} := \boldsymbol{M_{ECI}}\boldsymbol{U_{ECI}}^{\boldsymbol{-1}}\boldsymbol{M_{ECI}}^{\boldsymbol{T}}.$$

Figure 6: Country Ranking vs ECI



Notes: This figure shows a scatter plot with the country ranking on the vertical axis and a country's ranking based on the Economic Complexity Index on the horizontal axis. The data refers to year 2019.

product-specific subsets from a large and heterogeneous set of 'capabilities'.[19]

While Theorem 1 and Proposition 1 provide a flexible micro-foundation for the ECI, starting from a binary matrix based on RCAs might not be innocuous. For one thing, in a world with trade frictions, using RCAs does not in general allow to uncover productivities—or comparative advantages more broadly—from the data. For another, the log-supermodularity of $\boldsymbol{A}$ is not in general preserved when binarizing the country-product matrix. The country ranking considered here overcomes these potential pitfalls by building on a workhorse gravity framework. Surprisingly though, despite the different starting points, the ranking is similar to the ECI, as may be seen from Figure 6 (the rank correlation is .96). This further confirms that these measures are highly robust. And it lends additional support to the large literature that uses the ECI—and the Product Complexity Index (PCI)—in empirical studies—see literature review. I turn to measuring product complexity next.

6.2 Ranking Products by their Complexity

The main focus of the paper is on ranking countries by their distance to frontier. Yet, the same logic can also be used to rank products by their complexity. This may be seen most clearly when reverting to the micro-foundation of Section 2.3.2. Proposition 1 shows how a log-supermodular country-country matrix can be derived from a log-supermodular country-product matrix. By the same token, one can derive

[19] In particular, in their model, the probability that a country can make a product is log-supermodular in natural notions of country and product complexity: the probability that a country has any given capability and the number of capabilities that a product requires—see their Equation (20).

a log-supermodular product-product similarity matrix

$$B_{s's} = \sum_{i \in \mathcal{I}} \tilde{T}_i^{s'} \tilde{\rho}_i^{s'} \tilde{T}_i^{s} \tilde{\rho}_i^{s}. \tag{11}$$

Hence, starting from the estimated exporter-product fixed effects, it is possible to derive a ranking of products by their complexity by first forming the sample analogue of $\boldsymbol{B}$ and by then computing the second eigenvector of

$$\boldsymbol{L_B y} = \lambda \boldsymbol{D_B y},$$

where $\boldsymbol{D_B}$ denotes the diagonal matrix with entries $D_{B,ss}$ equal to the respective row sum of $\boldsymbol{B}$, and $\boldsymbol{L_B} := \boldsymbol{D_B} - \boldsymbol{B}$ the Laplacian matrix of $\boldsymbol{B}$.

Table 4 shows the top and bottom 15 products according to the ranking derived from using PPML in the first step regression. This ranking of product complexity is more prone to noise. It ranks high some small product categories that are exported by few countries with high capability.[20] This may be due to an underlying 'complexity' or due to 'idiosyncratic' forces unrelated to complexity. My framework is flexible enough to allow for such idiosyncratic forces. Yet, while such forces plausibly balance out when ranking 130 countries based on more than 1200 products, the opposite is not necessarily the case.[21]

Nevertheless, the product ranking is robust,[22] and it may serve as a flexible structural alternative to proxies for product complexity that have previously been used (e.g. Levchenko 2007; Costinot 2009b; Schetter 2020). The bottom 15 products, for example, are in line with an intuitive understanding of low product complexity, and this is also the case more broadly. More formally, we know from Figure 5, that—on balance—it is the case that countries closer to the frontier have a comparative advantage in complex products. And the product-product similarity matrix $\hat{\boldsymbol{B}}$ is also log-supermodular as shown in Supplementary Material S4.4.

The ranking also correlates highly with product characteristics that may be associated with complexity, as shown in Table 5—see the table notes and Supplementary

[20] Out of the top-15 products, for example, 4 products (2705, 8106, 5507, and 1183) are among the 50 products with lowest total exports.

[21] As an alternative, I could aggregate products into fewer categories. An advantage of the approach here is that this could, in principle, easily be done, i.e., it allows computing a product complexity ranking that is tailored to specific needs. Aggregating product categories would, however, imply lumping together products that are potentially fairly different. For the purpose of this paper, I therefore consider a ranking of 4-digit HS product categories. This has the additional advantage of forming the counterpart to the country ranking.

[22] The rank correlation between the baseline product ranking and (i) the one using OLS is .971 (ii) the one using RCAs is .975 and (iii) the PCI is .832. Further robustness checks are provided in Supplementary Material S5.2.2.

Table 4: Product Complexity Ranking

HS4d	Description	PPMLrank	OLSrank	RCArank	PCIrank
8401	Nuclear reactors and related equipment	1	1	7	98
2705	Non-petroleum gases	2	32	18	5
7111	Platinum clad metals	3	10	2	27
2850	Hydrides, nitrides, azides, silicides and borides	4	8	5	49
3705	Photographic film, developed	5	16	1	1
7506	Nickel plates	6	11	4	11
8106	Bismuth	7	13	19	291
3601	Propellant powders	8	9	14	118
5507	Artificial staple fibers, processed	9	21	46	163
2812	Halides of nonmetals	10	29	10	29
9110	Clock movements, complete, unassembled	11	45	3	10
7221	Bars of stainless steel, hot-rolled	12	35	17	58
8109	Zirconium	13	7	21	209
8601	Electric trains	14	4	33	119
7109	Gold clad metals	15	6	15	22
⋮	⋮	⋮	⋮	⋮	⋮
6105	Men's shirts, knit	1201	1147	1207	1172
801	Cashew nuts & coconuts	1202	1205	1210	1211
6305	Bags for packing goods	1203	1189	1204	1178
1212	Seaweeds & edible vegetable products	1204	1202	1184	1182
804	Avocados, pineapples, mangos, etc.	1205	1206	1214	1176
5201	Raw cotton	1206	1213	1213	1210
1802	Cocoa residues	1207	1212	1044	1215
1202	Peanuts	1208	1210	1208	1204
803	Bananas and plantains	1209	1214	1209	1201
708	Legumes	1210	1211	1202	1180
1211	Plants used in perfumery, pharmacy or insecticide	1211	1203	1200	1190
714	Tubers	1212	1208	1211	1200
1207	Other oil seeds	1213	1209	1203	1207
2401	Unmanufactured tobacco	1214	1207	1212	1197
1801	Cocoa beans	1215	1215	1215	1214

This table shows rankings of product complexity for the year 2019 using trade data at the HS4d classification level. Rankings are shown for the top 15 and bottom 15 products according to the baseline ranking using PPML in the first step. PPML refers to this ranking. OLS to the ranking derived from using OLS, and analogously for RCA. PCI refers to the product complexity index.

Material S4.1 for further details. This is particularly true for the required on-the-job learning (log(RL)), and for the skill intensity of an industry (log(SI)), but also for the intermediate input diversification (log(ID)) and the investment to output ratio (log(I/Y)). On the contrary, the ranking does not—or even negatively— correlate with the capital intensity of an industry (log(CI)) and its reliance on external financing (EF). Through the lense of the micro-foundation in Section 2.3.2 this suggests that countries closer to the frontier tend to specialize in skill intensive products.[23]

[23] This is in line with Malmberg (2020) who shows that the exports of richer countries are more skill-intensive, and Buera et al. (2022) who document that increases in income are associated with a shift towards skill-intensive industries.

Table 5: Product Complexity and Product Characteristics

	PPML	log(RL)	log(SI)	log(ID)	log(I/Y)	log(CI)	EF
PPML	1.00	0.82	0.54	0.40	0.27	-0.12	-0.03
log(RL)		1.00	0.16	0.11	0.07	0.06	0.23
log(SI)			1.00	0.35	0.03	-0.55	0.10
log(ID)				1.00	0.18	-0.14	0.02
log(I/Y)					1.00	0.19	-0.03
log(CI)						1.00	-0.06
EF							1.00

This table shows rank correlations between the ranking of product complexity for 1995 and different product characteristics. I consider the product ranking for 1995—the first year in my sample—because the product characteristics are for 1992 or before. RL is a measure of required on-the-job learning for a worker to become fully productive for 40 different SIC72 industries taken from Costinot (2009b). All other product characteristics are from Levchenko (2007) and they are measures of an industry's skill intensity (SI), intermediate input diversification (ID), investment-output ratio (I/Y), capital intensity (CI), and external finance dependence (EF), respectively. The latter is based on Rajan and Zingales (1998) and is in levels because it has negative values. Further details on these measures and on concordances with the product complexity ranking are provided in Appendix S4.1.

7 Conclusion

This paper presents a structural ranking of countries by their distance to frontier based on a standard fixed effects gravity regression. At the heart of this ranking is a generalized version of a 'technology gap' model: The ranking is based on the assumption that countries closer to the frontier are relatively—in a 'diff-in-diff' sense—better in some products than in others. I provide supportive evidence for this idea and show how it can be taken to the data without functional form assumptions nor the need of taking a stance on the nature of the products that countries at the frontier are good at producing.

The country ranking is rooted in comparative as opposed to absolute advantage, and it therefore provides information that is fundamentally different from GDP per capita. The ranking can thus provide a novel perspective on the development process of countries, disentangling changes in incomes from progress in the deep underlying productive capabilities of an economy. Future work may set out to analyze these dynamics more carefully.

As part of the analysis, I proposed a general, non-parametric approach for ranking items according to a log-supermodular structure. The underlying framework provides a flexible micro-foundation for the Economic Complexity Index (Hidalgo and Hausmann, 2009). More generally, this approach allows to learn about structural drivers of comparative advantage in trade and beyond.

Appendix

A Proofs

A.1 Proof of Proposition 1

It needs to be shown that for every quadruple of countries (i, i', k, k') such that $i < i'$ and $k < k'$ it holds

$$A_{ik} \cdot A_{i'k'} > A_{ik'} \cdot A_{i'k}.$$

This follows first from noting that $X_i^s := \tilde{T}_i^s \rho_i^s$ is log-supermodular as well by Assumption $1'$ and the fact that log-supermodularity is preserved under multiplication (e.g. Costinot 2009a, Lemma 1). And second, from the following chain of (in)equalities:

$$\begin{aligned}
A_{ik'} \cdot A_{i'k} &= \left[\frac{1}{S}\sum_{s\in\mathcal{S}} X_i^s X_{k'}^s\right] \cdot \left[\frac{1}{S}\sum_{s\in\mathcal{S}} X_{i'}^s X_k^s\right] \\
&= \left[\frac{1}{S}\sum_{s\in\mathcal{S}} \frac{X_i^s}{X_{i'}^s}\frac{X_{k'}^s}{X_k^s} X_{i'}^s X_k^s\right] \cdot \left[\frac{1}{S}\sum_{s\in\mathcal{S}} X_{i'}^s X_k^s\right] \\
&< \left[\frac{1}{S}\sum_{s\in\mathcal{S}} \frac{X_i^s}{X_{i'}^s} X_{i'}^s X_k^s\right] \cdot \left[\frac{1}{S}\sum_{s\in\mathcal{S}} \frac{X_{k'}^s}{X_k^s} X_{i'}^s X_k^s\right] \qquad \text{(A.1)} \\
&= \left[\frac{1}{S}\sum_{s\in\mathcal{S}} X_i^s X_k^s\right] \cdot \left[\frac{1}{S}\sum_{s\in\mathcal{S}} X_{i'}^s X_{k'}^s\right] \\
&= A_{ik} \cdot A_{i'k'}.
\end{aligned}$$

The inequality follows from noting first that $X_{i'}^s X_k^s > 0$, second that $\frac{X_i^s}{X_{i'}^s}$ is decreasing in s while $\frac{X_{k'}^s}{X_k^s}$ is increasing in s by Assumption $1'$, and from applying Chebyshev's Sum Inequality (Hardy et al., 1934, Theorem 43).[24] Inequality (A.1) shows the desired result.

□

A.2 Proof of Theorem 1

I show the desired result by means of two lemmata. In particular, note that the eigenvector corresponding to the second smallest eigenvalue of (5) solves the following constrained minimization problem (e.g. Chung 1997; Shi and Malik 2000; Belkin and

[24] Note that $\frac{X_i^s}{X_{i'}^s}$ and $\frac{X_{k'}^s}{X_k^s}$ are both non-constant, i.e., the inequality is indeed strict.

Niyogi 2003)

$$
\begin{aligned}
\arg\min \ & \boldsymbol{y^T L y} \\
s.t. \ & \boldsymbol{y^T D y} = 1 \\
& \boldsymbol{y^T D 1} = 0,
\end{aligned}
\tag{A.2}
$$

where **1** denotes a vector of ones and where the objective in this minimization problem can be rewritten as

$$\boldsymbol{y^T L y} = \frac{1}{2}\sum_{ij}(y_i - y_j)^2 A_{ij}. \tag{A.3}$$

My strategy is therefore to show first that if I augment minimization problem (A.2) by an additional constraint $\boldsymbol{y} \in \mathcal{A}$, where $\mathcal{A}$ is a closed set, the optimal solution to the augmented minimization problem is either the second eigenvector of (5), or it must be on the boundary of $\mathcal{A}$ (Lemma 1). Second, I find a closed set $\mathcal{A}$ such that the optimal solution to the augmented minimization problem is both strictly monotonic and in the interior of $\mathcal{A}$ (Lemma 2). The desired result then follows.

