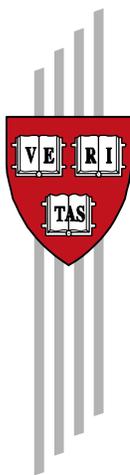


# **A Structural Ranking of Economic Complexity**

Ulrich Schetter

CID Research Fellow and Graduate Student  
Working Paper No. 119  
November 2019

© Copyright 2019 Schetter, Ulrich; and the President and Fellows of  
Harvard College



## **Working Papers**

Center for International Development  
at Harvard University

# A Structural Ranking of Economic Complexity\*

Ulrich Schetter

CID at Harvard University  
Cambridge, MA 02138  
ulrich\_schetter@hks.harvard.edu

This Version: November 2019

## Abstract

We propose a structural alternative to the Economic Complexity Index (ECI, Hidalgo and Hausmann 2009; Hausmann et al. 2011) that ranks countries by their complexity. This ranking is tied to comparative advantages. Hence, it reveals information different from GDP per capita on the deep underlying economic capabilities of countries. Our analysis proceeds in three main steps: (i) We first consider a simplified trade model that is centered on the assumption that countries' global exports are log-supermodular (Costinot, 2009a), and show that a variant of the ECI correctly ranks countries (and products) by their complexity. This model provides a general theoretical framework for ranking nodes of a weighted (bipartite) graph according to some underlying unobservable characteristic. (ii) We then embed a structure of log-supermodular productivities into a multi-product Eaton and Kortum (2002)-model, and show how our main insights from the simplified trade model apply to this richer set-up. (iii) We finally implement our structural ranking of economic complexity. The derived ranking is robust and remarkably similar to the one based on the original ECI.

**Keywords** bipartite graph · economic complexity · international trade · laplacian matrix · log-supermodularity · monotonic eigenvector · ranking

**JEL Classification** F10 · F11 · F14 · O49

---

\*I thank Andres Gomez, Ricardo Hausmann, Filipp Levikov, Marc Muendler, Ralph Ossa, David Torun and seminar participants at the University of St. Gallen and at the Harvard Growth Lab 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.

# 1 Introduction

The Economic Complexity Index (ECI) (Hidalgo and Hausmann, 2009; Hausmann et al., 2011) assesses the economic complexity of countries that is revealed through the products they make. In essence, it considers a country’s economy to be complex if it successfully exports complex products. The ECI has been shown to be a good indicator of both a country’s current economic strength and its future growth prospects. Yet, while the ECI is based on an intuitive narrative, it is less clear how the underlying logic is reflected in the general equilibrium of international trade, and what precisely the ECI measures. In this paper, we start from the motivating rationale of the ECI and assume that comparative advantages are rooted in a complementarity between countries’ economic complexity and products’ complexity following Costinot (2009a).<sup>1</sup> We embed this structure into a multi-product Eaton and Kortum (2002)-model and show that a structural variant of the ECI correctly ranks countries—and products, for that matter—by their economic complexity. This ranking is tied to comparative advantages as opposed to absolute advantages. Hence, it reveals information different from GDP per capita on the deep underlying economic capabilities of countries.

In a free-trade world, a complementarity between country and product characteristics implies that countries’ exports are log-supermodular, i.e. complex economies export relatively more of complex products. In turn, this implies that when equipped with a measure for product complexity, countries’ economic complexities could directly be inferred from the pattern of international specialization—a complex economy being one which concentrates its exports in complex products. We do not have good measures for product complexity, however. The revolutionary insight underlying the ECI is that such measures are, in fact, not needed to learn about countries’ economic complexities. These complexities may instead be inferred from the similarities of countries’ exports. The basic idea being that countries with similar (different) export baskets should have similar (different) levels of economic complexity. We show that with log-supermodular productivities this is indeed the case and how we can exploit the ensuing pattern of countries’ similarities to reveal their ranking of economic complexity. As part of our analysis, we propose a general theoretical framework for ranking nodes in a weighted (bipartite) graph according to some underlying unobservable characteristic.

We begin with a brief discussion of the Economic Complexity Index and its mathematical

---

<sup>1</sup>More generally, we assume that there is a complementarity between some country and some product characteristic, the exact nature of which will not matter for our analysis. For concreteness, we follow Hidalgo and Hausmann (2009); Hausmann et al. (2011) and call the country characteristic ‘economic complexity’ and the product characteristic ‘complexity’.

foundations in Section 2. The ECI was originally introduced as an iterative algorithm that considers an economy as being complex if it successfully exports complex products, where a product is considered complex if it is exported by economically complex countries (Hidalgo and Hausmann, 2009). It turns out that this procedure ranks countries based on the similarities of their export baskets. In fact, it is asymptotically equivalent to first forming a symmetric country-country matrix  $\mathbf{A}$  that indicates for each pair of countries the similarity of their export baskets, and to then ranking countries according to the eigenvector corresponding to the second smallest eigenvalue of

$$\mathbf{L}\mathbf{y} = \lambda\mathbf{D}\mathbf{y}, \quad (1)$$

where  $\lambda$  is an eigenvalue,  $\mathbf{y}$  the corresponding eigenvector,  $\mathbf{D}$  is a diagonal matrix with element  $D_{ii}$  equal to the  $i^{\text{th}}$  row sum of  $\mathbf{A}$ , and  $\mathbf{L} := \mathbf{D} - \mathbf{A}$  is the Laplacian matrix of  $\mathbf{A}$  (Hausmann et al., 2011; Caldarelli et al., 2012; Mealy et al., 2019). Our structural alternative to the ECI ranks countries according to this same eigenvector, but based on a structurally estimated matrix  $\mathbf{A}$  as opposed to an ad-hoc matrix based on Revealed Comparative Advantages (Balassa, 1965).

Section 3 presents the main theoretical result of our paper. In this section, we consider a stylized trade model that is centered on the assumption that countries' global exports at the product level,  $X_i^s$ , are log-supermodular in countries' economic complexity  $i$  and products' complexity  $s$ . That is, for every pair of countries  $i' > i$  and products,  $s' > s$ , we have

$$\frac{X_{i'}^{s'}}{X_{i'}^s} > \frac{X_i^{s'}}{X_i^s}. \quad (2)$$

Condition (2) implies that complex countries export relatively more of complex products, in line with the guiding rationale of the ECI. Because this is true for all countries, it implies in turn that the export baskets of complex countries are relatively more similar to the export baskets of other complex countries than to the export baskets of less complex countries and vice versa. Formally, we show in Lemma 1 that the country-country similarity matrix  $\mathbf{A}$  with elements

$$A_{i'i} := \frac{1}{S} \sum_{\hat{s} \in \mathcal{S}} X_{i'}^{\hat{s}} \cdot X_i^{\hat{s}}$$

inherits the log-supermodularity of the  $X_i^s$ , that is, for every quadruple of countries  $i' > i$  and  $k' > k$  it holds

$$\frac{A_{i'k'}}{A_{i'k}} > \frac{A_{ik'}}{A_{ik}}. \quad (3)$$

The key point is that this log-supermodularity imposes sufficient structure on country similarities to imply that the second eigenvector of (1) correctly ranks countries by their underlying

economic complexity. Precisely, we show in Theorem 1 that for every positive and symmetric matrix  $\mathbf{A}$  satisfying Condition (3), the second eigenvector of (1) is strictly monotonic. We provide Monte Carlo Simulations adding random noise to such matrices and show that this monotonicity is very robust as long as the size of the matrix is not too small relative to the size of the random shock. In other words, the second eigenvector correctly ranks countries by their complexity even if the empirically derived matrix  $\mathbf{A}$  is not everywhere log-supermodular—and in fact even if locally it satisfies Condition (3) only marginally more often 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.

In the remainder of the paper, we use our insights from Section 3 to develop a structural alternative to the ECI based on a workhorse trade model. Importantly, however, Theorem 1 not only allows to rank countries by their unobservable economic complexity, but our work provides a general theoretical framework for ranking nodes in a weighted unipartite graph—or, when combined with Lemma 1, a weighted bipartite graph. To illustrate this point, we briefly discuss how our insights can readily be applied to rank academic journals by their prestige, or politicians on a left-to-right scale, for example, at the end of Section 3.

In Section 4, we outline the economic model underlying our structural ranking of economic complexity and characterize equilibrium trade flows. We consider a multi-product (or industry) Eaton and Kortum (2002) model, where countries differ in their economic complexity  $i$  and products differ in their complexity  $s$ .<sup>2</sup> The exact nature of these country and product characteristics is not of importance. The key point is that we follow Costinot (2009a) and Costinot and Vogel (2015) in assuming that the country-product specific fundamental productivity  $\tilde{T}_i^s$  is log-supermodular. To accommodate additional sources of comparative advantages at the product level, we augment this fundamental productivity by an idiosyncratic component. In other words, the exporter-product specific location parameter of the Fréchet distribution is given by  $T_i^s := \tilde{T}_i^s \cdot \epsilon_i^s$ . We further allow for zero trade flows at the exporter-product level, assuming that they are governed by the same complementarity between country and product complexity as the fundamental productivities. That is, we assume that economically complex countries are relatively (in a ‘diff-in-diff’ sense) more likely to be exporting the complex products and, if they do, they tend to have a relatively higher

---

<sup>2</sup>Throughout, we follow the nomenclature in Hidalgo and Hausmann (2009) and speak of products, which are available in many different varieties. This is also consistent with the fact that we later on consider trade at the 4-digit HS-level. In terms of our modeling choices, however, these products correspond to what is typically referred to as sectors or industries in the international trade literature.

productivity in these products.