Throughout, I use λ_k to denote the k^{th} smallest eigenvalue of the generalized eigenproblem (5), and $\boldsymbol{u}_k$ to denote the corresponding eigenvector which I will henceforth refer to as the k^{th} eigenvector of (5).

Lemma 1

For every closed set $\mathcal{A}$ such that set $\mathcal{Y} := \left\{\boldsymbol{y} \in \mathbb{R}^I : \boldsymbol{y^T D y} = 1, \boldsymbol{y^T D 1} = 0, \boldsymbol{y} \in \mathcal{A}\right\}$ is nonempty, vector $\boldsymbol{y}^*$ defined as

$$
\begin{aligned}
\boldsymbol{y}^* := \arg\min \ & \boldsymbol{y^T L y} \\
s.t. \ & \boldsymbol{y} \in \mathcal{Y},
\end{aligned}
$$

is either $\boldsymbol{u}_2$, or it is on the boundary of set $\mathcal{A}$.

Proof:

Substituting $\boldsymbol{z} := \boldsymbol{D}^{1/2}\boldsymbol{y}$, we get

$$\boldsymbol{y}^* = \boldsymbol{D}^{-1/2}\boldsymbol{z}^*,$$

where

$$\boldsymbol{z}^* := \underset{\boldsymbol{z^T z}=1, \boldsymbol{z^T D}^{1/2}\boldsymbol{1}=0, \boldsymbol{D}^{-1/2}\boldsymbol{z}\in\mathcal{A}}{\arg\min} \boldsymbol{z^T \tilde{L} z}, \tag{A.4}$$

with $\boldsymbol{\tilde{L}} := \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$. Let us then consider this problem instead.

$\tilde{\boldsymbol{L}}$ is symmetric. Let $\boldsymbol{v}_k$ be the eigenvector corresponding to λ_k, the k^{th} smallest eigenvalue of $\tilde{\boldsymbol{L}}$. Note that λ_k is the exact same eigenvalue as previously defined and that $\boldsymbol{v}_k = \boldsymbol{D}^{1/2}\boldsymbol{u}_k$. It follows that $\boldsymbol{v}_1 = \boldsymbol{D}^{1/2}\mathbf{1}$.[25] Hence, constraint $\boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{D}^{1/2}\mathbf{1} = 0$ requires $\boldsymbol{z}$ to be orthogonal to the first eigenvector of $\tilde{\boldsymbol{L}}$.

Now, suppose that $\boldsymbol{u}_2 \in \mathcal{A}$. The Courant-Fischer Minimax Theorem then immediately implies that $\boldsymbol{z}^* = \boldsymbol{v}_2$ and, hence, $\boldsymbol{y}^* = \boldsymbol{u}_2$ (Shi and Malik 2000; Golub and van Loan 2013, Theorem 8.1.2), which proves our desired result for the case of $\boldsymbol{u}_2 \in \mathcal{A}$.

To show the desired result for the case of $\boldsymbol{u}_2 \notin \mathcal{A}$, we proceed by contradiction. Using the eigendecomposition of $\tilde{\boldsymbol{L}}$, we get

$$\boldsymbol{z}^{\boldsymbol{T}}\tilde{\boldsymbol{L}}\boldsymbol{z} = \boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{z} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{z})^{\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{z},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with element $\Lambda_{kk} = \lambda_k$ and $\boldsymbol{V}$ is the matrix whose k^{th} column is the eigenvector of $\tilde{\boldsymbol{L}}$ corresponding to λ_k, normalized to have length 1. Substituting $\boldsymbol{r} := \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{z}$ therefore yields that

$$\boldsymbol{y}^* = \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r}^*,$$

where[26]

$$\boldsymbol{r}^* := \underset{\boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{r}=1, \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{e}_1=0, \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r}\in\mathcal{A}}{\arg\min} \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{r}. \tag{A.5}$$

Now, suppose by way of contradiction that $\boldsymbol{y}^* = \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r}^* \in \mathcal{A}^{\circ}$, where $\mathcal{A}^{\circ}$ denotes the interior of set $\mathcal{A}$. On the one hand, we have $\boldsymbol{r}^* \in \operatorname{span}(\{\boldsymbol{e}_2, \boldsymbol{e}_3, ..., \boldsymbol{e}_n\})$. On the other hand, $\boldsymbol{u}_2 \notin \mathcal{A}$ and, hence, $\boldsymbol{r}^* \nparallel \boldsymbol{e}_2$, by assumption. Hence, there exists an $\tilde{\boldsymbol{r}}$ with

$$\tilde{r}_j := \begin{cases} r_j^* + dr_2 & \text{if } j = 2 \\ r_j^* + dr_k & \text{if } j = k \\ r_j^* & \text{otherwise} \end{cases}$$

[25]Note that $\boldsymbol{u}_1 = \mathbf{1}$ is the eigenvector corresponding to the smallest eigenvalue $\lambda_1 = 0$ of the generalized eigenproblem (5).

[26]To see this, note that using $\boldsymbol{r} := \boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{z}$ allows re-writing the objective in (A.4) as

$$\boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{r}.$$

Moreover, the orthogonality of $\boldsymbol{V}$ implies that $\boldsymbol{z} = \boldsymbol{V}\boldsymbol{r}$ and, hence,

$$\boldsymbol{y} = \boldsymbol{D}^{-1/2}\boldsymbol{z} = \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r}$$
$$\boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{z} = (\boldsymbol{V}\boldsymbol{r})^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{r} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{r} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{r},$$

where the last equality in the second line follows again from the fact that $\boldsymbol{V}$ is orthogonal. Lastly, using

$$\boldsymbol{z}^{\boldsymbol{T}}\boldsymbol{D}^{1/2}\mathbf{1} = (\boldsymbol{V}\boldsymbol{r})^{\boldsymbol{T}}\boldsymbol{D}^{1/2}\mathbf{1} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{D}^{1/2}\mathbf{1} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{e}_1,$$

where $\boldsymbol{e}_i$ denotes the i^{th} unit vector, implies that $\boldsymbol{y}^* = \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r}^*$ with $\boldsymbol{r}^*$ as defined in (A.5). In the above, the last equality follows from the fact that $\boldsymbol{v}_1 := \boldsymbol{D}^{1/2}\mathbf{1}$ is the first eigenvector of $\tilde{\boldsymbol{L}}$.

for some $k > 2$, such that

$$
\begin{aligned}
&(\tilde{r}_k)^2 < (r_k^*)^2, \\
&(\tilde{r}_2)^2 + (\tilde{r}_k)^2 = (r_2^*)^2 + (r_k^*)^2,
\end{aligned}
$$

and

$$\tilde{\boldsymbol{y}} := \boldsymbol{D}^{-1/2}\boldsymbol{V}\tilde{\boldsymbol{r}} \in \mathcal{A}.$$

Clearly, $\tilde{\boldsymbol{r}} \in \left\{\boldsymbol{r} \in \mathbb{R}^I : \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{r} = 1, \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{e}_1 = 0, \boldsymbol{D}^{-1/2}\boldsymbol{V}\boldsymbol{r} \in \mathcal{A}\right\}$. Moreover, $\lambda_i < \lambda_j$ for all $i < j$, implies that

$$\tilde{\boldsymbol{r}}^{\boldsymbol{T}}\boldsymbol{\Lambda}\tilde{\boldsymbol{r}} = \sum_{k=1}^{I} \lambda_k \tilde{r}_k^2 < \sum_{k=1}^{I} \lambda_k {r_k^*}^2 = \boldsymbol{r}^{*\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{r}^*,$$

a contradiction to $\boldsymbol{r}^*$—and, hence, $\boldsymbol{y}^*$—being optimal.[27] This concludes the proof of the lemma.

□

Lemma 2

The optimal solution to minimization problem

$$
\begin{aligned}
&\arg\min \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{L}\boldsymbol{y} \\
&\quad s.t.\ \boldsymbol{y} \in \mathcal{Y},
\end{aligned}
$$

where $\mathcal{Y} := \left\{\boldsymbol{y} \in \mathbb{R}^I : \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{D}\boldsymbol{y} = 1, \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{D}\mathbf{1} = 0, y_i \leq y_j\ \forall i \leq j\right\}$, *is such that* $y_i < y_j$ *for all* $i < j$.

Proof:

Note first that set $\mathcal{Y}$ is compact.[28] Hence, the continuous function $\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{L}\boldsymbol{y}$ attains a minimum on the set $\mathcal{Y}$ by the Extreme Value Theorem. It remains to be shown that this minimum is such that $y_i < y_j$ for all $i < j$. To do so, I proceed by contradiction.

Let $\boldsymbol{y}^*$ be the solution to the above minimization problem and suppose by way of contradiction that $\boldsymbol{y}^* \notin \mathcal{A}^\circ$, where $\mathcal{A} := \left\{\boldsymbol{y} \in \mathbb{R}^I : y_i \leq y_j\ \forall i \leq j\right\}$. Then, there exists a set of $m \geq 2$ consecutive numbers $\{i, ..., i+m-1\} \subset \{1, ..., I\}$ such that $y_k^* = y_l^*$

[27] Strictly speaking, this assumes that λ_2 is unique. With multiplicity larger than one of this eigenvalue, our arguments imply that $\boldsymbol{y}^*$ must either be a linear combination of the eigenvectors corresponding to the second smallest eigenvalue, or it must be on the boundary of set $\mathcal{A}$. Lemma 2 then implies that all of the eigenvectors corresponding to the second smallest eigenvalue must be monotonic—see Footnote 29.

[28] The definition of $\mathcal{Y}$ immediately implies that it is closed. The constraint $\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{D}\boldsymbol{y} = 1$ implies that ${y_i}^2 \leq \frac{1}{D_{ii}}$ for all $i \in \mathcal{I}$ which proves that $\mathcal{Y}$ is bounded.

$\forall\, k, l \in \{i, ..., i+m-1\}$. Moreover, $y^*_{i-1} < y^*_i$ (if $i > 1$) and, similarly, $y^*_{i+m-1} < y^*_{i+m}$ (if $i+m-1 < I$), where it cannot be that both $i = 1$ and $i+m-1 = I$, for if not, $\boldsymbol{y^*} \notin \mathcal{Y}$. Let $j := i+m-1$ and consider an alternative vector $\tilde{\boldsymbol{y}}$ satisfying

$$\tilde{y}_k = \begin{cases} y^*_k & \text{if } k \neq i, j \\ y^*_k + dy_k & \text{if } k = i, j \end{cases}$$

with dy_i, dy_j small and where

$$D_{ii} dy_i = -D_{jj} dy_j. \tag{A.6}$$

Clearly, $\tilde{\boldsymbol{y}}^{\boldsymbol{T}} \boldsymbol{D1} = 0$. Moreover, totally differentiating $f(\boldsymbol{y}) := \boldsymbol{y^T L y}$ and using $dy_k = 0$ for $k \neq i, j$, we get

$$\begin{aligned} df(\boldsymbol{y}) &= \sum_{k\in\mathcal{I}} (y_i - y_k) A_{ik} dy_i - \sum_{k\in\mathcal{I}} (y_k - y_i) A_{ki} dy_i + \sum_{k\in\mathcal{I}} (y_j - y_k) A_{jk} dy_j - \sum_{k\in\mathcal{I}} (y_k - y_j) A_{kj} dy_j \\ &= 2\left[\sum_{k\in\mathcal{I}} (y_i - y_k) A_{ik} dy_i\right] + 2\left[\sum_{k\in\mathcal{I}} (y_j - y_k) A_{jk} dy_j\right], \end{aligned}$$

where the second equality follows from the symmetry of $\boldsymbol{A}$. Using (A.6) and the fact that $y^*_i = y^*_j$ this implies

$$\begin{aligned} df(\boldsymbol{y^*}) &= 2\left[\sum_{k\in\mathcal{I}} (y^*_j - y^*_k)\left(A_{jk} - A_{ik}\frac{D_{jj}}{D_{ii}}\right)\right] dy_j \\ &= 2\left[\sum_{k\in\mathcal{I}} (y^*_j - y^*_k)\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right) A_{jk}\right] dy_j. \end{aligned} \tag{A.7}$$

Now, $(y^*_j - y^*_k)$ is decreasing in k by the definition of $\mathcal{Y}$, and $\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right)$ increasing by the log-supermodularity of $\boldsymbol{A}$ and the fact that $j > i$. Moreover, $A_{jk} > 0$ for all j, k. Chebyshev's Sum Inequality (Hardy et al., 1934, Theorem 43) therefore implies that

$$\begin{aligned} \left[\sum_{k\in\mathcal{I}} (y^*_j - y^*_k)\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right) A_{jk}\right] &< \frac{\left[\sum_{k\in\mathcal{I}} (y^*_j - y^*_k) A_{jk}\right] \cdot \left[\sum_{k\in\mathcal{I}}\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right) A_{jk}\right]}{\sum_{k\in\mathcal{I}} A_{jk}} \\ &= 0, \end{aligned} \tag{A.8}$$

where the equality follows from the fact that

$$\left[\sum_{k\in\mathcal{I}}\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right) A_{jk}\right] = D_{jj}\sum_{k\in\mathcal{I}}\left[\frac{A_{jk}}{D_{jj}} - \frac{A_{ik}}{D_{ii}}\right] = 0.$$

The inequality in (A.8) is strict because both $(y^*_j - y^*_k)$ and $\left(1 - \frac{A_{ik}}{A_{jk}}\frac{D_{jj}}{D_{ii}}\right)$ are non-constant. Equation (A.7) and Inequality (A.8) imply that for $dy_j > 0$ but small, moving from $\boldsymbol{y^*}$ to $\boldsymbol{\tilde{y}}$ strictly decreases the objective function. $\boldsymbol{\tilde{y}}$ is, however, not feasible as it violates constraint $\boldsymbol{y^T D y} = 1$. In particular,

$$\begin{aligned}
\boldsymbol{\tilde{y}^T D \tilde{y}} &= \sum_{k \in \mathcal{I}} \tilde{y}_k^2 D_{kk} \\
&= \boldsymbol{y^{*T} D y^*} + 2y_i^* dy_i D_{ii} + {dy_i}^2 D_{ii} + 2y_j^* dy_j D_{jj} + {dy_j}^2 D_{jj} \\
&= \boldsymbol{y^{*T} D y^*} + {dy_i}^2 D_{ii} + {dy_j}^2 D_{jj} \\
&> 1,
\end{aligned}$$

where the last equality follows from using Equation (A.6) in combination with $y_i^* = y_j^*$, and the inequality follows from $\boldsymbol{y^{*T} D y^*} = 1$. It follows, however, that we can scale $\boldsymbol{\tilde{y}}$ by some factor $\beta \in (0, 1)$ such that $\beta \boldsymbol{\tilde{y}^T D} \beta \boldsymbol{\tilde{y}} = 1$. Clearly, the vector $\beta \boldsymbol{\tilde{y}} \in \mathcal{Y}$. Moreover,

$$\beta \boldsymbol{\tilde{y}^T L} \beta \boldsymbol{\tilde{y}} = \beta^2 \boldsymbol{\tilde{y}^T L \tilde{y}} < \boldsymbol{\tilde{y}^T L \tilde{y}} < \boldsymbol{y^{*T} L y^*}, \tag{A.9}$$

where the first inequality follows from Equation (A.3) in combination with the fact that $\boldsymbol{A}$ is positive valued and that $\boldsymbol{\tilde{y}}$ is non-constant. Inequality (A.9) is a contradiction to $\boldsymbol{y^{*T} L y^*}$ being minimal. This concludes the proof of the lemma.

□

Lemmata 1 and 2 jointly imply Theorem 1. In particular, according to Lemma 2

$$\boldsymbol{y^*} := \underset{\boldsymbol{y^T D y}=1, \boldsymbol{y^T D 1}=0, y_i \leq y_j \ \forall i \leq j}{\arg\min} \boldsymbol{y^T L y}$$

is in the interior of set $\mathcal{A} := \left\{ \boldsymbol{y} \in \mathbb{R}^I : y_i \leq y_j \ \forall i \leq j \right\}$. On the one hand, this implies that $\boldsymbol{y^*}$ must be strictly monotonic by the definition of set $\mathcal{A}$. On the other hand, the fact that $\boldsymbol{y^*}$ is in the interior of set $\mathcal{A}$ implies that it must be the eigenvector corresponding to the second smallest eigenvalue of (5), by Lemma 1.[29]

□

References

Acemoglu, D., Antràs, P., and Helpman, E. (2007). Contracts and technology adoption. *The American Economic Review*, 97(3):916–943.

[29]Note that Inequality (A.9) is strict, i.e., moving from the boundary of set $\mathcal{A}$ to the interior strictly decreases the objective function. It follows that in case of multiplicity of λ_2, all associated eigenvectors must be strictly monotonic, for if not, moving in the direction of the non-monotonic eigenvector would allow to approach the boundary of set $\mathcal{A}$ without changing the objective function.

Angrist, N., Goldberg, P. K., and Jolliffe, D. (2021). Why is growth in developing countries so hard to measure? *Journal of Economic Perspectives*, 35(3):215–242.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers - International Monetary Fund*, 16(1):159.

Atkin, D., Costinot, A., and Fukui, M. (2021). Globalization and the ladder of development: Pushed to the top or held at the bottom? Working Paper 29500, NBER.

Balassa, B. (1965). Trade liberalisation and 'revealed' comparative advantage. *The Manchester School*, 33(2):99–123.

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., and Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, 51(3):104450.

Balland, P.-A. and Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, 93(1):1–23.

Baragwanath, K., Goldblatt, R., Hanson, G., and Khandelwal, A. K. (2021). Detecting urban markets with satellite imagery: An application to India. *Journal of Urban Economics*, 125:103173.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396.

Buera, F. J., Kaboski, J. P., Rogerson, R., and Vizcaino, J. I. (2022). Skill-biased structural change. *The Review of Economic Studies*, 89(2):592–625.

Bustos, S., Gomez, C., Hausmann, R., and Hidalgo, C. A. (2012). The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLOS ONE*, 7(11):e49393. doi:10.1371/journal.pone.0049393.

Bustos, S. and Yildirim, M. A. (2020). Production ability and economic growth. *Research Policy*, 104153.

Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., and Tacchella, A. (2012). A network analysis of countries' export flows: Firm grounds for the building blocks of the economy. *PLoS ONE*, 7(10):e47278.

Caselli, F. (2005). Chapter 9: Accounting for cross-country income differences. In Aghion, P. and Durlauf, S. N., editors, *Handbook of Economic Growth*, volume 1, pages 679–741. Elsevier.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics*, 82(2):152–167.

Chung, F. R. K. (1997). *Spectral Graph Theory.* American Mathematical Society, Providence, RI.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5):2031–2059.

Correia, S. (2017). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Mimeo.

Correia, S., Guimarães, P., and Zylkin, T. (2019a). ppmlhdfe: Fast poisson estimation with high-dimensional fixed effects. arxiv:1903.01690.

Correia, S., Guimarães, P., and Zylkin, T. (2019b). Verifying the existence of maximum likelihood estimates for generalized linear models. arxiv:1903.01633.

Costinot, A. (2009a). An elementary theory of comparative advantage. *Econometrica*, 77(4):1165–1192.

Costinot, A. (2009b). On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264.

Costinot, A., Donaldson, D., and Komunjer, I. (2012). What goods do countries trade? A quantitative exploration of Ricardo's ideas. *Review of Economic Studies*, 79(2):581–608.

Costinot, A. and Vogel, J. (2010). Matching and inequality in the world economy. *Journal of Political Economy*, 118(4):747–786.

Costinot, A. and Vogel, J. (2015). Beyond Ricardo: Assignment models in international trade. *Annual Review of Economics*, 7(1):31–62.

Cuñat, A. and Melitz, M. J. (2012). Volatility, labor market flexibility, and the pattern of comparative advantage. *Journal of the European Economic Association*, 10(2):225–254.

Davis, D. R. and Dingel, J. I. (2020). The comparative advantage of cities. *Journal of International Economics*, 123:103291.

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *The Review of Economics and Statistics*, 87(1):1–19.

Deaton, A. and Heston, A. (2010). Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics*, 2(4):1–35.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the

Penn World Table. *American Economic Review*, 105(10):3150–3182.

Feenstra, R. C., Ma, H., Neary, P. J., and Prasada Rao, D. (2013). Who shrunk China? Puzzles in the measurement of real GDP. *The Economic Journal*, 123(573):1100–1129.

Fouss, F., Saerens, M., and Shimbo, M. (2016). *Algorithms and Models for Network Data and Link Analysis.* Cambridge University Press.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–3153.

Gersbach, H., Schetter, U., and Schmassmann, S. (2019). From local to global: A unified theory of public basic research. Discussion Paper 13833, CEPR.

Golub, G. H. and van Loan, C. F. (2013). *Matrix Computations.* The Johns Hopkins University Press, Baltimore, MD, 4th edition.

Gomez-Lievano, A. and Patterson-Lomba, O. (2021). Estimating the drivers of urban economic complexity and their connection to economic performance. *Royal Society Open Science*, 8(9):210670.

Grossman, G. M., Helpman, E., and Kircher, P. (2017). Matching, sorting, and the distributional effects of international trade. *Journal of Political Economy*, 125(1):224–264.

Hall, K. M. (1970). An r-dimensional quadratic placement algorithm. *Management Science*, 17(3):219–229.

Hall, R. E. and Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1):83–116.

Hanson, G. H., Lind, N., and Muendler, M.-A. (2015). The dynamics of comparative advantage. Working Paper 21753, NBER.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934). *Inequalities.* Cambridge University Press, London, UK.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., and Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. *World Development*, 93:75–93.

Harvard Growth Lab (2021). *Atlas of Economic Complexity.* https://dataverse.harvard.edu/dataverse/atlas (retrieved in August 2021).

Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simoes, A., and Yildirim, M. A. (2011). *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* https://atlas.media.mit.edu/atlas/.

Hausmann, R., Hwang, J., and Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12(1):1–25.

Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4, chapter 3, pages 131–195. Elsevier.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring eonomic growth from outer space. *American Economic Review*, 102(2):994–1028.

Hendricks, L. and Schoellman, T. (2022). Skilled labor productivity and cross-country income differences. *American Economic Journal: Macroeconomics*, forthcoming.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 3(2):92–113.

Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.

Hsieh, C.-T. and Klenow, P. J. (2010). Development accounting. *American Economic Journal: Macroeconomics*, 2(1):207–223.

Javorcik, B. S., Turco, A. L., and Maggioni, D. (2018). New and improved: Does FDI boost production complexity in host countries? *The Economic Journal*, 128(614):2507–2537.

Jerven, M. (2013). *Poor Numbers. How We Are Misled by African Development Statistics and What to Do about It.* Cornell University Press, Ithaca, NY.

Johnson, S., Larson, W., Papageorgiou, C., and Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics*, 60(2):255–274.

Jones, C. I. (2016). Chapter 1: The facts of economic growth. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 3–69. Elsevier.

Krugman, P. (1985). A 'technology gap? model of international trade. In Jungenfelt, K. and Hague, D., editors, *Structural Adjustment in Developed Open Economies*, pages 35–61. Palgrave Macmillan UK, London.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

Levchenko, A. A. and Zhang, J. (2016). The evolution of comparative advantage:

Measurement and welfare implications. *Journal of Monetary Economics*, 78:96–111.

Lo Turco, A. and Maggioni, D. (2020). The knowledge and skill content of production complexity. *Research Policy*, 104059.

Maggioni, D., Lo Turco, A., and Gallegati, M. (2016). Does product complexity matter for firms' output volatility? *Journal of Development Economics*, 121:94–109.

Malmberg, H. (2020). Skill bias and development – evidence from trade and industry data. Technical report.

Martinez, L. R. (2022). How much should we trust the dictator's GDP growth estimates? *Journal of Political Economy*, forthcoming.

McNerney, J., Li, Y., Gomez-Lievano, A., and Neffke, F. (2021). Bridging the short-term and long-term dynamics of economic structural change. Technical report. arXiv: 2110.09673.

Mealy, P., Farmer, J. D., and Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, 5(1):eaau1705.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600.

O'Clery, N., Yildirim, M. A., and Hausmann, R. (2021). Productive ecosystems and the arrow of development. *Nature Communications*, 12(1):1479.

Petralia, S., Balland, P.-A., and Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, 46(5):956–969.

Pinkovskiy, M. and Sala-i Martin, X. (2016). Lights, camera ... income! Illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, 131(2):579–631.

Poncet, S. and Starosta de Waldemar, F. (2013). Export upgrading and growth: The prerequisite of domestic embeddedness. *World Development*, 51:104–118.

Rajan, R. G. and Zingales, L. (1998). Financial dependence and growth. *The American Economic Review*, 88(3):559–586.

Ravallion, M. (2003). Measuring aggregate welfare in developing countries: How well do national accounts and surveys agree? *The Review of Economics and Statistics*, 85(3):645–652.

Rossi, F. (2022). The relative efficiency of skilled labor across countries: Measurement and interpretation. *American Economic Review*, 112(1):235–266.

Sala-i Martin, X. and Artadi, E. V. (2004). *The Global Competitiveness Report 2004-2005*, chapter 1.3, pages 51–80. Palgrave Macmillan. Downloaded from http://www.

salaimartin.com/media/pdf/1.3_The_Global_Comp_Index.pdf in August 2019.

Sampson, T. (2014). Selection into trade and wage inequality. *American Economic Journal: Microeconomics*, 6(3):157–202.

Sattinger, M. (1975). Comparative advantage and the distributions of earnings and abilities. *Econometrica*, 43(3):455–468.

Schetter, U. (2020). Quality differentiation, comparative advantage, and international specialization across products. Research Fellow and Graduate Student Working Paper 126, CID at Harvard University. http://dx.doi.org/10.2139/ssrn.3091581.

Sciarra, C., Chiarotti, G., Ridolfi, L., and Laio, F. (2020). Reconciling contrasting views on economic complexity. *Nature Communications*, 11(1):3352.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Springer Science+Business Media.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Shikher, S. (2017). The impact of educated labor on technology adoption and comparative advantage. Mimeo.

Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641–658.

Subramanian, A. (2019). India's GDP mis-estimation: Likelihood, magnitudes, mechanisms, and implications. Technical Report 354.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports*, 2:723. DOI: 10.1038/srep00723.

Tørsløv, T., Wier, L., and Zucman, G. (2018). The missing profits of nations. Working Paper 24701, NBER, Cambridge, MA.

van Dam, A., Dekker, M., Morales-Castilla, I., Rodríguez, M. Á., Wichmann, D., and Baudena, M. (2021). Correspondence analysis, spectral clustering and graph embedding: Applications to ecology and economic complexity. *Scientific Reports*, 11(1):8926.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

Whittaker, R. H. (1967). Gradient analysis of vegetation. *Biological Reviews*, 42(2):207–264.

Wooldridge, J. M. (1999a). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

Wooldridge, J. M. (1999b). Quasi-likelihood methods for count data. In *Handbook of Applied Econometrics Volume 2: Microeconomics*, pages 321–368. John Wiley & Sons, Ltd.