We discuss how we can rank countries by their economic complexity in Section 5. In a world as described by our trade model, this can be achieved by applying Theorem 1 to a similarity matrix  $\mathbf{A}$  with elements

$$A_{i'i} := \mathbb{E} \left[ \frac{1}{S} \sum_{\hat{s} \in \mathcal{S}} z_i^{\hat{s}} T_{i'}^{\hat{s}} \cdot z_i^{\hat{s}} T_i^{\hat{s}} \right] \quad (4)$$

where  $z_i^s$  is a binary random variable that indicates whether country  $i$  is making product  $s$  or not. Interestingly, the same need not be true for the ECI, which is based on a binary country-product matrix that indicates for each country the set of products for which it has a Revealed Comparative Advantage (RCA) of at least 1 according to the Balassa (1965) measure.

While we cannot observe matrix  $\mathbf{A}$  as defined in (4) from the data, we discuss how we can estimate it in Section 6. In particular, in a first step we can estimate the country-product specific productivities  $T_i^s$  up to a normalization for each country and product from a fixed effects regression of bilateral trade flows (Costinot et al., 2012). We estimate these fixed effects using both OLS and PPML, respectively. In a second step, we use the estimated  $\hat{T}_i^s$  to form the sample analogue of matrix  $\mathbf{A}$ . To rank countries, we finally compute the eigenvector corresponding to the second smallest eigenvalue of (1). Our OLS estimator ranks Japan, South Korea, and Switzerland at the top, and Yemen, Sudan, and Malawi at the bottom of a list of 127 countries included in our sample. This ranking is remarkably robust: The rank correlation with the one derived from using PPML in the first step is larger than .995, and even with the original ECI that starts from a binary country-product matrix indicating country-product pairs with RCA of at least one it has a rank correlation of .96. Hence, our work suggests that while theoretically the original ECI may fail to correctly rank countries in a world with trade frictions, this may be less of a concern in practice. It may therefore also help explaining the astounding success of the ECI in measuring economic strength and future growth potential. Importantly, this ranking of countries by their economic complexity is fundamentally different from a ranking by their GDP per capita.<sup>3</sup> The reason is simple: our notion of economic complexity is tied to comparative advantages as opposed to absolute advantages. Hence, the structural variant of the ECI proposed here may reveal important and novel information on the deep underlying economic capabilities of countries.

Analogous to the original Economic Complexity Index, the exact same reasoning used to rank

---

<sup>3</sup>One way of seeing this is by noting that the normalized exporter-product fixed effects do not capture GDP per capita (the wage).

countries by their economic complexity also allows to rank products by their complexity. We discuss this and present rankings at the 2-digit HS classification level in Section 7. The product ranking is somewhat less robust, which may not come as a surprise given that we use export data from 127 countries to evaluate the similarities of 97 products. But yet, this ranking may serve as an alternative to proxies typically used in the literature.<sup>4</sup>

Our paper contributes to several strands of literature. We build on the works by Hidalgo and Hausmann (2009), Hausmann et al. (2011), and Mealy et al. (2019) on the one hand, and by Eaton and Kortum (2002), Costinot (2009a), and Costinot et al. (2012) on the other, and propose a structural ranking of countries by their economic complexity. While nothing in particular hinges on the interpretation of our country characteristic as ‘economic complexity’, our ranking is based on international trade data and, hence, our work contributes to the literature measuring the ‘economic complexity’ of countries based on trade data (Hausmann et al., 2007; Hidalgo and Hausmann, 2009; Hausmann et al., 2011; Tacchella et al., 2012; Morrison et al., 2017; Albeaik et al., 2017; Servedio et al., 2018). To the best of our knowledge, this paper is the first to start from a theoretical model of how ‘economic complexity’—or, more generally, countries’ economic strength—is reflected in international trade flows, and to then show that and how the ranking of economic complexity can be uncovered from the data.

Our ranking is closely related to the Economic Complexity Index (Hidalgo and Hausmann, 2009; Hausmann et al., 2011): It differs in that we start from a structural country-country similarity matrix. It is then, however, based on the exact same generalized eigenproblem of the respective matrix. The same is true for the product rankings. Moreover, in spite of the substantial differences in the way the similarity matrices are constructed, the derived rankings are highly correlated. Hence, our work lends support to applications of the Economic Complexity Index in empirical studies (e.g. Hausmann et al. 2011; Poncet and Starosta de Waldemar 2013; Hartmann et al. 2017; Petralia et al. 2017; Javorcik et al. 2018) and in numerous policy reports, and it may guide the way for more structural applications of these concepts in future. It further provides an alternative to proxies for product complexity used in the literature (e.g. Levchenko 2007; Costinot 2009b; Schetter 2019).

More generally, our ranking may be seen as a ranking of countries according to their deep underlying capabilities, technologies, and know-how that allow them to be competitive in

---

<sup>4</sup>According to our structural ranking using OLS in the first step, the three most complex products are: ‘Nuclear reactors, boilers, machinery and mechanical appliances’ (84), ‘Electrical machinery and equipment and parts thereof’ (85), and ‘Photographic or cinematographic goods’ (37). The three least complex products are: ‘Ores, slag, and ash’ (26), ‘Oil seeds and oleaginous fruits’ (12), and ‘Coffee, tea, mate and spices’ (09).

complex products. This ranking is conceptually very different from e.g. the Global Competitiveness Index (GCI, Sala-i Martin and Artadi 2004): While the GCI assesses competitiveness based on a multitude of observable determinants, we follow Hidalgo and Hausmann (2009) and measure the competitiveness that is revealed through what countries actually do. Our ranking is also conceptually different from a ranking of countries based on their GDP per capita, and our work may thus provide a novel perspective on economic development, allowing to separate growth in income from advances in the deep underlying productive capabilities of an economy.

To derive our structural ranking, we follow Costinot et al. (2012); Hanson et al. (2015); Levchenko and Zhang (2016) and consider a multi-product (sector) Eaton and Kortum (2002)-model, which allows extracting productivities at the country-product level from a fixed effects gravity regression.<sup>5</sup> As opposed to these papers, however, we do not use the estimated productivities to learn about the importance of Ricardian comparative advantage for trade and welfare, or to study time trends in comparative advantage. Rather, we show that these estimated productivities can be used to learn about the deep underlying economic complexity of countries and products, respectively, that drive comparative advantage at the country-product level.<sup>6</sup>

To derive our main theoretical result, we consider a simplified trade model first. Our analysis of this model provides a general theoretical framework for ranking nodes in a weighted (bipartite) graph. A large literature ranks nodes according to their importance for the network—their centrality (e.g. Katz 1953; Freeman 1977; Bonacich 1987; Brin and Page 1998; Kitsak et al. 2010).<sup>7</sup> In the economics literature, centrality-based rankings have been proposed to identify individuals that are important for fast diffusion of innovation (Banerjee et al., 2013), to design policies for conflict resolution (König et al., 2017), building state capability (Acemoglu et al., 2015), and fostering innovation (König et al., 2018), for example, and, more

---

<sup>5</sup>The fixed effects regression is consistent with alternative foundations for the gravity equation based on e.g. Armington (1969), Krugman (1980), Melitz (2003) (see Head and Mayer (2014)). We think of countries' economic complexity and products' complexity as being reflected in productivities, and we therefore follow the above papers in interpreting these fixed effects through the lens of an Eaton and Kortum (2002)-model.

<sup>6</sup>Hence, our paper also differs from previous work that tests for a complementarity between a country and a product characteristic using proxies for these characteristics (e.g. Levchenko 2007; Nunn 2007; Cuñat and Melitz 2012). Closer to our work is Costinot (2009b) who uses a proxy for product complexity to construct a measure of 'revealed institutional quality' of countries, assuming that there is a complementarity between the two. While in principle we could follow a similar approach here, it would imply that the quality of the derived country ranking hinges on the quality of the product proxy used. We therefore follow a different approach and show how we can exploit the assumed log-supermodularity to reveal the underlying ranking of economic complexity without relying on an ad-hoc proxy for product complexity.

<sup>7</sup>See Jackson (2008) and Liao et al. (2017) for overviews of these measures, and Bloch et al. (2019) for an axiomatic foundation of some of these measures.

generally, to identify ‘key players’ in a network (Ballester et al., 2006). Our focus is different: We assume that nodes—countries in our case—differ in some unobservable characteristic—their economic complexity—and then seek to rank them according to this characteristic.<sup>8</sup> This ranking is based on the similarities of nodes to each other, which can mathematically be described as a graph or network, but we are not interested in the importance of individual nodes for the network or even the network as such.<sup>9</sup>

Finally, our work is related to spectral graph theory (Chung, 1997). More to the point, the eigenvector that we use to rank countries—and nodes in a weighted graph more generally—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). We show that this is actually true in a global sense if  $\mathbf{A}$  is log-supermodular.

## 2 Mathematical Foundations of the Economic Complexity Index

In this section, we briefly review the Economic Complexity Index (ECI) and the underlying mathematical algorithm. We will highlight that the ECI is, in fact, equivalent to a generalized eigenvector of a country-country matrix that summarizes the similarity of their export baskets. We will study this eigenvector in the next section and later on use it to develop our structural variant of the ECI.

The Economic Complexity Index is a measure of countries’ economic strength (and products’ complexity) based on export data (Hidalgo and Hausmann, 2009; Hausmann et al., 2011). Its motivation is as intuitive as it is compelling: If we observe that a given product is produced in a country, this reveals that the country has the capability to provide all necessary inputs for production and to use them competitively. Hence, the set of products that a country makes is informative about its capabilities. Analogously, the set of countries that successfully export a given product is informative about its production requirements. Guided by this logic, Hidalgo and Hausmann (2009) suggest that a complex country is one that exports complex products

---

<sup>8</sup>The key point is that this country characteristic is unobservable. In that sense, our work also differs from e.g. Perry and Reny (2016), who propose an axiomatic approach to ranking scientists based on their observable publications and citations.