World Bank (2021a). *World Development Indicators (database)*. https://datatopics.worldbank.org/world-development-indicators/ (retrieved on 2 August 2021).

World Bank (2021b). *Worldwide Governance Indicators (database)*. govindicators.org (retrieved on 28 October 2021).

Yildirim, M. A. (2021). Sorting, matching, and economic complexity. CID Faculty Working Paper 392.

Supplementary Material

S1 Detailed Rankings

S1.1 Full List of Country Rankings

Table S1: Country Rankings of Economic Complexity

Country	PPML	PPMLrank	OLS	OLSrank	RCA	RCArank	ECI	ECIrank
JPN	1.57	1	1.49	1	1.66	1	2.15	1
KOR	1.41	2	1.33	2	1.47	2	1.79	4
DEU	1.38	3	1.24	4	1.34	4	1.84	3
CHE	1.34	4	1.27	3	1.42	3	2.04	2
CZE	1.29	5	1.23	5	1.29	6	1.56	9
AUT	1.27	6	1.20	7	1.23	8	1.60	7
FIN	1.25	7	1.21	6	1.31	5	1.47	12
SWE	1.24	8	1.17	8	1.23	9	1.65	6
BEL	1.24	9	1.14	10	1.16	13	1.28	17
GBR	1.21	10	1.08	11	1.18	10	1.59	8
FRA	1.20	11	1.04	18	1.16	14	1.33	15
USA	1.20	12	1.07	12	1.18	11	1.53	10
SGP	1.19	13	1.05	15	1.25	7	1.77	5
SVN	1.18	14	1.16	9	1.16	12	1.47	11
ITA	1.16	15	1.01	21	1.14	16	1.16	19
CHN	1.13	16	1.06	14	1.15	15	1.05	22
HUN	1.08	17	1.02	20	1.10	17	1.28	16
SVK	1.07	18	1.04	17	1.08	18	1.28	18
NLD	1.05	19	0.94	25	0.98	22	1.10	20
POL	1.04	20	0.96	24	1.03	20	1.03	24
NOR	1.04	21	1.05	16	1.07	19	0.92	27
ISR	1.02	22	1.02	19	0.96	26	1.40	14
IRL	1.00	23	1.06	13	1.02	21	1.41	13
DNK	0.99	24	0.99	22	0.97	23	1.04	23
ESP	0.98	25	0.81	34	0.93	29	0.77	34
MYS	0.97	26	0.93	27	0.97	25	1.09	21
RUS	0.96	27	0.91	30	0.92	30	0.54	43
EST	0.94	28	0.92	28	0.93	28	0.99	25
LTU	0.93	29	0.87	31	0.89	31	0.89	30
ROU	0.93	30	0.93	26	0.97	24	0.98	26
CAN	0.90	31	0.98	23	0.94	27	0.78	33
THA	0.87	32	0.79	37	0.87	32	0.89	29
PRT	0.86	33	0.85	33	0.79	38	0.58	41
HRV	0.86	34	0.87	32	0.81	34	0.82	31
BGR	0.84	35	0.91	29	0.80	36	0.64	38
IND	0.81	36	0.79	35	0.80	35	0.50	44
LVA	0.80	37	0.79	36	0.76	40	0.75	35
BLR	0.77	38	0.75	40	0.84	33	0.80	32
AUS	0.74	39	0.78	38	0.67	44	-0.00	65
MEX	0.74	40	0.71	45	0.78	39	0.91	28
ARE	0.74	41	0.77	39	0.65	45	0.08	62
TUR	0.74	42	0.72	42	0.74	41	0.57	42
BRA	0.73	43	0.73	41	0.79	37	0.35	49

Table S1: Country Rankings of Economic Complexity

Country	PPML	PPMLrank	OLS	OLSrank	RCA	RCArank	ECI	ECIrank
ZAF	0.71	44	0.71	44	0.71	43	0.37	48
UKR	0.69	45	0.71	43	0.74	42	0.42	47
SAU	0.61	46	0.66	48	0.64	46	0.74	36
NZL	0.59	47	0.69	46	0.61	47	0.42	46
PHL	0.58	48	0.61	49	0.57	48	0.63	39
GRC	0.56	49	0.66	47	0.55	49	0.22	53
BIH	0.51	50	0.52	52	0.55	50	0.60	40
CYP	0.49	51	0.45	54	0.38	54	0.69	37
VNM	0.46	52	0.55	51	0.48	51	-0.02	67
IDN	0.46	53	0.56	50	0.43	52	-0.12	72
KWT	0.43	54	0.46	53	0.37	55	0.17	54
QAT	0.39	55	0.40	56	0.19	65	0.22	52
TUN	0.37	56	0.43	55	0.36	57	0.25	51
ARG	0.35	57	0.28	59	0.39	53	0.00	64
CHL	0.32	58	0.37	57	0.33	59	-0.02	68
BHR	0.31	59	0.22	65	0.37	56	0.42	45
COL	0.30	60	0.25	61	0.34	58	0.16	56
CRI	0.29	61	0.22	64	0.29	60	0.30	50
PAN	0.24	62	0.21	66	0.25	62	0.11	58
KAZ	0.21	63	0.13	71	0.28	61	-0.04	69
MKD	0.18	64	0.26	60	0.20	64	0.13	57
IRN	0.16	65	0.24	62	0.12	69	-0.25	75
JOR	0.15	66	0.22	63	0.16	68	0.08	59
GEO	0.12	67	0.17	67	0.11	70	-0.06	70
EGY	0.10	68	0.29	58	0.18	66	-0.02	66
URY	0.09	69	0.13	70	0.21	63	0.16	55
MUS	0.07	70	0.09	73	0.09	71	-0.13	73
SLV	0.01	71	0.12	72	0.18	67	0.08	60
LBN	-0.05	72	0.09	74	-0.03	74	0.07	63
MAR	-0.05	73	0.14	69	-0.07	77	-0.58	90
PAK	-0.07	74	0.15	68	0.00	73	-0.41	83
OMN	-0.10	75	-0.12	81	-0.03	75	-0.07	71
ARM	-0.11	76	-0.04	77	-0.14	80	-0.36	80
GTM	-0.12	77	-0.01	76	0.03	72	-0.30	78
LKA	-0.15	78	0.05	75	-0.05	76	-0.48	85
MDA	-0.17	79	-0.05	79	-0.09	78	-0.31	79
TTO	-0.19	80	-0.24	82	-0.20	82	0.08	61
PER	-0.21	81	-0.05	78	-0.14	81	-0.43	84
DZA	-0.24	82	-0.46	89	-0.78	98	-0.92	101
KEN	-0.29	83	-0.10	80	-0.09	79	-0.25	74
COG	-0.36	84	-0.35	84	-0.55	90	-0.71	92
AGO	-0.37	85	-0.56	93	-1.22	113	-1.25	117
AZE	-0.38	86	-0.37	86	-0.52	88	-0.89	100
ALB	-0.46	87	-0.36	85	-0.36	83	-0.38	81
KHM	-0.48	88	-0.39	87	-0.54	89	-0.77	95
DOM	-0.48	89	-0.40	88	-0.43	85	-0.28	77
BGD	-0.50	90	-0.29	83	-0.57	94	-0.92	102
UZB	-0.52	91	-0.53	92	-0.46	87	-0.69	91
GAB	-0.55	92	-0.64	97	-0.96	107	-1.10	112
ECU	-0.56	93	-0.51	90	-0.60	95	-0.94	103
CIV	-0.57	94	-0.52	91	-0.64	97	-1.02	108
ZMB	-0.58	95	-0.68	98	-0.43	84	-0.55	88
HND	-0.64	96	-0.57	94	-0.43	86	-0.54	87

Table S1: Country Rankings of Economic Complexity

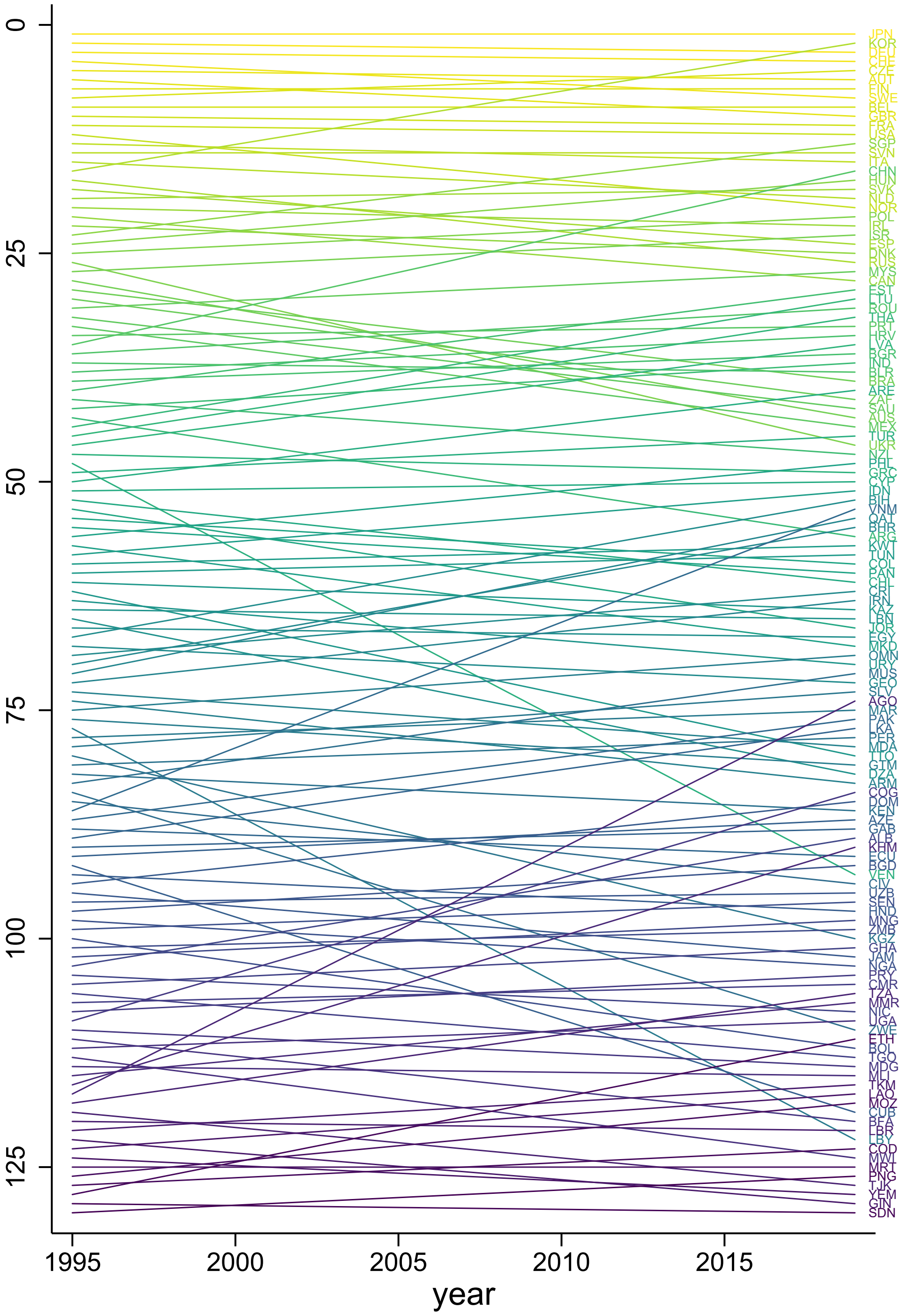
Country	PPML	PPMLrank	OLS	OLSrank	RCA	RCArank	ECI	ECIrank
SEN	-0.66	97	-0.61	95	-0.63	96	-0.79	96
GHA	-0.68	98	-0.70	100	-0.78	100	-1.03	109
JAM	-0.71	99	-0.79	103	-0.56	92	-0.27	76
KGZ	-0.73	100	-0.75	102	-0.56	91	-0.56	89
MNG	-0.73	101	-0.63	96	-0.78	99	-1.01	107
MMR	-0.82	102	-0.70	99	-0.81	101	-0.94	104
NGA	-0.84	103	-0.87	104	-1.33	117	-1.64	126
PRY	-0.89	104	-0.73	101	-0.57	93	-0.38	82
VEN	-0.94	105	-1.09	110	-1.21	111	-1.12	114
NIC	-1.03	106	-0.89	105	-0.84	102	-0.87	99
TGO	-1.07	107	-1.19	113	-0.94	106	-0.75	93
BOL	-1.09	108	-1.02	108	-0.88	103	-0.95	105
ZWE	-1.12	109	-1.11	111	-0.92	105	-0.83	97
COD	-1.15	110	-0.97	106	-1.21	112	-1.00	106
MDG	-1.15	111	-1.00	107	-1.00	108	-1.07	110
LAO	-1.22	112	-1.27	116	-1.15	110	-0.87	98
CMR	-1.24	113	-1.13	112	-1.56	120	-1.58	125
TKM	-1.28	114	-1.31	117	-0.90	104	-0.77	94
LBR	-1.29	115	-1.09	109	-1.94	125	-1.53	124
BFA	-1.30	116	-1.23	115	-1.32	116	-1.52	123
TZA	-1.34	117	-1.21	114	-1.11	109	-1.11	113
ETH	-1.51	118	-1.32	118	-1.31	115	-1.29	118
MLI	-1.53	119	-1.60	120	-2.05	126	-2.03	129
CUB	-1.54	120	-1.97	125	-1.55	119	-0.52	86
MWI	-1.54	121	-1.49	119	-1.28	114	-1.20	115
LBY	-1.68	122	-1.74	123	-1.91	124	-1.30	119
UGA	-1.69	123	-1.65	121	-1.41	118	-1.33	120
MOZ	-1.83	124	-1.81	124	-1.71	122	-1.40	122
PNG	-1.87	125	-2.31	127	-2.27	128	-1.75	127
MRT	-2.00	126	-1.68	122	-1.80	123	-1.36	121
TJK	-2.16	127	-2.42	128	-1.66	121	-1.07	111
GIN	-2.17	128	-2.04	126	-2.65	130	-2.39	130
YEM	-2.69	129	-2.95	129	-2.08	127	-1.21	116
SDN	-2.75	130	-3.13	130	-2.48	129	-1.76	128

This table shows country rankings for the year 2019 using trade data at the HS4d classification level. PPML (OLS) corresponds to the second eigenvector of $\boldsymbol{Ly} = \lambda \boldsymbol{Dy}$ using PPML (OLS) in the first-step regression. RCA to the second eigenvector using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. For $X \in \{PPML, OLS, RCA, ECI\}$, 'Xrank' shows the ranking and 'X' the associated eigenvector. All eigenvectors have been normalized to have mean 0 and standard deviation 1.

S1.2 Country Rankings Over Time

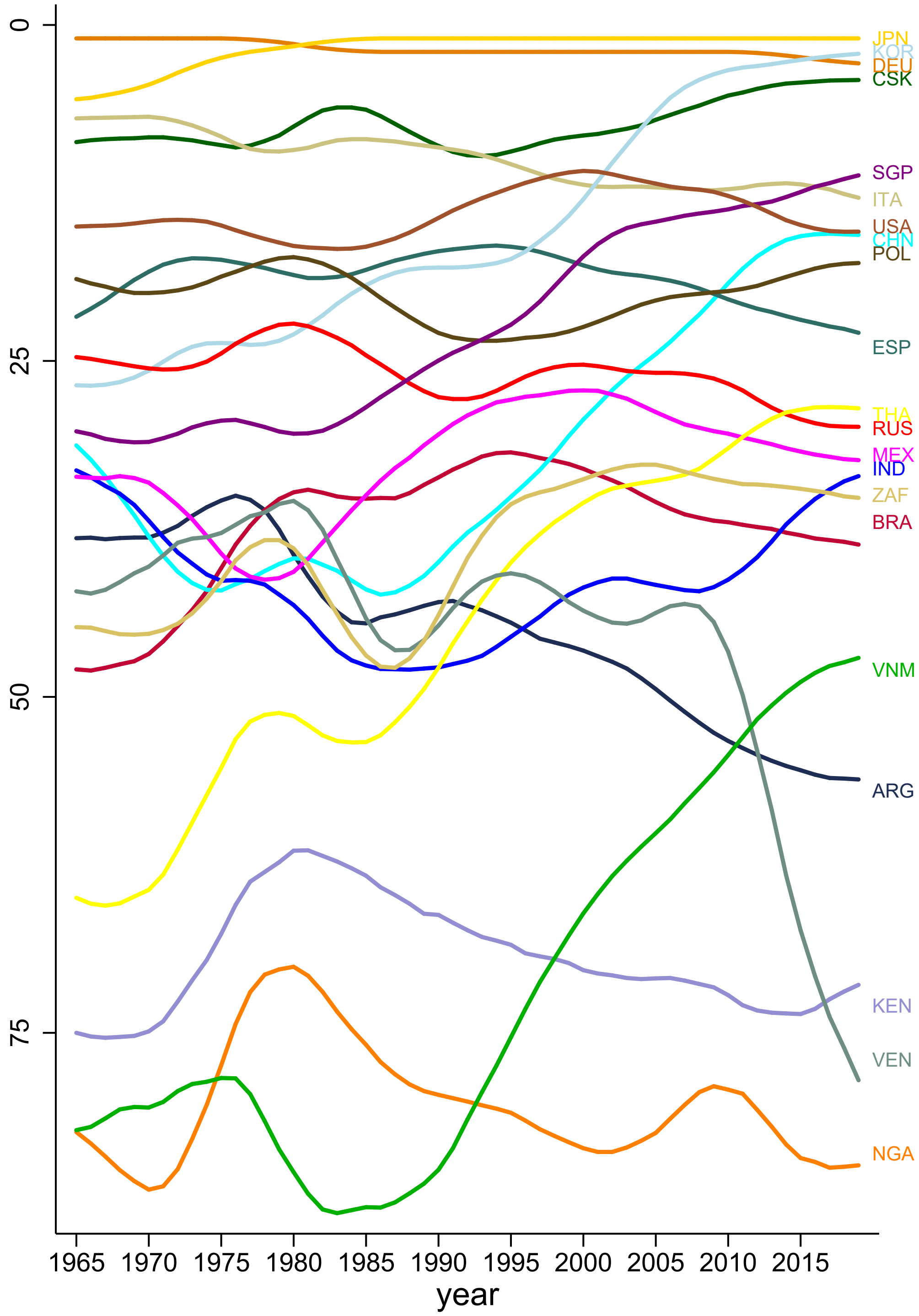
The following figures summarize the evolution over time of the country rankings. Further details are provided in the notes of the figures.

Figure S1: Changes of Country Rankings from 1995 to 2019



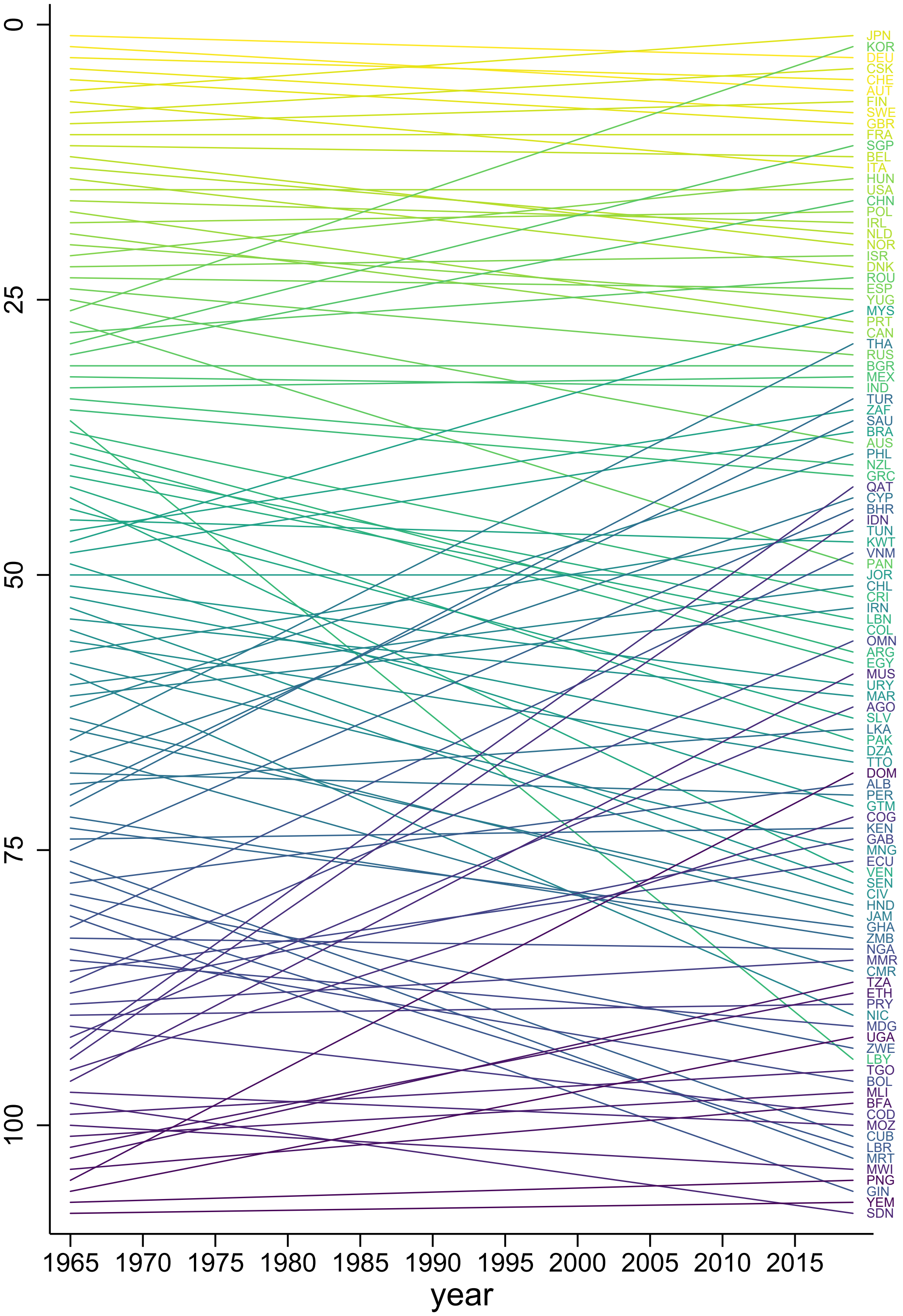
Notes: The figure shows for all countries the change of their 11-year weighted moving average ranking from 1995 to 2019 (the endpoints shown reflect one-sided moving averages). Lines are color-coded based on their 1995 ranking. Weights: $\{1/36, 2/36, ..., 6/36, 5/36, ..., 1/36\}$.

Figure S2: Country Rankings from 1965 to 2019



Notes: The figure shows for a selection of countries their 11-year weighted moving average ranking from 1965 to 2019 with weights $\{1/36, 2/36, ..., 6/36, 5/36, ..., 1/36\}$. Rankings have been computed using SITC trade data for a balanced panel of 108 countries. The list of countries may be seen from Figure S3 and has been derived as follows: (i) The following countries have been consistently aggregated: (a) former Soviet Union member states as RUS; (b) former republics of Yugoslavia as YUG; (c) Czech Republic and Slovakia as CSK; (d) Sudan and South Sudan as SDN. (ii) All countries were dropped that were not consistently part of the list of countries included in the SITC country ranking of the Atlas of Economic Complexity (I thank Sebastian Bustos for sharing the list of countries).

Figure S3: Changes of Country Rankings from 1965 to 2019



Notes: The figure shows for all countries the change of their 11-year weighted moving average ranking from 1965 to 2019 (the endpoints shown reflect only one-sided moving averages). Rankings have been computed using SITC trade data. Lines are color-coded based on their 1965 ranking. Weights: $\{1/36, 2/36, ..., 6/36, 5/36, ..., 1/36\}$.

S2 Details on Numerical Simulations of Section 3.4

This appendix provides details and further results for the Monte Carlo simulation of Section 3.4.

As discussed in Section 3.4, the figures shown in Table 1 are based on $80k$ randomly generated matrices $\boldsymbol{A}$—$10k$ for each column. To generate these matrices, start from symmetric supermodular matrices $\tilde{\boldsymbol{A}}$, add noise to these matrices, and finally exponentiate them elementwise.

To generate the supermodular matrices $\tilde{\boldsymbol{A}}$, I make use of the fact that local (log-)supermodularity is necessary and sufficient for a matrix to be (log-) supermodular. That is, for every 2 by 2 block of $\boldsymbol{A}$ with elements in contiguous rows and columns, $i' > i$ and $k' > k$, respectively, it must hold

$$A_{ik} \cdot A_{i'k'} > A_{i'k} \cdot A_{ik'}.$$

Hence, to randomly draw a positive supermodular and symmetric $I \times I$ matrix $\tilde{\boldsymbol{A}}$, I proceed as follows:

1. Randomly draw an $I \times I$ matrix $\boldsymbol{R}$ with elements R_{ij} from a uniform distribution on $[0, 1]$.
2. Randomly draw an index i from the discrete uniform distribution with support $i \in \{1, 2, ..., I\}$.
3. Set $\tilde{A}_{i,:} = R_{i,:} * 100$ and $\tilde{A}_{:,i} = R_{i,:}^T * 100$.[30]
4. Fill $\tilde{\boldsymbol{A}}$ by choosing $\tilde{A}_{l,k} = \tilde{A}_{k,l}$ as follows:
   (a) Set element $\tilde{A}_{k,l} = \tilde{A}_{k,l-1} + \tilde{A}_{k+1,l} - \tilde{A}_{k+1,l-1} - |R_{k,l}|$ and element $\tilde{A}_{l,k} = \tilde{A}_{k,l}$ for $k = i-1, i-2, ..., 1$ and $l = i+1, i+2, ..., I$.
   (b) Set element $\tilde{A}_{k,l} = \tilde{A}_{k+1,l} + \tilde{A}_{k,l+1} - \tilde{A}_{k+1,l+1} + |R_{k,l}|$ and element $\tilde{A}_{l,k} = \tilde{A}_{k,l}$ for $k = i-1, i-2, ..., 1$ and $l = i-1, i-2, ..., k$.
   (c) Set element $\tilde{A}_{k,l} = \tilde{A}_{k,l-1} + \tilde{A}_{k-1,l} - \tilde{A}_{k-1,l-1} + |R_{k,l}|$ and element $\tilde{A}_{l,k} = \tilde{A}_{k,l}$ for $k = i+1, i+2, ..., I$ and $l = k, k+1, ..., I$.

[30] The elements are scaled to increase the variance of elements in the initial row / column vis-à-vis the variance of shocks that govern the log-supermodularity of the matrix (see next step). This is to emphasize that log-supermodularity is a 'diff-in-diff' condition and not concerned with absolute sizes of elements in different rows and columns. Simulation results are virtually the same when using a scaling factor of 1.