<sup>9</sup>One way of seeing that our ranking is not concerned with a country’s centrality in the network is by noting that in a simple Ricardian model of international trade, log-supermodularity of productivities—our main assumption underlying our structural alternative to the ECI—gives rise to a ‘ladder’ of international specialization (Costinot, 2009a).

and vice versa. To measure these complexities, they propose an iterative algorithm that is based on a binary country-product matrix that indicates for each country the set of products of which it is a significant exporter. They consider a country to be a significant exporter of a product if it has a Revealed Comparative Advantage (RCA) of at least one according to the Balassa (1965) measure.

It turns out that asymptotically this iterative procedure ranks countries by an eigenvector of a country-country similarity matrix, and this eigenvector is, in fact, used for the Economic Complexity Index (ECI) (Hausmann et al., 2011). In particular, let  $\mathbf{M}$  denote the  $I \times S$  binary country-product matrix with entry  $M_{is} = 1$  if country  $i$  has an RCA of at least 1 in product  $s$  and  $M_{is} = 0$  otherwise. Further, let  $\mathbf{U}$  be the  $S \times S$  diagonal matrix with entry  $U_{ss}$  equal to the ubiquity of product  $s$ , i.e.  $U_{ss}$  is the sum of the  $s$ th column of matrix  $\mathbf{M}$ . We can use these matrices to generate a positive and symmetric country-country similarity matrix

$$\mathbf{A} := \mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T,$$

where here and below we use a superscript  $T$  to denote the transpose of a matrix. Matrix  $\mathbf{A}$  specifies for each pair of countries  $i, i'$ , the number of products that they have in common, with each product weighted by the inverse of its ubiquity. The ECI is the eigenvector corresponding to the second smallest eigenvalue of the generalized eigenproblem (Hausmann et al., 2011; Mealy et al., 2019)

$$\mathbf{L}\mathbf{y} = \lambda\mathbf{D}\mathbf{y}, \tag{5}$$

where  $\mathbf{D}$  is the diagonal matrix with diagonal entries equal to the respective row sum of  $\mathbf{A}$ , and  $\mathbf{L} := \mathbf{D} - \mathbf{A}$  is the Laplacian matrix of  $\mathbf{A}$ .<sup>10,11</sup> This eigenvector—which we henceforth simply refer to as the second eigenvector of (5)—solves the following minimization problem (e.g. Chung 1997; Shi and Malik 2000; Belkin and Niyogi 2003):

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

---

<sup>10</sup>This generalized eigenvector is equivalent to the eigenvector corresponding to the second largest eigenvalue of matrix  $\mathbf{D}^{-1}\mathbf{A}$ , where  $\mathbf{D}$  is the same matrix as in (5), i.e. it is a diagonal matrix with countries' diversities on the diagonal (Mealy et al., 2019). Hausmann et al. (2011) use this representation to define the ECI. Our subsequent work will build on the generalized eigenproblem and, hence, we consider this representation instead.

<sup>11</sup>The iterative algorithm proposed in Hidalgo and Hausmann (2009) actually converges to the first eigenvector, which is a vector of ones. Hidalgo and Hausmann (2009) stop after  $N$  iterations and rescale the derived vector to have standard deviation of 1. This rescaled vector converges to the second eigenvector (see Caldarelli et al. (2012) for a discussion).

where here and below we use  $\mathbf{1}$  to denote a vector of ones.

Hence, the Economic Complexity Index ultimately ranks countries by reducing the rich structure of similarities of countries’ export baskets as summarized in  $\mathbf{A}$  to a single dimension. It is therefore not obvious to what extent the ECI is informative about the deep underlying economic capabilities of countries, and the associated literature so far lacks a thorough understanding of (i) whether such information is entailed in the similarities of countries’ exports and (ii) if so, whether the second eigenvector of (5) can reveal this information. In the remainder of the paper, we show that the answer to both questions is yes if the guiding rationale of the ECI is correct, i.e. if it is indeed the case that economically complex countries tend to export complex products. In particular, we show that a variant of the ECI correctly ranks countries by their economic complexity if we assume that the fundamental productivity of a country in a product is log-supermodular such that—on balance—economically complex countries are relatively more productive in complex products. Heuristically, note that the objective in (6) can be rewritten as

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

which suggests that the second eigenvector of (5) tends to assign similar values  $y_i$  and  $y_j$  to similar countries, i.e. to pairs of countries with large values  $A_{ij}$ .<sup>12</sup> Indeed, this eigenvector has previously been proposed as a dimensionality reduction algorithm that ‘optimally preserves local neighborhood information’ (Belkin and Niyogi, 2003). We show in the next section that in a ‘log-supermodular world’ this is actually true globally, that is, the second eigenvector ranks countries in accordance with the deep underlying economic complexity that drives their similarity.

### 3 A General Theoretical Framework for Ranking Nodes in a Weighted (Bipartite) Graph

In this section, we present a general theoretical framework for ranking nodes in a weighted (bipartite) graph according to some underlying unobservable characteristic. Our main focus is on developing a structural alternative to the ECI. We will therefore introduce this general framework by means of a stylized version of our economic model from the next section.

---

<sup>12</sup>The two constraints in minimization problem (6) essentially rule out trivial solutions: The first constraint rules out solutions where all values of  $y$  are zero or arbitrarily close to zero, while the second constraint rules out solutions where the entries in  $y$  are different from zero but all the same.

This allows introducing in a transparent way the connection between our main economic assumption and the ability of the second eigenvector of (5) to correctly rank countries by their complexity. Importantly, however, our main insights from this section not only apply to ranking countries and products, but, in fact, more generally to ranking nodes in a weighted (bipartite) graph. We will revert to this point at the end of Section 3.2, and begin with introducing some definitions that will prove useful in our subsequent discussions.

### 3.1 Definitions

Our theory will be centered on strictly log-supermodular matrices, which we define as follows:

**Definition 1 (Log-supermodular matrix)**

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

$$\frac{M_{r'c'}}{M_{r'c}} > \frac{M_{rc'}}{M_{rc}}. \quad (8)$$

Definition 1 may most easily be understood by means of a simple example. In particular, it states that a matrix  $\mathbf{M}$  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  $\mathbf{M}$  are log-supermodular, where, recall, a block of matrix  $\mathbf{M}$  is defined as follows:

**Definition 2 (Block of matrix)**

A block of matrix  $\mathbf{M}$  is a submatrix formed by the elements in the intersection of contiguous rows and columns of  $\mathbf{M}$ .

In the next section, we will show that for every log-supermodular matrix the solution to problem (6) is monotonic—i.e. the eigenvector corresponding to the second smallest eigenvalue of (5) is monotonic, where we define a monotonic vector as follows:

**Definition 3 (Monotonic vector)**

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

With these definitions at hands, we now turn to the main theoretical result of our paper.

**3.2 Theory**

Consider a world with  $I$  countries and  $S$  products (or industries). Suppose that countries differ by some characteristic, which we call a country's economic complexity, for concreteness. Similarly, suppose that products differ by some characteristic, which we call their complexity. To simplify notation, we will henceforth identify countries by their rank of economic complexity  $i \in \{1, 2, \dots, I\}$  and products by their rank of complexity  $s \in \{1, 2, \dots, S\}$  from the lowest to the highest, i.e. for any pair of countries  $i' > i$ , country  $i'$  is more complex than country  $i$ , and analogously for products. Importantly, however, we think of these characteristics and the implied rankings as being unobservable. In fact, our goal is precisely to uncover this ranking from the data.

The exact interpretation of  $i$  and  $s$  is not important. The key point is that we follow Costinot (2009a) and assume that there is a complementarity between  $i$  and  $s$  such that a high- $i$  country has a comparative advantage in a high- $s$  product. We will embed this structure in a multi-sector Eaton and Kortum (2002) model in the next section, and show how we can exploit the ensuing equilibrium trade flows to correctly rank countries and products. For now, we simply assume that comparative advantages are one-for-one reflected in countries' aggregate sales of a product, which we denote by  $X_i^s > 0$ . Precisely, we assume that countries' exports are strictly log-supermodular:

**Assumption 1**

Let  $\mathbf{X}$  be the  $I \times S$  positive matrix with element  $X_{is} = X_i^s$  equal to the global sales of country  $i$  and product  $s$ . Matrix  $\mathbf{X}$  is strictly log-supermodular.

In essence, Assumption 1 implies that high- $i$  countries have relatively higher exports in high- $s$  products. Because this holds true for all countries, it implies in turn that the export basket of an economically complex country is relatively more similar to the export baskets of other complex countries than to the export baskets of less complex countries and vice versa. In particular, let  $\mathbf{A}$  be the positive and symmetric  $I \times I$  country-country similarity matrix with

element

$$A_{ii'} := \frac{1}{S} \sum_{s \in \mathcal{S}} X_i^s \cdot X_{i'}^s. \quad (9)$$

As we show in the following lemma, this matrix inherits the log-supermodularity of matrix  $\mathbf{X}$ :

**Lemma 1**

*Matrix  $\mathbf{A}$  as defined in (9) is strictly log-supermodular.*

The proof of Lemma 1 is given in Appendix A.1. In words, Lemma 1 compares the export baskets of two countries  $i' > i$  based on how similar they are to the export baskets of countries  $k' > k$ . In essence, Lemma 1 implies that—when compared to a less complex country  $i$ —the export basket of country  $i'$  is systematically more similar to the export baskets of other complex countries ( $k'$ ) than to the ones of less complex countries ( $k$ ). As we show in the following theorem, this imposes sufficient structure such that the second eigenvector of (5) correctly ranks countries by their economic complexity when applied to matrix  $\mathbf{A}$ .

**Theorem 1**

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

$$\mathbf{L}\mathbf{y} = \lambda\mathbf{D}\mathbf{y} \quad (10)$$

*is strictly monotonic.*

The proof of Theorem 1 is given in Appendix A.2. Theorem 1 is the main theoretical result of our paper. It provides a general theoretical framework for ranking nodes of a weighted unipartite graph—or, when combined with Lemma 1, nodes in a weighted bipartite graph—according to some unobservable underlying characteristic.

To simplify the exposition, we have assumed that rows in matrix  $\mathbf{X}$  and, hence, rows and columns in matrix  $\mathbf{A}$ , are already ordered according to the underlying economic complexity. It is in such case that matrix  $\mathbf{A}$  is indeed log-supermodular. This is, of course, not true in general. In fact, it is precisely this order that we would like to uncover from the data, and what we might expect to have is a matrix  $\mathbf{A}$  that can be made log-supermodular by appropriate permutations of rows and columns. Note that any such permutation results in the exact same permutation of the elements of the second eigenvector of (10). Theorem 1

therefore tells us that we can uncover the ranking of economic complexity by rearranging rows and columns of  $\mathbf{A}$  such that the second eigenvector of (10) is indeed monotonic.

Four remarks are in order: First, while the underlying bipartite graph provides a compelling foundation for why matrix  $\mathbf{A}$  is log-supermodular, it is worth noting that Theorem 1 does not hinge on this foundation: We can readily apply Theorem 1 to any unipartite graph whose adjacency matrix is log-supermodular.

Second, of course the eigenvector allows to correctly rank countries up to sign only. Heuristically, this is the case because we rank countries based on the similarity of their export baskets, and this similarity does not embed information on the sign of this ranking. Mathematically, this is, in fact, inherent to using an eigenvector for this ranking, and is also the case for the Economic Complexity Index (Hidalgo and Hausmann, 2009; Hausmann et al., 2011). In practice, it implies that we need some additional information to determine the sign of the ranking, i.e. to determine which countries should be ranked on top. Importantly, this is probably more of a theoretical concern, rather than a real issue in practical applications, where the underlying theory readily lends itself to a strong prior regarding the direction of the ranking. In case of our complexity ranking, for example, we can determine the direction of the ranking by requiring that industrialized countries be ranked high.

Third, it is important to note that while our main focus is on developing a structural alternative to the Economic Complexity Index and while we therefore think of  $\mathbf{A}$  as a country-country similarity matrix here, nothing in particular hinges on this interpretation. In fact, Theorem 1 is a general result for ranking nodes in a graph and—when combined with Lemma 1—nodes in one part of a bipartite graph. In particular, we can think of the  $X_i^s$  as the elements of an  $I \times S$  adjacency matrix  $\mathbf{X}$  of a bipartite graph. Lemma 1 and Theorem 1 apply to any such graph as long as the elements of the adjacency matrix satisfy Assumption 1. For instance, we may assume that talented scientists are systematically more successful at publishing in prestigious journals, or that ‘left-wing’ politicians have a systematically higher probability of accepting ‘left-wing’ policies. If so, our work can readily be applied to rank politicians on a ‘left-to-right’ scale or academic journals according to their prestige.

Finally, in practical applications it is unlikely that matrix  $\mathbf{A}$  is indeed perfectly log-supermodular. Hence, an ensuing question is whether the result in Theorem 1 is robust to deviations from the perfectly log-supermodular structure of matrix  $\mathbf{A}$ , i.e. whether it holds up in situations where Condition (8) is not satisfied for all quadruples of elements in the intersections of pairs of rows and columns. We turn to this issue next.

### 3.3 Robustness

To evaluate the robustness of our result in Theorem 1 to variations of matrix  $\mathbf{A}$ , we perform a simple Monte Carlo study that involves  $80k$  randomly drawn  $100 \times 100$  matrices— $10k$  for each column in Table 1. To simulate matrices  $\mathbf{A}$ , we start from a randomly drawn symmetric matrix  $\tilde{\mathbf{A}}$  that is supermodular, i.e. it is log-supermodular when exponentiating it element-wise. Details on how we generate this matrix are provided in Appendix B. For our purposes here it suffices to note that all 2 by 2 blocks 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  $\tilde{\mathbf{A}}$  we have

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

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

In a second step, we add to matrix  $\tilde{\mathbf{A}}$  another symmetric random matrix, whose elements 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. Note that in the rightmost columns these shocks are large when compared to the margin with which 2 by 2 blocks are log-supermodular. Indeed, as we discuss in a second, these blocks are log-supermodular only marginally more often than expected for an iid random matrix.

Finally, we exponentiate the matrix element-wise to get our simulated matrix  $\mathbf{A}$ . Further details are provided in Appendix B.

For each random matrix  $\mathbf{A}$ , we then compute three statistics measuring how successful the second eigenvector of (10) is in ranking rows and columns of that matrix, and average these statistics over the  $10k$  random matrices in the respective column of Table 1. 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’). We finally present 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’). We summarize our findings in Table 1.

The first column in Table 1 shows our benchmark with no random shocks. By construction, all 2 by 2 blocks are log-supermodular and, hence, all random matrices  $\mathbf{A}$  in this column are log-supermodular. As predicted by Theorem 1, the second eigenvector of (10) always ranks

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 our  $80k$  randomly generated  $100 \times 100$  matrices— $10k$  for each column. Random matrices have been generated as described in the main text and further detailed in Appendix B. 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  $[1, 2, \dots, 100]$ —, where 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  $\mathbf{A}$  that are log-supermodular.

all rows (and columns) correctly.<sup>13</sup>

In the remaining columns of Table 1 we 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  $\mathbf{A}$  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  $\mathbf{A}$  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 very high. How is that possible? To see this, note that ‘Avg share LSM’ limits attention to 2 by 2 blocks of matrix  $\mathbf{A}$ . 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

<sup>13</sup>While this is not the main focus here, a perhaps interesting insight that also emerges from these simulations is that supermodularity, as opposed to log-supermodularity, does not impose sufficient structure for the second eigenvector to correctly rank rows and columns, even in the absence of random shocks.

elements in rows and columns  $i, i + 30$  and  $k, k + 30$  in more than 90% of the cases, even with random shocks 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. Indeed, as we show in Appendix B, smaller matrices, i.e. matrices with less information at greater distances to exploit, are less robust to adding noise.<sup>14</sup>

In summary, our simulations suggest that our ranking of rows and columns in positive and symmetric log-supermodular matrices is very robust to adding random noise. We provide further details and additional results in Appendix B.

## 4 Economic Model

In the previous section, we have outlined a general theoretical framework for ranking nodes in a weighted (bipartite) graph. In the remainder of the paper, we apply these insights and develop a structural alternative to the ECI. We begin with outlining the economic model.

We consider a multi-product (or industry) Eaton and Kortum (2002) model following Costinot et al. (2012).<sup>15</sup> There are  $I$  countries, indexed by  $i, j \in \mathcal{I}$ , and  $S$  products, indexed by  $s \in \mathcal{S}$ . To simplify notation, we assume that countries are ranked by their economic complexity from the least to the most complex economy such that for every  $i, i' \in \mathcal{I} : i < i'$  we have that country  $i'$  is economically more complex than country  $i$ . Similarly, products are ranked by their complexities from the least to the most complex such that for every  $s, s' \in \mathcal{S} : s < s'$  we have that product  $s'$  is more complex than product  $s$ . Importantly, however, we think of these characteristics as being unobservable and, in fact, we are ultimately interested in finding a way of ranking countries—and products, for that matter—according to their complexities.

We follow Costinot (2009a) and Costinot and Vogel (2015) in assuming that the country-product specific fundamental productivities  $\tilde{T}_i^s$  are log-supermodular. We augment these fundamental productivities by idiosyncratic productivity components at the exporter-product level,  $T_i^s = \tilde{T}_i^s \epsilon_i^s$ , and allow for zeros at the exporter-product level as will be detailed below. 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  $j$ . As standard in

---

<sup>14</sup>An interesting insight that emerges from these considerations is that, indeed, for a (noisy) log-supermodular matrix, the eigenvector corresponding to the second smallest eigenvalue of (10) preserves global and not just local neighborhood information, as suggested in Belkin and Niyogi (2003).

<sup>15</sup>The literature on international trade typically refers to the upper-tier level of goods-differentiation as industries (or sectors) and the lower-tier level as varieties within a given industry. To be consistent with the nomenclature chosen in Hidalgo and Hausmann (2009); Hausmann et al. (2011), we refer to the upper-tier level as products and the lower-tier level as varieties of a given product.

the trade literature, we normalize  $d_{ii} = 1$  and assume that trade costs satisfy the triangular inequality. There is perfect competition in all markets.

## 4.1 Households

Country  $i$  is populated by  $L_i$  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  $\alpha^s$ ,  $\sum_{s \in \mathcal{S}} \alpha^s = 1$ , and the lower-tier being CES over a continuum of measure one of varieties within products with elasticity of substitution  $\sigma$ . Accordingly, the total expenditure in country  $i$  on variety  $\omega$  of product  $s$  is

$$x_i^s(\omega) = \left[ \frac{p_i^s(\omega)}{P_i^s} \right]^{1-\sigma} \alpha^s w_i L_i,$$

where  $w_i$  is the wage rate and, hence,  $w_i L_i$  total income in country  $i$ , and where

$$P_i^s := \left[ \int_0^1 p_i^s(\omega)^{1-\sigma} d\omega \right]^{\frac{1}{1-\sigma}}$$

is the CES-price index for product  $s$  in country  $i$ .