Table S2: Robustness with 50×50 Matrices

	Upper bound of uniform distribution							
	0.0	0.3	1.0	3.0	10.0	50.0	100.0	500.0
Avg rank correlation	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.979
Avg share rows/columns correct	1.000	1.000	1.000	1.000	1.000	0.950	0.698	0.154
Share of iterations all correct	1.000	1.000	1.000	1.000	1.000	0.282	0.000	0.000
Avg share LSM	1.000	0.929	0.773	0.607	0.533	0.507	0.503	0.501

This table shows summarizing statistics for $80k$ randomly generated 50×50 matrices—$10k$ for each column. All else is the same as described in the notes of Table 1.

Table S3: Robustness with 10×10 Matrices

	Upper bound of uniform distribution							
	0.0	0.3	1.0	3.0	10.0	50.0	100.0	500.0
Avg rank correlation	1.000	1.000	1.000	0.999	0.960	0.435	0.301	0.219
Avg share rows/columns correct	1.000	1.000	1.000	0.987	0.611	0.168	0.138	0.122
Share of iterations all correct	1.000	1.000	1.000	0.939	0.090	0.000	0.000	0.000
Avg share LSM	1.000	0.926	0.766	0.603	0.532	0.507	0.504	0.500

This table shows summarizing statistics for $80k$ randomly generated 10×10 matrices—$10k$ for each column. All else is the same as described in the notes of Table 1.

This procedure results in a matrix that is positive, symmetric and supermodular. I add to this matrix another $I \times I$ symmetric random matrix $\boldsymbol{S}$, drawn iid from a uniform distribution with lower bound 0 and upper bound as shown in the respective column-header in Table 1. All elements of this matrix are then normalized by 1/5 of the largest element.[31] Finally, the matrix is exponentiated element-by-element to derive at the positive and symmetric matrix $\boldsymbol{A}$. For each of these matrices, I then compare the ranking of rows and columns implied by the eigenvector corresponding to the second smallest eigenvalue of (5) to the 'true' ranking of the underlying log-supermodular matrix $\tilde{\boldsymbol{A}}$, i.e., to a vector with elements $[1, 2, ..., I]$, where I is the size of the matrix. To determine the sign of the eigenvector, I require that the sum of its first three elements must be positive, analogous to some outside information that we might use in practical applications.

Summarizing statistics for these simulations are provided in Table 1. The insights from this table do not hinge on the assumption of a uniform distribution, and the main message is the same when using e.g. normal distributions instead. The ranking is, however, somewhat less robust to noise when considering smaller matrices as shown in Tables S2

[31] This normalization avoids very large values in the final matrix $\boldsymbol{A}$ that might cause computational problems. The normalization does not affect the generalized eigenvectors (5) of matrix $\boldsymbol{A}$.

and S3, respectively. Considering the discussion of Section 3.4 this may not come as a surprise. The noise added to the log-supermodular matrix has a bigger impact on log-supermodularity of neighboring elements than on the log-supermodularity of elements at greater distances. For large enough matrices, the second vector exploits this structure at greater distances. For small matrices this is not possible. Still, the results in Tables S2 and S3 confirm that the ranking is very robust to random noise as long as the noise is not too big relative to the size of matrix $\boldsymbol{A}$.

S3 Micro-foundation for Log-supermodular Country-Country Matrix Based on 'Nested' Matrix

This appendix provides a micro-foundation for a log-supermodular country-country matrix based on a perfectly nested country-product matrix, as discussed in Section 6.1.

Consider a binary country-product matrix $\boldsymbol{M}$ that indicates for each country the set of products that it makes. Suppose that this matrix is perfectly nested. That is, for every $i' > i$ it holds that $\mathcal{S}_i \subset \mathcal{S}_{i'}$, where $\mathcal{S}_i$ denotes the set of products that country i makes (i.e., with element $M_{is} = 1$). Then, the country-country similarity matrix $\boldsymbol{A}$ with element $A_{i'i}$

$$A_{i'i} := \sum_{s \in \mathcal{S}} M_{i's} M_{is}$$

is log-supermodular as the following proposition shows:

Proposition 2

For every two pairs of countries, $i' > i$ and $k' > k$ it holds

$$A_{i'k'} A_{ik} \geq A_{i'k} A_{ik'}.$$

The inequality is strict whenever $i' > k$ and $i < k'$.

Proof: Note first that with a nested structure it holds for $i' > i$

$$\sum_{s \in \mathcal{S}} M_{i's} M_{is} = \sum_{s \in \mathcal{S}} M_{is} =: d_i.$$

Now, three cases need to be distinguished.

(i) $i < i' < k < k'$

$$A_{i'k'} A_{ik} = d_i d_{i'} = A_{i'k} A_{ik'}$$

(ii) $i < k < i' < k'$

$$A_{i'k'} A_{ik} = d_i d_{i'} > d_i d_k = A_{i'k} A_{ik'}$$

(iii) $i < k < k' < i'$

$$A_{i'k'}A_{ik} = d_i d_{k'} > d_i d_k = A_{i'k}A_{ik'}$$

This completes the proof. □

S4 Empirical Appendix

This appendix provides further details on Table 5 (S4.1), variants of Table 2 (S4.2), a variant of Figure 5 with a separate line for each country (S4.3), and a test of log-supermodularity of the product-product similarity matrix underlying the product ranking of Section 6.2 (S4.4).

S4.1 Details on Table 5

This appendix provides further details on Table 5.

'RL' is a measure of required on-the-job learning taken from Costinot (2009b). All other product characteristics are from Levchenko (2007). With the exception of 'EF', these measures are at the 4-digit SIC87 level and based on the NBER-CES Manufacturing Industry Database for the year 1992. 'CI' is capital intensity measured as 1 minus an industry's labor share in value added. 'SI' is skilled labor intensity measured as the share of non-production workers in employment times the labor share in value added. 'ID' is a measure of intermediate input diversification, i.e., 1 over the Herfindahl index of intermediate input use. 'I/Y' is the investment to output ratio computed as investment over (valued added + cost of materials). Lastly, 'EF' is the Rajan and Zingales (1998) measure of external finance dependence computed based on COMPUSTAT at the 3-digit SIC87 level.

Using these product characteristics, the correlations have been computed as follows: (i) 6-digit HS product codes have been matched to SIC87 industry codes using the concordance from WITS (https://wits.worldbank.org/product_concordance.html). In case of the required learning (RL), these have been further matched to SIC72 codes using the concordance table provided by the NBER (https://data.nber.org/nberces/nberces5811/). (ii) The 4-digit HS complexity ranks have been assigned to each of the corresponding 6-digit HS product codes. (iii) Weighted averages of these 6-digit product complexity ranks have been computed for each industry of the respective product characteristics, where the weights are the total global export shares of 6-digit

HS products in the respective industry. (iv) The correlation between the aggregated complexity rank and the respective product characteristic has been computed.

S4.2 Additional Correlates with Country Ranking

Table S4 shows a variant of Table 2 using different country measures as detailed in the notes of the table. Table S5 shows how the change from 1995 to 2019 in a country's rank correlates with changes in the country characteristics of Table 2.

Table S4: Distance to Frontier and Sources of Comparative Advantage in Complex Products

	PPML	log(RD)	ftw85	findevt	logH_L	logK_L
PPML	1.00	0.82	0.71	0.73	0.82	0.86
log(RD)		1.00	0.79	0.61	0.65	0.74
ftw85			1.00	0.63	0.63	0.65
findevt				1.00	0.64	0.72
logH_L					1.00	0.82
logK_L						1.00

This table shows rank correlations between the country ranking of distance to frontier for 1995 and different country characteristics. Log physical and human capital ('log(K_L)' and 'log(H_L)', respectively) is from Hall and Jones (1999) and refers to 1988. 'ftw85' is a measure of 'Legal System and Property Rights' for 1985 from Gwartney and Lawson (2004). 'findevt' is credit extended by banks and non-bank financial intermediaries to the private sector divided by GDP, averaged over 1980-89 from Beck et al. (2000). 'log(RD)' is the 19-year centered moving average of a country's log share of GDP invested in R&D for 1995 from World Bank (2021a). I thank Davin Chor for sharing the data on log(K_L), log(H_L), ftw85, and findevt.

Table S5: Changes in Distance to Frontier and Changes in Country Characteristics

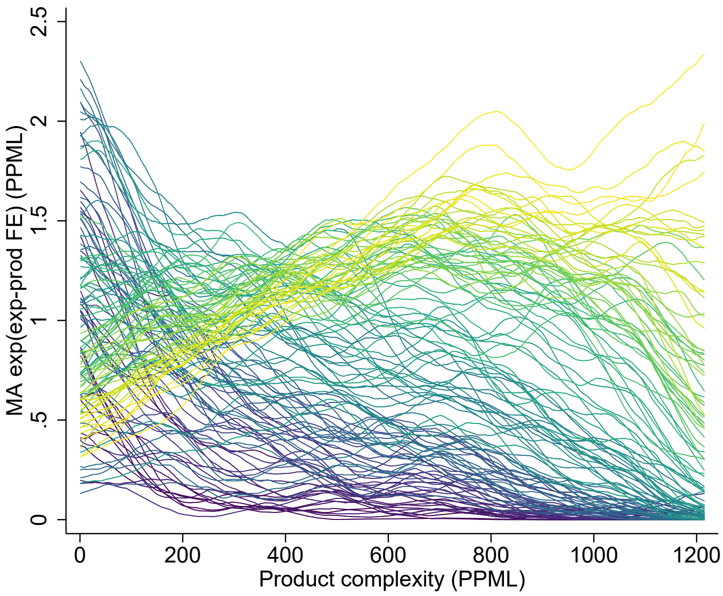
	dPPML	dlog(RD)	dGE	dlog(HC)	dlog(K/L)	dlog(TFP)
dPPML	1.00	0.20	0.39	-0.05	0.40	0.41
dlog(RD)		1.00	0.08	0.11	0.16	-0.11
dGE			1.00	-0.08	-0.02	0.39
dlog(HC)				1.00	0.18	-0.00
dlog(K/L)					1.00	0.11
dlog(TFP)						1.00

This table shows correlations between changes in the baseline ranking of distance to frontier (PPML) and changes in different country characteristics. The country characteristics are as detailed in the notes of Table 2. All changes are from 1995 to 2019, with the exception of the government effectiveness indicator (GE), for which there is no data for 1995 and, hence, the change from 1996 to 2019 has been considered.

S4.3 Specialization of Countries Across Products

Figure S4 shows a variant of Figure 5 with a separate line for each country.

Figure S4: Country Specialization Across Products



Notes: The figure shows line plots summarizing country-specialization across products. Products are ordered by their complexity from least to most complex, where the complexity metric from Section 6.2 has been used. Lines are color-coded by country capability with brighter / more yellow indicating higher capability. Each line represents one country and shows the 201-product centered weighted moving average of its exponentiated normalized exporter-product fixed effect (weights: $\{1/(101)^2, 2/(101)^2, ..., 101/(101)^2, 100/(101)^2, ..., 1/(101)^2\}$).

S4.4 Log-supermodularity Test for Product-Product Similarity Matrix

Table S6 summarizes a test of log-supermodularity following Davis and Dingel (2020) for the product-product similarity matrix $\hat{\boldsymbol{B}}$ that has been used for the product-complexity ranking of Section 6.2. To save computational power, the tests start from grouping the 1215 products into 243 bins.

Table S6: Test of Log-supermodularity of Product-Product Similarity Matrix

Bins	# comparisons	Success rate (p-value)
243	432253503	0.846 (0.000)
122	27235890	0.902 (0.000)
61	1673535	0.943 (0.000)
24	37950	0.983 (0.000)
12	2145	0.998 (0.000)
4	15	1.000 (0.000)

This table shows results from a test of log-supermodularity following Davis and Dingel (2020) of the product-product similartiy matrix $\hat{\boldsymbol{B}}$ underlying the product-complexity ranking of Section 6.2. Column 'bins' refers to the number of bins in which the 1215 products have bin grouped and aggregated. Column '# comparisons' specifies the number of inequality tests performed in the respective row. Column 'Success rate' specifies the share of these inequality-tests that satisfied the condition of log-supermodularity. p-values are based on 2000 iterations of a null-model, where in each iteration (i) matrix $\boldsymbol{M}$ with element $M_{is} = \hat{T}_i^s z_i^s$ has been randomly reshuffled s.t. row and column sums were preserved; (ii) the random matrix has been rearranged based on the same eigen-procedure used for the data matrix; (iii) the same test of LSM has been applied as for the data matrix.

S5 Robustness of Rankings

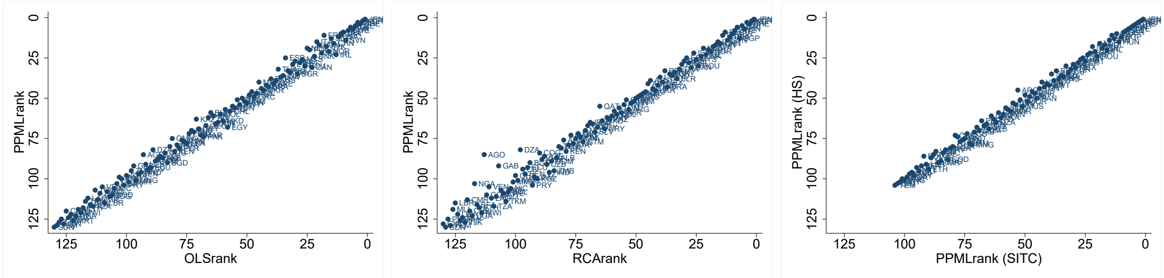
This part of the appendix shows robustness of the country and product ranking. Section S5.1 presents the main robustness checks and shows that the country ranking is robust to the choice of the first-step estimator and with respect to the sample selection and the industry classification considered, respectively. Section S5.2 shows that the rankings are robust to the various data cleaning steps and normalizations, specifically: (i) the minimum number of destinations with non-zero exports per exporter-product; (ii) the minimum cut-off for export values at the exporter-importer-product level; (iii) the normalization of the exporter-product fixed effects; (iv) the winsorizing threshold for the exporter-product fixed effects; and (v) with respect to the year of analysis.