## 4.2 Production

Production is constant returns to scale using labor as the only input. There is a continuum of varieties  $\omega \in [0, 1]$  of each product  $s$ . We use  $\varphi_i^s(\omega)$  to denote the constant productivity of producing variety  $\omega$  of product  $s$  in country  $i$ , and assume that it is drawn independently for each triplet  $(i, s, \omega)$  from a Fréchet distribution with dispersion parameter  $\theta > 0$  and country-product specific location parameter  $T_i^s > 0$

$$F_i^s(\varphi) = \exp[-\varphi^{-\theta} T_i^s], \quad \forall \varphi > 0.$$

The location parameter has two components: Country  $i$ 's fundamental productivity in product  $s$ ,  $\tilde{T}_i^s > 0$ , and an idiosyncratic productivity component,  $\epsilon_i^s$ , that is independently distributed across countries and products with strictly positive support and mean one

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

Here and below we use  $\mathbb{E}[\cdot]$  to denote the expectation operator. The fundamental productivities  $\tilde{T}_i^s$  capture comparative advantages arising from the systematic relationship between a

country’s economic complexity and the complexity of a product, while the idiosyncratic component captures all other sources of comparative advantage at the product level.<sup>16</sup> Following Costinot (2009a) and Costinot and Vogel (2015), we assume that the fundamental productivity is log-supermodular in a country’s economic complexity and a product’s complexity, that is:

**Assumption 2**

$$\frac{\tilde{T}_{i'}^{s'}}{\tilde{T}_{i'}^s} > \frac{\tilde{T}_i^{s'}}{\tilde{T}_i^s}, \quad \forall i' > i \in \mathcal{I}, s' > s \in \mathcal{S}$$

In words: the above condition implies that there is a complementarity between countries’ economic complexity and products’ complexity such that—on balance—a more complex economy is relatively more productive in the complex products. This fundamental pattern of comparative advantages will be reflected in trade flows and will eventually allow identifying countries’ economic complexity from trade data.

Zeros are prevalent in international trade at the country-product level.<sup>17</sup> To accommodate these, we assume that country  $i$  is active in product  $s$  with probability  $\rho_i^s > 0$ . These probabilities are guided by the same complementarity that governs the fundamental productivities, that is, an economically more complex country is relatively more likely to be active in the complex products:

**Assumption 3**

$$\frac{\rho_{i'}^{s'}}{\rho_{i'}^s} > \frac{\rho_i^{s'}}{\rho_i^s}, \quad \forall i' > i \in \mathcal{I}, s' > s \in \mathcal{S}$$

We may think of  $\rho_i^s$  as e.g. the probability that country  $i$  has acquired the product-specific know-how or technologies needed to make product  $s$ . In what follows, it will come in handy to introduce a binary random variable  $z_i^s$  that takes on value of one with probability  $\rho_i^s$  and zero otherwise, and that indicates whether country  $i$  is active in product  $s$ . We assume that the realization of  $z_i^s$  is independent across  $i$  and  $s$ .

In addition to zeros at the country-product level, there are zeros at the bilateral product level, i.e. countries export a product to a subset of destinations only. We will discuss these

---

<sup>16</sup>Of course, in a multi-product Eaton and Kortum (2002)-model, there are also comparative advantages within products at the variety-level. See Costinot et al. (2012) for a discussion.

<sup>17</sup>In our estimations below, we consider exports at the 4-digit HS classification level. Our cleaned dataset has  $\sim 44\%$  zeros at the exporter-product level.

further in Section 6 below. For now it suffices to denote by  $\mathcal{I}_j^s$  the set of countries that have strictly positive exports of product  $s$  to destination country  $j$ .

### 4.3 Equilibrium Trade Flows

We now characterize equilibrium trade flows in our economy, before turning to measuring economic complexity in the following sections.

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 variety. With a Fréchet distribution of productivities, this implies for the probability that country  $i \in \mathcal{I}_j^s$  is the lowest-cost provider of any given variety of product  $s$  to country  $j$  the following well-known expression (cf. Eaton and Kortum 2002 and Costinot et al. 2012)

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

Moreover, with a Fréchet distribution of productivities, the distribution of prices conditional on being the lowest-cost provider of a variety of product  $s$  to destination country  $j$  is the same for all source countries  $i$ . In turn, this implies that country  $i$ 's total sales of product  $s$  to country  $j$  are given by

$$x_{ij}^s = \frac{(w_i d_{ij}^s)^{-\theta} T_i^s}{\sum_{\hat{i} \in \mathcal{I}_j^s} (w_{\hat{i}} d_{\hat{i}j}^s)^{-\theta} T_{\hat{i}}^s} \alpha^s L_j w_j. \quad (11)$$

The key observation for our purposes is that equilibrium trade flows are intimately related to productivities. In particular, we have for any importer  $j$ , any pair of exporters  $i$  and  $i'$ , and any pair of products  $s$  and  $s'$  that they both ship to  $j$ ,

$$\frac{x_{i'j}^{s'}/x_{ij}^{s'}}{x_{i'j}^s/x_{ij}^s} = \frac{T_{i'}^{s'}/T_i^{s'}}{T_{i'}^s/T_i^s} \cdot \left[ \frac{d_{i'j}^{s'}/d_{ij}^{s'}}{d_{i'j}^s/d_{ij}^s} \right]^{-\theta}. \quad (12)$$

Expression (12) is at the heart of why a variant of the ECI can correctly rank countries—and products, for that matter—according to their economic complexity based on trade data. Ignore, for the sake of the argument, the idiosyncratic component in  $T_i^s$ , i.e. suppose that  $T_i^s$  is log-supermodular. As long as trade costs do not introduce a bias, Equation (12) then implies that economically complex countries systematically specialize in the complex products and vice versa.<sup>18</sup> Because this is true for all countries, it also implies that countries' export baskets

<sup>18</sup>This may be seen from considering the case of

$$d_{ij}^s = d_{ij} d_j^s,$$

are similar to the export baskets of countries with comparable levels of economic complexity. As we have shown in Section 3, this is precisely what allows the second eigenvector of (10) to correctly rank countries by their complexity. When applied to an appropriate country-country similarity matrix, this will also be true with stochastic  $T_i^s$  and zeros at the exporter-product level. We show this next.

## 5 A Structural Ranking of Economic Complexity

In this section, we describe how we can rank countries by their economic complexity in a world as described by our model. We begin with re-considering the measure originally proposed in Hidalgo and Hausmann (2009); Hausmann et al. (2011). As we will show, this measure may fail to correctly rank countries in a world with trade frictions. We then propose an alternative measure based on our theoretical model.

### 5.1 Economic Complexity Index in a World with Trade Frictions

As already discussed in Section 2, the ECI starts from a binary country-product matrix indicating for each country the products for which it has a Revealed Comparative Advantage (RCA) of at least one according to the Balassa (1965) measure. The RCA of country  $i$  for product  $s$  is defined as

$$RCA_i^s = \frac{\frac{X_i^s}{\sum_{\hat{s} \in \mathcal{S}} X_i^{\hat{s}}}}{\frac{\sum_{i \in \mathcal{I}} X_i^s}{\sum_{i \in \mathcal{I}} \sum_{\hat{s} \in \mathcal{S}} X_i^{\hat{s}}}},$$

where  $X_i^s$  are total global sales of product  $s$  by country  $i$ . According to our model, this simplifies to

$$RCA_i^s = \frac{X_i^s}{w_i L_i \alpha^s}, \quad (13)$$

where the equality follows from balanced trade, which implies that  $\sum_{\hat{s} \in \mathcal{S}} X_i^{\hat{s}} = w_i L_i$ , and from the fact that the expenditure share of product  $s$  is  $\alpha^s$ . Now suppose that there was free trade. Then, for any pair of exporters  $i$  and  $i'$  and any pair of products  $s$  and  $s'$  we would have

$$\frac{RCA_{i'}^{s'} / RCA_i^{s'}}{RCA_{i'}^s / RCA_i^s} = \frac{X_{i'}^{s'} / X_i^{s'}}{X_{i'}^s / X_i^s} = \frac{T_{i'}^{s'} / T_i^{s'}}{T_{i'}^s / T_i^s}, \quad (14)$$

which implies (Costinot et al., 2012, Corollary 1)

$$\frac{x_{i'j}^{s'}}{x_{i'j}^s} \geq \frac{x_{ij}^{s'}}{x_{ij}^s} \Leftrightarrow \frac{T_{i'}^{s'}}{T_{i'}^s} \geq \frac{T_i^{s'}}{T_i^s}.$$

where, for the purpose of our discussions here, we have simplified by assuming that  $\rho_i^s = 1$  for all  $i, s$ . The second equality follows from using free trade in Equation (11) and summing over export destinations. Equation (14) shows that in a world without trade frictions, the RCA inherits the log-supermodularity of the country-product specific productivities  $T_i^s$ . In principle, this would allow to infer countries' economic complexities from a country-product matrix of RCAs. Yet, this is not necessarily true in a world with trade frictions.<sup>19</sup> And even in a free-trade world, the ECI may fail to correctly rank countries by their complexity because the log-supermodularity of the RCAs is not necessarily preserved when discretizing the country-product matrix of RCAs.

Importantly, however, these potential problems are tied to the way the country-country similarity matrix is constructed, and they do not reflect conceptual shortcomings of the ECI. That is, we can correctly rank countries by applying Theorem 1 to an alternative similarity matrix. To show this, it will be instructive to first consider the case where  $T_i^s$  and  $\rho_i^s$  are known, before turning to their estimation in the next section.

## 5.2 A Structural Variant of the Economic Complexity Index

Remember that  $T_i^s$  is random, that is

$$T_i^s = \tilde{T}_i^s \epsilon_i^s,$$

where  $\tilde{T}_i^s$  is the fundamental productivity that is governed by the complementarity between a country's economic complexity and a product's complexity, and where  $\epsilon_i^s$  is an independently distributed idiosyncratic source of comparative advantage at the country-product level with mean one. This 'error' term may imply that a country with low economic complexity has a high productivity for a complex product. If anything, we can therefore hope to exploit the

---

<sup>19</sup>This may most easily be seen by means of a stylized example. In particular, consider a world with 4 countries  $i < i' < k < k'$  where, as before, these countries are identified by their respective economic complexities. Suppose that there is free trade between countries  $i$  and  $i'$  and between countries  $k$  and  $k'$ , but no trade between these pairs of countries. Suppose further that all countries are of equal size, i.e. that  $L_l = L$  for all  $l \in \{i, i', k, k'\}$ , and that for all sectors  $s$  it holds that

$$\frac{T_{i'}^s}{T_i^s} = \frac{T_{k'}^s}{T_k^s}.$$

In this stylized example we have for all products  $s$

$$RCA_i^s = RCA_k^s \quad \text{and} \quad RCA_{i'}^s = RCA_{k'}^s,$$

and a similarity matrix based on RCAs will not be log-supermodular, i.e. we cannot apply Theorem 1 to correctly rank countries.

structure imposed on trade flows by the fundamental productivities (and the probabilities of positive exports  $\rho_i^s$ ). It turns out that we can do so by jointly considering countries' exports across all products. In particular, let us define the country-country similarity matrix  $\mathbf{A}$  with elements

$$A_{ii'} := \mathbb{E} \left[ \frac{1}{S} \sum_{s \in \mathcal{S}} T_i^s z_i^s T_{i'}^s z_{i'}^s \right] = \frac{1}{S} \sum_{s \in \mathcal{S}} \mathbb{E} \left[ \tilde{T}_i^s \epsilon_i^s z_i^s \tilde{T}_{i'}^s \epsilon_{i'}^s z_{i'}^s \right] = \frac{1}{S} \sum_{s \in \mathcal{S}} \tilde{T}_i^s \rho_i^s \tilde{T}_{i'}^s \rho_{i'}^s. \quad (15)$$

$\tilde{T}_i^s$  and  $\rho_i^s$  are both log-supermodular, by Assumption 2 and 3, respectively. Hence,  $\tilde{T}_i^s \rho_i^s$  is log-supermodular as well, and  $\mathbf{A}$  is log-supermodular by Lemma 1. Applying Theorem 1 to matrix  $\mathbf{A}$  as defined in (15) therefore allows to correctly rank countries by their economic complexity. We summarize these insights in the following proposition.

**Proposition 1**

*Consider matrix  $\mathbf{A}$  as defined in Equation (15). The eigenvector corresponding to the second smallest eigenvalue of the generalized eigenproblem (16)*

$$\mathbf{L}\mathbf{y} = \lambda\mathbf{D}\mathbf{y} \quad (16)$$

*correctly ranks countries by their economic complexity up to sign. As before,  $\mathbf{D}$  denotes the diagonal matrix with entries  $D_{ii}$  equal to the respective row sum of  $\mathbf{A}$ , and  $\mathbf{L} := \mathbf{D} - \mathbf{A}$  the Laplacian matrix of  $\mathbf{A}$ .*

Proposition 1 presents our structural alternative to the ECI. As previously noted, this alternative uses the exact same eigenvector as the original ECI, but based on a structural country-country similarity matrix. In the next section, we implement this alternative and compare the ensuing country ranking to the one implied by the original ECI.

## 6 Estimated Country Rankings

In this section, we implement the ranking of economic complexity proposed in Proposition 1. Matrix  $\mathbf{A}$  as defined in Equation (15) is not directly observable from the data. We therefore begin with a discussion of the estimation of this matrix.

### 6.1 Estimating Matrix $\mathbf{A}$

Matrix  $\mathbf{A}$  as defined in Equation (15) can be estimated using a simple two-step estimator. In a first step, we can follow Costinot et al. (2012) and estimate the country-product specific

productivities  $T_i^s$  using a fixed effects regression for log trade flows

$$\ln x_{ij}^s = \delta_{ij} + \delta_j^s + \delta_i^s + v_{ij}^s, \quad (17)$$

where  $\delta_{ij}$ ,  $\delta_j^s$ , and  $\delta_i^s$  are importer-exporter, importer-product, and exporter-product fixed effects, respectively. In particular, according to our theoretical model log trade-flows satisfy

$$\ln x_{ij}^s = \delta_{ij} + \delta_j^s + \ln(T_i^s) + v_{ij}^s,$$

where  $v_{ij}^s$  is an error that captures e.g. idiosyncratic components of variable trade costs and is assumed to be orthogonal to the regressors. In such case, we can estimate  $T_i^s$  by first estimating Equation (17) using OLS and then exponentiating the estimated country-product fixed effects

$$\hat{T}_i^{s,OLS} = \exp(\hat{\delta}_i^{s,OLS}). \quad (18)$$

As noted in Costinot et al. (2012), this allows estimating the  $T_i^s$  up to normalization by some reference country and some reference product, i.e. it allows estimating

$$\frac{T_i^s / T_i^{\hat{s}}}{T_{\hat{i}}^s / T_{\hat{i}}^{\hat{s}}}$$

for some reference country  $\hat{i}$  and some reference product  $\hat{s}$ . We will further discuss this normalization in Section 6.3. Importantly, the exact choice of this normalization does not matter for our asymptotic ability to rank countries based on our estimated matrix  $\mathbf{A}$ .

In our implementation of the country ranking below, we will use the same data as the one used for computation of the original ECI. That is, we start from bilateral trade flows at the 4-digit *HS* level, and include 127 countries in our sample. This dataset includes many zeros. In our theoretical set-up, we have introduced zeros at the country-product level, which are systematically related to a country's economic complexity and a product's complexity in the same way as the fundamental productivities  $\tilde{T}_i^s$ . In addition, there are zeros at the bilateral-product level. The OLS-estimate then hinges on the assumption that  $v_{ij}^s$  is orthogonal to the regressors also when conditioning on  $x_{ij}^s > 0$ .

As an alternative, we can follow Silva and Tenreyro (2006) and estimate

$$x_{ij}^s = \exp(\delta_{ij} + \delta_j^s + \delta_i^s) + \tilde{v}_{ij}^s \quad (19)$$

using Poisson Pseudo Maximum Likelihood (PPML), which allows for zero trade flows at the

bilateral product level.<sup>20</sup> Under the assumption that  $\mathbb{E} [\tilde{v}_{ij}^s | \delta_{ij}, \delta_j^s, \delta_i^s] = 0$ ,  $T_i^s$  can then again be estimated as

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

We use both the OLS and the PPML estimator. As we will show in the next section and in Online Appendix D, the country rankings are very robust to using either of them, with—*ceteris paribus*—rank correlations between the two implied rankings of around 0.99 or higher across a broad set of robustness checks.

Equipped with our estimated  $\hat{T}_i^s$ , we can in a second step estimate matrix  $\mathbf{A}$  using its sample analogue. In particular, we can estimate element  $A_{ii'}$  as

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

where  $z_i^s$  is a binary variable that takes on value of one if country  $i$  is exporting product  $s$  at all (i.e. if we were able to estimate  $T_i^s$ ) 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 our first step regressions are consistent.<sup>21</sup>

## 6.2 Data

To estimate economic complexity, we use data on bilateral trade flows at the product level as provided by the Atlas of Economic Complexity.<sup>22</sup> This data covers more than 200 countries and is available for several years at the 4-digit HS classification level (1239 products). In our baseline specification, we use data for year 2016. From this data, we exclude all importers and exporters that are not part of the list of 127 countries included in the country rankings

---

<sup>20</sup>As noted in Hanson et al. (2015), when using PPML we can interpret the exporter product fixed effects as technologies at the country-product level in the Eaton et al. (2012) model. This model explicitly allows for zeros in international trade by considering a discrete number of random productivity draws by country (and product). It gives rise to a gravity equation in expected trade shares that can be estimated using a Multinomial Pseudo Maximum Likelihood Estimator (Eaton et al., 2012). With destination fixed effects, this estimator is equivalent to the Poisson Pseudo Maximum Likelihood Estimator (Sotelo, 2019). We do not have data on home shares and therefore estimate the gravity equation using levels. As noted by Sotelo (2019), using levels is also consistent with the Eaton et al. (2012) model, and it is asymptotically equivalent to using trade shares, with the estimators differing only in the way observations are weighted.

<sup>21</sup>This follows from rewriting  $\hat{A}_{ii'}$  as

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

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

<sup>22</sup>The data was downloaded from <http://www.atlas.cid.harvard.edu> in March 2019.

available on <http://www.atlas.cid.harvard.edu>.<sup>23</sup> To reduce noise, we then set to 0 export values of less than USD 1000 at the bilateral-product level, and drop countries' exports of a given product if they are not shipped to at least 3 destinations in that year (after dropping export values of less than USD 1000). We provide robustness checks with regards to the choice of the year and the data cleaning thresholds in Online Appendix D.

### 6.3 Rankings

To derive our country ranking of economic complexity, we first estimate Equation (17) by OLS and Equation (19) by PPML, respectively. To do so, we use the Stata *reghdfe* (Correia, 2017) and *ppmlhdfe* command (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. We normalize the estimated exporter-product fixed effects such that for every country  $i$  and every product  $s$  it holds

$$\sum_{\hat{s} \in \mathcal{S}_i} \delta_i^{\hat{s}} = 0 \quad (20a)$$

$$\sum_{\hat{i} \in \mathcal{I}^s} \delta_i^s = 0. \quad (20b)$$

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. Importantly, the exact choice of the normalization will not matter for our asymptotic ability to rank countries by their economic complexity, which, remember, relies solely on the log-supermodularity of  $T_i^s$  (and  $\rho_i^s$ ), and is therefore invariant to the normalization. We choose normalization (20) to balance countries and products and to avoid that normalized productivities scale with the random  $T_i^s$  in one reference country and product. We provide robustness checks with regards to the normalization in Online Appendix D.

To derive our estimated  $\hat{T}_i^s$ , we exponentiate the normalized exporter-product fixed effects, and then take their square root, because our country-country similarity matrix is based on a

---

<sup>23</sup>The list of countries was downloaded in March 2019 and may be seen from Table C.1. To come up with this list, the Center for International Development 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.

quadratic form.<sup>24</sup> Finally, the estimated  $\hat{T}_i^s$  are highly skewed to the right.<sup>25</sup> To avoid that our country rankings are heavily influenced by outliers, we therefore censor our estimated  $\hat{T}_i^s$  at the top. In our baseline specification, we set all values above the 95<sup>th</sup> percentile equal to the value at the 95<sup>th</sup> percentile, treating missing  $\hat{T}_i^s$  as zeros. We vary this threshold in Online Appendix D and show that the implied country rankings are similar across different choices for the threshold.