S5.1 Main Robustness Checks for Country Ranking

Panels (a) and (b) of Figure S5 show that the country ranking is highly robust to the choice of the estimator for the exporter-product fixed effects. These figures locate each country in a scatter plot with its rank derived from using PPML in the first step regression on the vertical axis and its rank derived from using OLS in the first

Figure S5: Robustness of Country Rankings

(a) PPML vs OLS (b) PPML vs RCA (c) HS vs SITC



Notes: Figure (a) shows a scatter plot with a country's rank using PPML in the first step regression on the vertical axis and its rank using OLS in the first step regression on the horizontal axis. Figure (b) shows a country's rank using the RCA in the first step on the horizontal axis. Figure (c) starts from the rankings using 4-digit HS trade data for 130 countries on the vertical axis, and using 4-digit SITC trade data for 108 countries on the horizontal axis. Based on these rankings, it then recomputes a country's rank among the set of 104 countries that are included in both the HS- and the SITC-based country rankings. That is, the list of countries in the SITC-based ranking minus Czechoslovakia, the Soviet Union, Sudan, and Yugoslavia (which have consistently been aggregated for the SITC-based ranking). All data refer to year 2019.

step regression on the horizontal axis (panel (a)); panel (b) considers a country's rank derived from using Revealed Comparative Advantages (RCA, Balassa 1965) on the horizontal axis. These figures show tight relationships, with a rank correlation of .995 (panel (a)) and of .982 (panel (b)), respectively.

Panel (c) of Figure S5 compares the baseline ranking using 4-digit HS trade data with a ranking using 4-digit SITC trade data—the latter has been used for the long-run analysis in Supplementary Material S1.2. The rankings not only differ in that they use different product (or industry) classifications, but also different sets of countries, with the SITC-based ranking restricted to a set of 108 countries (or country aggregates) that I can observe for the entire period from 1965 to 2019. Nevertheless, the rankings are very similar with a rank correlation of .996.

Lastly, as shown in Section 6.1, the ranking is even similar when using a binarized country-product matrix based on RCAs.

S5.2 Additional Robustness Checks

This section presents additional robustness checks for the country ranking of Section 5 and for the product ranking of Section 6.2. In turn, it varies the thresholds used for data-cleaning, the censoring threshold for outliers, the normalization of the exporter-product fixed effects, and the year. For each of these robustness checks, rank correlations for the implied country and product rankings, respectively, across different specifications are provided. Apart from the respective robustness check under consideration, data and data cleaning choices are the same as in the baseline specification. That is, I use bilateral trade data for 130 exporters and importers at the 4-digit HS level for the year 2019. Export values of less than USD 1000 at the bilateral-product level are dropped, as well as all of a country's exports of a given product if it does not sell this product to at least 3 destinations. The estimated exporter-product fixed effects are normalized such that for every country i and every product s it holds

$$\sum_{\hat{s}\in\mathcal{S}_i} \delta_i^{\hat{s}} = 0$$
$$\sum_{\hat{i}\in\mathcal{I}^s} \delta_{\hat{i}}^{s} = 0 \ .$$

I then take the square root of the exponentiated fixed effects to account for the fact that the objective is a quadratic form. To correct for outliers, the top 5% of the exporter-product fixed-effects are finally winsorized.

Further details on the various robustness checks are provided in the notes of the respective table.

S5.2.1 Robustness of Country Ranking

Table S7: Robustness of Country Ranking to Minimum Threshold for Number of Export Destinations by Exporter-Product

	cut_1				cut_3				cut_5			
	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI
PPML_cut_1	1.00	0.98	0.98	0.95	0.99	0.98	0.97	0.96	0.98	0.97	0.96	0.96
OLS_cut_1		1.00	0.98	0.95	0.99	1.00	0.99	0.96	0.99	0.99	0.98	0.96
RCA_cut_1			1.00	0.98	0.98	0.98	0.99	0.98	0.98	0.97	0.98	0.98
ECI_cut_1				1.00	0.95	0.94	0.97	0.99	0.94	0.94	0.95	0.97
PPML_cut_3					1.00	1.00	0.99	0.96	1.00	0.99	0.99	0.97
OLS_cut_3						1.00	0.99	0.96	0.99	1.00	0.99	0.96
RCA_cut_3							1.00	0.98	0.99	0.99	0.99	0.98
ECI_cut_3								1.00	0.96	0.95	0.96	0.99
PPML_cut_5									1.00	0.99	0.99	0.97
OLS_cut_5										1.00	0.99	0.96
RCA_cut_5											1.00	0.97
ECI_cut_5												1.00

This table shows rank correlations between different country rankings. PPML (OLS) refers to the country ranking using PPML (OLS) in the first-step regression. RCA refers to the country ranking using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. 'cut_x' indicates that prior to the first-step regression, exporter-product observations have been dropped if the product has not been shipped to at least x destinations. All other specifications are as described at the onset of this appendix.

Table S8: Robustness of Country Ranking to Minimum Threshold for Tradeflows at the Bilateral Product Level

	minx_0				minx_1000				minx_10000			
	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI
PPML_minx_0	1.00	0.99	0.99	0.96	1.00	1.00	0.99	0.96	1.00	0.99	0.98	0.96
OLS_minx_0		1.00	0.99	0.96	0.99	1.00	0.99	0.96	0.99	0.99	0.99	0.96
RCA_minx_0			1.00	0.98	0.99	0.99	1.00	0.98	0.99	0.99	1.00	0.98
ECI_minx_0				1.00	0.96	0.96	0.98	1.00	0.97	0.96	0.98	1.00
PPML_minx_1000					1.00	1.00	0.99	0.96	1.00	0.99	0.98	0.96
OLS_minx_1000						1.00	0.99	0.96	0.99	0.99	0.99	0.96
RCA_minx_1000							1.00	0.98	0.99	0.99	1.00	0.98
ECI_minx_1000								1.00	0.97	0.96	0.98	1.00
PPML_minx_10000									1.00	1.00	0.99	0.97
OLS_minx_10000										1.00	0.99	0.96
RCA_minx_10000											1.00	0.98
ECI_minx_10000												1.00

This table shows rank correlations between different country rankings. PPML (OLS) refers to the country ranking using PPML (OLS) in the first-step regression. RCA refers to the country ranking using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. 'minx_x' indicates that prior to the first-step regression, export values of less than USD x at the bilateral-product level have been dropped. All other specifications are as described at the onset of this appendix.

Table S9: Robustness of Country Ranking to Normalization of Exporter-Product Fixed Effects

	norm_lsum			norm_cossim			norm_vECI			norm_nsqrt			
	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	ECI
PPML_norm_lsum	1.00	1.00	0.99	0.96	0.98	0.98	0.93	1.00	0.99	0.99	0.99	0.98	0.97
OLS_norm_lsum		1.00	0.99	0.96	0.98	0.98	0.94	0.99	1.00	0.99	0.99	0.99	0.97
RCA_norm_lsum			1.00	0.98	0.98	0.98	0.96	0.99	0.99	1.00	0.98	0.98	0.99
PPML_norm_cossim				1.00	0.97	0.96	0.96	0.95	0.95	0.97	0.96	0.96	0.99
OLS_norm_cossim					1.00	1.00	0.98	0.98	0.98	0.98	0.98	0.98	0.97
RCA_norm_cossim						1.00	0.98	0.97	0.98	0.98	0.97	0.98	0.97
PPML_norm_vECI							1.00	0.93	0.94	0.95	0.93	0.95	0.96
OLS_norm_vECI								1.00	0.99	0.99	0.99	0.98	0.97
RCA_norm_vECI									1.00	0.99	0.98	0.98	0.97
PPML_norm_nsqrt										1.00	0.98	0.98	0.98
OLS_norm_nsqrt											1.00	0.99	0.97
RCA_norm_nsqrt												1.00	0.97
ECI													1.00

This table shows rank correlations between different country rankings. PPML (OLS) refers to the country ranking using PPML (OLS) in the first-step regression. RCA refers to the country ranking using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. 'norm_x' indicates which normalization has been used. 'norm_lsum' denotes the baseline normalization. All other normalizations start from this baseline. 'norm_cossim' denotes the cosine similarity, that is, the country-country similarity matrix $\hat{\boldsymbol{A}}$ with elements

$$\hat{A}_{ii'} = \frac{\sum_{s \in \mathcal{S}} z_i^s \hat{T}_i^s z_{i'}^s \hat{T}_{i'}^s}{\sqrt{\left[\sum_{s \in \mathcal{S}} z_i^s \hat{T}_i^s z_i^s \hat{T}_i^s\right] \cdot \left[\sum_{s \in \mathcal{S}} z_{i'}^s \hat{T}_{i'}^s z_{i'}^s \hat{T}_{i'}^s\right]}}.$$

In rows and columns 'norm_vECI', each product is normalized by its 'ubiquity' when computing the country-country similarity matrix, that is matrix $\hat{\boldsymbol{A}}$ has elements

$$\hat{A}_{ii'} = \sum_{s \in \mathcal{S}} \frac{z_i^s \hat{T}_i^s z_{i'}^s \hat{T}_{i'}^s}{\sum_{\hat{i} \in \mathcal{I}} z_{\hat{i}}^s \hat{T}_{\hat{i}}^s}.$$

Finally, in rows and columns 'norm_nsqrt', the exponentiated exporter-product fixed effects are used directly, instead of the square-root thereof. All other specifications are as described at the onset of this appendix.

Table S10: Robustness of Country Ranking to Winsorizing Threshold for Outliers of Exporter-Product Fixed Effects

	cens_90			cens_92.5			cens_95			cens_97.5			cens_99			
	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	ECI
PPML_cens_90	1.00	0.99	0.99	0.95	1.00	0.99	0.99	1.00	0.99	0.98	0.99	0.99	0.98	0.98	0.98	0.98
OLS_cens_90		1.00	0.99	0.95	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.98	0.98	0.98	0.98
RCA_cens_90			1.00	0.98	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.98	0.98	0.99
PPML_cens_92.5				1.00	0.95	0.96	0.98	0.96	0.96	0.98	0.96	0.97	0.98	0.97	0.97	0.98
OLS_cens_92.5					1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98
RCA_cens_92.5						1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.98	0.99	0.98
PPML_cens_95							1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.98	0.98	0.99
OLS_cens_95								1.00	1.00	0.99	1.00	0.99	0.99	0.99	0.99	0.98
RCA_cens_95									1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.98
PPML_cens_97.5										1.00	0.99	0.99	1.00	0.98	0.98	1.00
OLS_cens_97.5											1.00	1.00	0.99	0.99	0.99	0.99
RCA_cens_97.5												1.00	0.99	0.99	1.00	0.99
PPML_cens_99													1.00	0.99	0.99	1.00
OLS_cens_99														1.00	0.99	0.99
RCA_cens_99															1.00	0.99
ECI																1.00

This table shows rank correlations between different country rankings. PPML (OLS) refers to the country ranking using PPML (OLS) in the first-step regression. RCA refers to the country ranking using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. 'cens_x' indicates that the normalized exporter-product fixed effects have been winsorized at the x^{th} percentile. All other specifications are as described at the onset of this appendix.

Table S11: Rank Correlations of Country Rankings Across Different Years

	year_2015				year_2016				year_2017				year_2018				year_2019			
	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI	PPML	OLS	RCA	ECI
PPML_2015	1.00	0.99	0.99	0.96	0.99	0.99	0.99	0.96	0.99	0.99	0.98	0.96	0.98	0.98	0.97	0.95	0.99	0.98	0.97	0.95
OLS_2015		1.00	0.99	0.96	0.99	0.99	0.99	0.95	0.99	0.99	0.99	0.96	0.98	0.98	0.98	0.95	0.98	0.98	0.98	0.95
RCA_2015			1.00	0.98	0.99	0.98	0.99	0.96	0.98	0.98	0.99	0.96	0.97	0.97	0.98	0.96	0.98	0.98	0.98	0.96
ECI_2015				1.00	0.95	0.95	0.96	0.98	0.95	0.95	0.96	0.97	0.94	0.94	0.95	0.97	0.95	0.94	0.96	0.97
PPML_2016					1.00	0.99	0.99	0.96	0.99	0.99	0.99	0.96	0.98	0.98	0.97	0.95	0.99	0.98	0.97	0.94
OLS_2016						1.00	0.99	0.95	0.99	0.99	0.99	0.96	0.98	0.98	0.97	0.95	0.99	0.99	0.97	0.94
RCA_2016							1.00	0.97	0.99	0.99	0.99	0.97	0.98	0.97	0.98	0.96	0.98	0.98	0.98	0.96
ECI_2016								1.00	0.96	0.95	0.96	0.99	0.94	0.94	0.95	0.97	0.96	0.95	0.97	0.98
PPML_2017									1.00	1.00	0.99	0.97	0.99	0.98	0.98	0.96	0.99	0.99	0.98	0.95
OLS_2017										1.00	0.99	0.97	0.98	0.99	0.98	0.96	0.99	0.99	0.98	0.95
RCA_2017											1.00	0.98	0.98	0.98	0.98	0.97	0.99	0.98	0.99	0.96
ECI_2017												1.00	0.96	0.95	0.97	0.98	0.97	0.96	0.98	0.98
PPML_2018													1.00	1.00	0.99	0.96	0.99	0.98	0.97	0.94
OLS_2018														1.00	0.99	0.96	0.98	0.98	0.97	0.94
RCA_2018															1.00	0.98	0.98	0.98	0.99	0.96
ECI_2018																1.00	0.96	0.95	0.97	0.98
PPML_2019																	1.00	1.00	0.99	0.96
OLS_2019																		1.00	0.99	0.96
RCA_2019																			1.00	0.98
ECI_2019																				1.00

This table shows rank correlations between different country rankings. PPML (OLS) refers to the country ranking using PPML (OLS) in the first-step regression. RCA refers to the country ranking using Revealed Comparative Advantages. ECI refers to the Economic Complexity Index. 'year_x' indicates that trade data for year x has been used. All other specifications are as described at the onset of this appendix.