We use the estimated  $\hat{T}_i^s$  to compute the country-country similarity matrix  $\hat{\mathbf{A}}$  with element

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

where  $z_i^s$  is as previously defined. Using this estimated similarity matrix in the generalized eigenproblem (16), eventually allows to derive our alternative measure of economic complexity as the eigenvector corresponding to the second smallest eigenvalue, which following Hausmann et al. (2011) we then normalize to have mean zero and standard-deviation one. To determine the order of the ranking, we choose Japan to be ranked at the top, which implies that industrialized countries are ranked high.

Table 2 shows the original Economic Complexity Index along with the implied ranking for the top 15 and bottom 15 countries in 2016, and contrasts these with the normalized eigenvectors and implied rankings derived from our PPML and our OLS estimator, respectively. The full ranking for our list of 127 countries is provided in Appendix C.

The country rankings are surprisingly similar across the three different estimators. The rank correlation between the different rankings is 0.96 or even higher, as shown in Table 3. Interestingly, if anything, our alternative estimators tend to rank Continental European EU member states higher when compared to the original ECI (see AUT, CZE, SVN, FRA), and countries like SGP, GBR, USA, IRL, ISR lower. This pattern is actually consistent with what one might have expected based on our economic theory: Continental European countries are exposed to relatively intense competition from other complex economies which, in turn, makes exporting of the complex products more demanding for them. As opposed to the RCA, our structural estimator can account for such differences in the trade environment. Still, it is remarkable how similar overall our structural alternatives are to the original ECI,

---

<sup>24</sup>Note that taking the square-root will again not impact the log-supermodularity of  $\hat{T}_i^s$ . It will, however, 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  $T + \delta$  in one product and  $T - \delta$  in the other product. We provide a robustness check with regards to this normalization in Online Appendix D.

<sup>25</sup>The skewness of our estimates  $\hat{T}_i^s$  is 22.6 when estimated using OLS and 37.3 when using PPML.

Table 2: Country Rankings of Economic Complexity

Country	ECI	ECIrank	PPML	PPMLrank	OLS	OLSrank
JPN	2.08	1	1.49	1	1.42	1
CHE	1.99	2	1.27	4	1.20	3
DEU	1.84	3	1.32	3	1.19	4
KOR	1.74	4	1.35	2	1.26	2
SGP	1.72	5	1.11	15	0.96	20
GBR	1.59	6	1.14	9	1.03	11
SWE	1.58	7	1.19	7	1.13	8
AUT	1.53	8	1.20	6	1.14	7
CZE	1.52	9	1.21	5	1.16	6
USA	1.48	10	1.11	14	1.01	13
FIN	1.46	11	1.18	8	1.16	5
IRL	1.38	12	0.95	22	1.01	14
ISR	1.37	13	0.94	23	0.92	25
SVN	1.36	14	1.12	13	1.11	9
FRA	1.31	15	1.14	11	1.00	15
⋮	⋮	⋮	⋮	⋮	⋮	⋮
TZA	-1.30	113	-1.32	112	-1.25	111
MDG	-1.31	114	-1.24	110	-1.22	107
LAO	-1.35	115	-1.59	120	-1.67	121
LBR	-1.38	116	-1.10	105	-0.73	97
COG	-1.38	117	-1.33	113	-1.51	118
TJK	-1.44	118	-1.87	122	-2.23	122
MWI	-1.45	119	-2.39	124	-2.49	125
MRT	-1.51	120	-1.67	121	-1.27	112
CMR	-1.51	121	-1.02	103	-0.96	102
YEM	-1.60	122	-2.90	127	-3.18	127
SDN	-1.66	123	-2.83	126	-3.08	126
CIV	-1.67	124	-1.47	117	-1.38	116
NGA	-1.78	125	-1.14	107	-1.13	106
PNG	-1.92	126	-2.20	123	-2.30	124
GIN	-2.10	127	-2.42	125	-2.28	123

This table shows country rankings of economic complexity for the year 2016 using trade data at the HS4d classification level. Rankings are shown for the top 15 and bottom 15 countries according to the original Economic Complexity Index. ECI refers to the original Economic Complexity Index, PPML (OLS) to the normalized eigenvector using PPML (OLS) in the first-step regression.

given that, remember, the original ECI starts from a binary country-product matrix that indicates for each country the set of products for which it has an RCA of at least 1.

Table 3: Rank Correlations Between Different Country Rankings

	ECI	PPML	OLS
ECI	1.00	0.97	0.96
PPML		1.00	1.00
OLS			1.00

This table shows rank correlations between the different rankings of economic complexity for the year 2016 using trade data at the HS4d classification level. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression.

## 7 Ranking Products by their Complexity

As already noted in the original paper by Hidalgo and Hausmann (2009); Hausmann et al. (2011), the same logic used to rank countries according to their economic complexity can also be used to rank products according to their complexity. In particular, our economic theory implies that, on balance, products of similar levels of complexity should be intensively exported by similar sets of countries. Indeed, the same reasoning as applied to our country-country similarity matrix  $\mathbf{A}$  implies that the product-product similarity matrix  $\mathbf{B}$  with elements

$$B_{ss'} := \mathbb{E} \left[ \frac{1}{I} \sum_{i \in \mathcal{I}} T_i^s z_i^s T_i^{s'} z_i^{s'} \right] = \frac{1}{I} \sum_{i \in \mathcal{I}} \tilde{T}_i^s \rho_i^s \tilde{T}_i^{s'} \rho_i^{s'} \quad (21)$$

is log-supermodular as well. Hence, the eigenvector corresponding to the second smallest eigenvalue of

$$\mathbf{L}_B \mathbf{y} = \lambda \mathbf{D}_B \mathbf{y}$$

correctly ranks products by their complexity. Analogous to the above,  $\mathbf{D}_B$  denotes the diagonal matrix with entries  $D_{ss}$  equal to the respective row sum of  $\mathbf{B}$ , and  $\mathbf{L}_B := \mathbf{D}_B - \mathbf{B}$  the Laplacian matrix of  $\mathbf{B}$ .

To derive our product ranking, we aggregate trade data to the 2-digit HS level, because we have 127 countries in our sample based on which to evaluate the similarities of products. We then follow the exact same procedure as outlined in Sections 6.2 and 6.3, with matrix  $\mathbf{B}$  replacing matrix  $\mathbf{A}$ . To determine the order of the ranking, we require ‘Nuclear reactors, boilers, machinery and mechanical appliances’ (84) to be ranked high.

The different rankings for the 15 most and least complex products according to the original Product Complexity Index (PCI) are shown in Table 4. The full ranking of products is provided in Appendix C. According to our structural ranking using OLS in the first step,

Table 4: Product Rankings of Complexity

Product	PCI	PCIranks	PPML	PPMLranks	OLS	OLSranks
37	2.37	1	1.83	2	1.53	3
82	2.01	2	0.66	29	1.01	17
84	1.89	3	0.64	30	1.75	1
92	1.72	4	0.62	32	0.53	35
59	1.46	5	1.64	4	1.33	7
87	1.41	6	0.78	23	1.42	5
95	1.37	7	1.11	10	1.13	13
90	1.31	8	0.76	24	1.50	4
30	1.23	9	0.79	22	0.97	19
86	1.19	10	0.59	34	0.42	38
83	1.18	11	0.96	15	1.21	10
73	1.13	12	0.20	48	1.22	9
94	1.06	13	0.82	20	1.10	15
38	1.05	14	0.32	43	0.92	21
32	1.01	15	0.64	31	0.64	30
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	-1.14	83	-0.64	69	-0.16	59
67	-1.16	84	0.43	38	-0.47	67
52	-1.17	85	-1.21	82	-1.31	83
46	-1.25	86	0.15	52	-0.75	71
25	-1.29	87	-1.04	78	-0.81	74
8	-1.30	88	-1.84	95	-1.66	91
74	-1.32	89	-1.42	84	-1.08	79
15	-1.34	90	-1.08	79	-1.22	82
75	-1.35	91	0.97	14	1.13	12
9	-1.36	92	-1.73	93	-1.80	95
12	-1.38	93	-1.76	94	-1.88	96
31	-1.55	94	-0.36	64	-0.41	66
71	-1.87	95	-1.50	88	-0.32	62
26	-2.40	96	-2.27	97	-2.14	97
27	-2.94	97	-1.87	96	-1.19	80

This table shows rankings of product complexity for the year 2016 using trade data at the HS4d classification level. Rankings are shown for the top 15 and bottom 15 products according to the original Product Complexity Index. PCI refers to the original Product Complexity Index, PPML (OLS) to the normalized eigenvector using PPML (OLS) in the first-step regression.

the three most complex products are: ‘Nuclear reactors, boilers, machinery and mechanical appliances’ (84), ‘Electrical machinery and equipment and parts thereof’ (85), and ‘Photographic or cinematographic goods’ (37). The three least complex products are: ‘Ores, slag,

Table 5: Rank Correlations Between Different Product Rankings

	PCI	PPML	OLS
PCI	1.00	0.72	0.84
PPML		1.00	0.84
OLS			1.00

This table shows rank correlations between the different rankings of product complexity for the year 2016 using trade data at the HS4d classification level. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression.

and ash’ (26), ‘Oil seeds and oleaginous fruits’ (12), and ‘Coffee, tea, mate and spices’ (09). The product rankings are somewhat less robust when compared to the country rankings (see Table 5 and the robustness checks in Online Appendix D). For instance, ‘Musical instruments, parts and accessories of such articles’ (92) are ranked 4<sup>th</sup> by the original PCI, but 35<sup>th</sup> according to our structural alternative using OLS in the first step. Given that we use 127 countries to evaluate the similarities of 97 products, this may not come as a surprise. Still, the product rankings capture important aspects of product complexity, and they may serve as an alternative to proxies previously used in the literature (e.g. Levchenko 2007; Costinot 2009b; Schetter 2019).