S5.2.2 Robustness of Product Ranking

Table S12: Robustness of Product Rankings to Minimum Threshold for Number of Export Destinations by Exporter-Product

	cut_1				cut_3				cut_5			
	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI
PPML_cut_1	1.00	0.92	0.97	0.81	0.95	0.92	0.92	0.85	0.90	0.90	0.86	0.85
OLS_cut_1		1.00	0.91	0.76	0.93	0.97	0.91	0.83	0.89	0.92	0.85	0.84
RCA_cut_1			1.00	0.87	0.93	0.91	0.94	0.91	0.89	0.89	0.87	0.91
PCI_cut_1				1.00	0.74	0.73	0.74	0.96	0.69	0.70	0.64	0.91
PPML_cut_3					1.00	0.97	0.97	0.83	0.98	0.97	0.94	0.86
OLS_cut_3						1.00	0.95	0.82	0.94	0.96	0.90	0.84
RCA_cut_3							1.00	0.84	0.96	0.95	0.97	0.87
PCI_cut_3								1.00	0.79	0.80	0.76	0.97
PPML_cut_5									1.00	0.98	0.97	0.85
OLS_cut_5										1.00	0.95	0.85
RCA_cut_5											1.00	0.83
PCI_cut_5												1.00

This table shows rank correlations between different product rankings. PPML (OLS) refers to the product ranking using PPML (OLS) in the first-step regression. RCA refers to the product ranking using Revealed Comparative Advantages. PCI refers to the Product Complexity Index. 'cut_x' indicates that prior to the first-step regression, all exporter-product observations have been dropped if the product has not been shipped to at least x destinations. All other specifications are as described at the onset of this appendix.

Table S13: Robustness of Product Ranking to Minimum Threshold for Tradeflows at the Bilateral Product Level

	minx_0				minx_1000				minx_10000			
	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI
PPML_minx_0	1.00	0.97	0.97	0.83	1.00	0.97	0.97	0.83	0.97	0.96	0.96	0.83
OLS_minx_0		1.00	0.95	0.82	0.97	1.00	0.95	0.82	0.94	0.96	0.93	0.82
RCA_minx_0			1.00	0.84	0.97	0.95	1.00	0.84	0.95	0.94	0.98	0.84
PCI_minx_0				1.00	0.83	0.82	0.84	1.00	0.85	0.80	0.82	0.99
PPML_minx_1000					1.00	0.97	0.97	0.83	0.98	0.97	0.96	0.83
OLS_minx_1000						1.00	0.95	0.82	0.94	0.96	0.93	0.82
RCA_minx_1000							1.00	0.84	0.95	0.94	0.98	0.84
PCI_minx_1000								1.00	0.85	0.80	0.82	0.99
PPML_minx_10000									1.00	0.98	0.97	0.86
OLS_minx_10000										1.00	0.96	0.81
RCA_minx_10000											1.00	0.83
PCI_minx_10000												1.00

This table shows rank correlations between different product rankings. PPML (OLS) refers to the product ranking using PPML (OLS) in the first-step regression. RCA refers to the product ranking using Revealed Comparative Advantages. PCI refers to the Product Complexity Index. 'minx_x' indicates that prior to the first-step regression, all export values of less than USD x at the bilateral-product level have been dropped. All other specifications are as described at the onset of this appendix.

Table S14: Robustness of Product Ranking to Normalization of Exporter-Product Fixed Effects

	norm_lsum			norm_cossim			norm_vECI			norm_nsqrt			
	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PCI
PPML_norm_lsum	1.00	0.97	0.97	1.00	0.97	0.97	0.96	0.93	0.96	0.93	0.79	0.90	0.83
OLS_norm_lsum		1.00	0.95	0.97	1.00	0.94	0.92	0.93	0.93	0.91	0.84	0.88	0.82
RCA_norm_lsum			1.00	0.98	0.95	1.00	0.91	0.89	0.94	0.91	0.77	0.92	0.84
PPML_norm_cossim				1.00	0.97	0.97	0.95	0.93	0.96	0.93	0.79	0.90	0.83
OLS_norm_cossim					1.00	0.94	0.92	0.92	0.93	0.91	0.84	0.88	0.82
RCA_norm_cossim						1.00	0.90	0.88	0.93	0.90	0.77	0.91	0.83
PPML_norm_vECI							1.00	0.98	0.99	0.86	0.71	0.81	0.78
OLS_norm_vECI								1.00	0.97	0.81	0.69	0.76	0.73
RCA_norm_vECI									1.00	0.87	0.74	0.85	0.83
PPML_norm_nsqrt										1.00	0.91	0.96	0.93
OLS_norm_nsqrt											1.00	0.87	0.89
RCA_norm_nsqrt												1.00	0.94
PCI													1.00

This table shows rank correlations between different product rankings. PPML (OLS) refers to the product ranking using PPML (OLS) in the first-step regression. RCA refers to the product ranking using Revealed Comparative Advantages. PCI refers to the Product Complexity Index. 'norm_x' indicates which normalization has been used. 'norm_lsum' denotes the baseline normalization. All other normalizations start from this baseline. 'norm_cossim' denotes the cosine similarity, that is, the product-product similarity matrix $\hat{\boldsymbol{B}}$ with elements

$$\hat{B}_{ss'} = \frac{\sum_{i\in\mathcal{I}} z_i^s \hat{T}_i^s z_i^{s'} \hat{T}_i^{s'}}{\sqrt{\left[\sum_{i\in\mathcal{I}} z_i^s \hat{T}_i^s z_i^s \hat{T}_i^s\right] \cdot \left[\sum_{i\in\mathcal{I}} z_i^{s'} \hat{T}_i^{s'} z_i^{s'} \hat{T}_i^{s'}\right]}}.$$

In rows and columns 'norm_vECI', each country is normalized by its 'diversity' when computing the product-product similarity matrix, that is matrix $\hat{\boldsymbol{B}}$ has elements

$$\hat{B}_{ss'} = \sum_{i\in\mathcal{I}} \frac{z_i^s \hat{T}_i^s z_i^{s'} \hat{T}_i^{s'}}{\sum_{\hat{s}\in\mathcal{S}} z_i^{\hat{s}} \hat{T}_i^{\hat{s}}}.$$

Finally, in rows and columns 'norm_nsqrt', the exponentiated exporter-product fixed effects are used directly, instead of the square-root thereof. All other specifications are as described at the onset of this appendix.

Table S15: Robustness of Product Ranking to Winsorizing Threshold for Outliers of Exporter-Product Fixed Effects

	cens_90			cens_92.5			cens_95			cens_97.5			cens_99			
	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PPML	OLS	RCA	PCI
PPML_cens_90	1.00	0.98	0.97	1.00	0.97	0.97	0.98	0.95	0.96	0.96	0.92	0.95	0.92	0.86	0.92	0.75
OLS_cens_90		1.00	0.95	0.97	0.99	0.95	0.96	0.97	0.94	0.93	0.93	0.92	0.89	0.87	0.89	0.72
RCA_cens_90			1.00	0.97	0.95	1.00	0.96	0.94	0.99	0.94	0.90	0.97	0.90	0.85	0.94	0.78
PPML_cens_92.5				1.00	0.98	0.97	1.00	0.97	0.97	0.98	0.94	0.96	0.94	0.89	0.94	0.79
OLS_cens_92.5					1.00	0.95	0.97	0.99	0.95	0.95	0.96	0.94	0.92	0.92	0.92	0.77
RCA_cens_92.5						1.00	0.97	0.94	1.00	0.95	0.92	0.98	0.92	0.87	0.95	0.81
PPML_cens_95							1.00	0.97	0.97	0.99	0.95	0.97	0.97	0.92	0.96	0.83
OLS_cens_95								1.00	0.95	0.96	0.99	0.94	0.94	0.96	0.93	0.82
RCA_cens_95									1.00	0.96	0.93	0.99	0.94	0.89	0.97	0.84
PPML_cens_97.5										1.00	0.96	0.97	0.99	0.94	0.97	0.87
OLS_cens_97.5											1.00	0.93	0.95	0.99	0.93	0.87
RCA_cens_97.5												1.00	0.96	0.90	0.99	0.87
PPML_cens_99													1.00	0.95	0.97	0.89
OLS_cens_99														1.00	0.91	0.89
RCA_cens_99															1.00	0.88
PCI																1.00

This table shows rank correlations between different product rankings. PPML (OLS) refers to the product ranking using PPML (OLS) in the first-step regression. RCA refers to the product ranking using Revealed Comparative Advantages. PCI refers to the Product Complexity Index. 'cens_x' indicates that normalized exporter-product fixed effects have been winsorized at the x^{th} percentile. All other specifications are as described at the onset of this appendix.

Table S16: Rank Correlations of Product Ranking Across Different Years

	year_2015				year_2016				year_2017				year_2018				year_2019			
	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI	PPML	OLS	RCA	PCI
PPML_2015	1.00	0.97	0.98	0.83	0.98	0.96	0.97	0.82	0.98	0.95	0.97	0.81	0.97	0.95	0.96	0.81	0.97	0.94	0.96	0.80
OLS_2015		1.00	0.95	0.82	0.96	0.98	0.95	0.81	0.96	0.97	0.94	0.80	0.95	0.97	0.93	0.80	0.94	0.96	0.93	0.79
RCA_2015			1.00	0.84	0.96	0.94	0.99	0.83	0.96	0.93	0.98	0.82	0.96	0.94	0.98	0.82	0.95	0.92	0.97	0.81
PCI_2015				1.00	0.82	0.81	0.83	0.96	0.82	0.82	0.84	0.95	0.82	0.82	0.83	0.95	0.82	0.82	0.83	0.94
PPML_2016					1.00	0.97	0.98	0.83	0.99	0.95	0.97	0.81	0.98	0.95	0.96	0.81	0.97	0.94	0.96	0.80
OLS_2016						1.00	0.95	0.83	0.96	0.98	0.95	0.81	0.95	0.97	0.93	0.81	0.95	0.96	0.93	0.80
RCA_2016							1.00	0.84	0.97	0.94	0.99	0.82	0.96	0.94	0.98	0.82	0.96	0.93	0.98	0.81
PCI_2016								1.00	0.83	0.83	0.84	0.96	0.81	0.82	0.82	0.95	0.82	0.82	0.82	0.94
PPML_2017									1.00	0.97	0.98	0.83	0.98	0.96	0.96	0.82	0.98	0.95	0.96	0.82
OLS_2017										1.00	0.95	0.83	0.95	0.97	0.93	0.82	0.95	0.97	0.93	0.81
RCA_2017											1.00	0.85	0.97	0.94	0.98	0.84	0.96	0.94	0.98	0.83
PCI_2017												1.00	0.81	0.82	0.82	0.96	0.82	0.81	0.83	0.95
PPML_2018													1.00	0.97	0.98	0.83	0.98	0.95	0.97	0.81
OLS_2018														1.00	0.95	0.83	0.96	0.97	0.94	0.81
RCA_2018															1.00	0.84	0.96	0.93	0.99	0.82
PCI_2018																1.00	0.83	0.82	0.83	0.96
PPML_2019																	1.00	0.97	0.98	0.83
OLS_2019																		1.00	0.95	0.82
RCA_2019																			1.00	0.84
PCI_2019																				1.00

This table shows rank correlations between different product rankings. PPML (OLS) refers to the product ranking using PPML (OLS) in the first-step regression. RCA refers to the product ranking using Revealed Comparative Advantages. PCI refers to the Product Complexity Index. 'year_x' indicates that trade data for year x has been used. All other specifications are as described at the onset of this appendix.

References

Balassa, B. (1965). Trade liberalisation and 'revealed' comparative advantage. *The Manchester School*, 33(2):99–123.

Beck, T., Demirgüç-Kunt, A., and Levine, R. (2000). A new database on the structure and development of the financial sector. *The World Bank Economic Review*, 14(3):597–605.

Costinot, A. (2009). On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264.

Davis, D. R. and Dingel, J. I. (2020). The comparative advantage of cities. *Journal of International Economics*, 123:103291.

Gwartney, J. D. and Lawson, R. (2004). *Economic freedom of the world: 2004 annual report*. Fraser Institute.

Hall, R. E. and Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1):83–116.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

Rajan, R. G. and Zingales, L. (1998). Financial dependence and growth. *The American Economic Review*, 88(3):559–586.

World Bank (2021). *World Development Indicators (database)*. https://datatopics.worldbank.org/world-development-indicators/ (retrieved on 2 August 2021).