## 8 Conclusion

In this paper, we proposed a structural variant of the Economic Complexity Index (Hidalgo and Hausmann, 2009; Hausmann et al., 2011) and showed that it correctly ranks countries according to their deep underlying economic strength. This ranking is rooted in comparative as opposed to absolute advantages, i.e. it is not necessarily reflected in countries’ GDP per capita. Our work may therefore allow for a novel perspective on the development process of countries, disentangling changes in incomes from progress in the deep underlying productive capabilities of an economy.

While our main focus was on ranking countries—and products, for that matter—by their complexity, along the way, we developed a general theoretical framework for ranking nodes in a weighted (bipartite) graph according to some unobservable characteristics. This framework may prove useful in other contexts where our main assumption of log-supermodularity is naturally satisfied. For example, we may postulate that talented scientists are systematically more successful at publishing in prestigious journals, or that ‘left-wing’ politicians have a

systematically higher probability of accepting ‘left-wing’ policies. If so, our work can readily be applied to rank politicians according to their ideological orientation or academic journals according to their prestige.

# Appendix

## A Proofs

### A.1 Proof of Lemma 1

We need to show 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 from the following chain of (in)equalities:

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

The inequality follows from noting first that  $X_{i'}^s X_k^s > 0$  and 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 then applying Chebyshev’s Sum Inequality (Hardy et al., 1934, Theorem 43).<sup>26</sup> Inequality (A.1) shows the desired result. □

---

<sup>26</sup>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.

## A.2 Proof of Theorem 1

We show the desired result by means of two lemmata. In particular, recall from (6) that the eigenvector corresponding to the second smallest eigenvalue of (10) solves a constrained minimization problem. Our strategy is therefore to show first that if we augment minimization problem (6) by an additional constraint  $\mathbf{y} \in \mathcal{A}$ , where  $\mathcal{A}$  is a closed set, the optimal solution to the augmented minimization problem is either the second eigenvector of (10), or it must be on the boundary of  $\mathcal{A}$  (Lemma 2). Second, we 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 3). The desired result then follows.

Throughout, we will use  $\lambda_k$  to denote the  $k^{\text{th}}$  smallest eigenvalue of the generalized eigenproblem (10), and  $\mathbf{u}_k$  to denote the corresponding eigenvector which we will henceforth refer to as the  $k^{\text{th}}$  eigenvector of (10).

### Lemma 2

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

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

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

### Proof:

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

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

where

$$\mathbf{z}^* := \arg \min_{\mathbf{z}^T \mathbf{z} = 1, \mathbf{z}^T \mathbf{D}^{1/2} \mathbf{1} = 0, \mathbf{D}^{-1/2} \mathbf{z} \in \mathcal{A}} \mathbf{z}^T \tilde{\mathbf{L}} \mathbf{z}, \quad (\text{A.2})$$

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

$\tilde{\mathbf{L}}$  is symmetric. Let  $\mathbf{v}_k$  be the eigenvector corresponding to  $\lambda_k$ , the  $k^{\text{th}}$  smallest eigenvalue of  $\tilde{\mathbf{L}}$ . Note that  $\lambda_k$  is the exact same eigenvalue as previously defined and that  $\mathbf{v}_k = \mathbf{D}^{1/2} \mathbf{u}_k$ . It follows that  $\mathbf{v}_1 = \mathbf{D}^{1/2} \mathbf{1}$ .<sup>27</sup> Hence, constraint  $\mathbf{z}^T \mathbf{D}^{1/2} \mathbf{1} = 0$  requires  $\mathbf{z}$  to be orthogonal to the first eigenvector of  $\tilde{\mathbf{L}}$ .

---

<sup>27</sup>Note that  $\mathbf{u}_1 = \mathbf{1}$  is the eigenvector corresponding to the smallest eigenvalue  $\lambda_1 = 0$  of the generalized eigenproblem (10).

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

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

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

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

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

where<sup>28</sup>

$$\mathbf{r}^* := \arg \min_{\mathbf{r}^T \mathbf{r} = 1, \mathbf{r}^T \mathbf{e}_1 = 0, \mathbf{D}^{-1/2} \mathbf{V} \mathbf{r} \in \mathcal{A}} \mathbf{r}^T \mathbf{\Lambda} \mathbf{r}. \quad (\text{A.3})$$

Now, suppose by way of contradiction that  $\mathbf{y}^* = \mathbf{D}^{-1/2} \mathbf{V} \mathbf{r}^* \in \mathcal{A}^\circ$ , where  $\mathcal{A}^\circ$  denotes the interior of set  $\mathcal{A}$ . On the one hand, we have  $\mathbf{r}^* \in \text{span}(\{\mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_n\})$ . On the other hand,  $\mathbf{u}_2 \notin \mathcal{A}$  and, hence,  $\mathbf{r}^* \not\parallel \mathbf{e}_2$ , by assumption. Hence, there exists an  $\tilde{\mathbf{r}}$  with

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

for some  $k > 2$ , such that

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

---

<sup>28</sup>To see this, note that using  $\mathbf{r} := \mathbf{V}^T \mathbf{z}$  allows re-writing the objective in (A.2) as

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

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

$$\begin{aligned} \mathbf{y} &= \mathbf{D}^{-1/2} \mathbf{z} = \mathbf{D}^{-1/2} \mathbf{V} \mathbf{r} \\ \mathbf{z}^T \mathbf{z} &= (\mathbf{V} \mathbf{r})^T \mathbf{V} \mathbf{r} = \mathbf{r}^T \mathbf{V}^T \mathbf{V} \mathbf{r} = \mathbf{r}^T \mathbf{r}, \end{aligned}$$

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

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

where  $\mathbf{e}_i$  denotes the  $i^{\text{th}}$  unit vector, implies that  $\mathbf{y}^* = \mathbf{D}^{-1/2} \mathbf{V} \mathbf{r}^*$  with  $\mathbf{r}^*$  as defined in (A.3). In the above, the last equality follows from the fact that  $\mathbf{v}_1 := \mathbf{D}^{1/2} \mathbf{1}$  is the first eigenvector of  $\tilde{\mathbf{L}}$ .

and

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

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

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

a contradiction to  $\mathbf{r}^*$ —and, hence,  $\mathbf{y}^*$ —being optimal.<sup>29</sup> This concludes the proof of the lemma. □

### Lemma 3

*The optimal solution to minimization problem*

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

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

#### Proof:

Note first that the set  $\mathcal{Y}$  is compact.<sup>30</sup> Hence, the continuous function  $\mathbf{y}^T \mathbf{L} \mathbf{y}$  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, we proceed by contradiction.

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

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

---

<sup>29</sup>Strictly speaking, this assumes that  $\lambda_2$  is unique. With multiplicity larger than one of this eigenvalue, our arguments imply that  $\mathbf{y}^*$  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 3 then implies that all of the eigenvectors corresponding to the second smallest eigenvalue must be monotonic—see Footnote 32.

<sup>30</sup>The definition of  $\mathcal{Y}$  immediately implies that it is closed. The constraint  $\mathbf{y}^T \mathbf{D} \mathbf{y} = 1$  implies that  $y_i^2 \leq \frac{1}{D_{ii}}$  for all  $i \in \mathcal{I}$  which proves that  $\mathcal{Y}$  is bounded.

with  $dy_i, dy_j$  small and where

$$D_{ii}dy_i = -D_{jj}dy_j. \quad (\text{A.4})$$

Clearly,  $\tilde{\mathbf{y}}^T \mathbf{D} \mathbf{1} = 0$ . Moreover, totally differentiating  $f(\mathbf{y}) := \mathbf{y}^T \mathbf{L} \mathbf{y}$  and using  $dy_k = 0$  for  $k \neq i, j$ , we get

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

where the second equality follows from the symmetry of  $\mathbf{A}$ . Using (A.4) and the fact that  $y_i^* = y_j^*$  this implies

$$\begin{aligned} df(\mathbf{y}^*) &= 2 \left[ \sum_{k \in \mathcal{I}} (y_j^* - y_k^*) \left( A_{jk} - A_{ik} \frac{D_{jj}}{D_{ii}} \right) \right] dy_j \\ &= 2 \left[ \sum_{k \in \mathcal{I}} (y_j^* - y_k^*) \left( 1 - \frac{A_{ik}}{A_{jk}} \frac{D_{jj}}{D_{ii}} \right) A_{jk} \right] dy_j. \end{aligned} \quad (\text{A.5})$$

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  $\mathbf{A}$  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

$$\begin{aligned} \left[ \sum_{k \in \mathcal{I}} (y_j^* - y_k^*) \left( 1 - \frac{A_{ik}}{A_{jk}} \frac{D_{jj}}{D_{ii}} \right) A_{jk} \right] &< \frac{\left[ \sum_{k \in \mathcal{I}} (y_j^* - y_k^*) A_{jk} \right] \cdot \left[ \sum_{k \in \mathcal{I}} \left( 1 - \frac{A_{ik}}{A_{jk}} \frac{D_{jj}}{D_{ii}} \right) A_{jk} \right]}{\sum_{k \in \mathcal{I}} A_{jk}} \\ &= 0, \end{aligned} \quad (\text{A.6})$$

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.6) 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.<sup>31</sup> Equation (A.5) and Inequality (A.6) imply that for  $dy_j > 0$  but small, moving from  $\mathbf{y}^*$  to  $\tilde{\mathbf{y}}$  strictly decreases the objective function.  $\tilde{\mathbf{y}}$  is, however, not feasible as it violates constraint

---

<sup>31</sup>Recall that we cannot have  $y_j^* = y_k^*$  for all  $k$  by the definition of set  $\mathcal{Y}$ .

$\mathbf{y}^T \mathbf{D} \mathbf{y} = 1$ . In particular,

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

where the last equality follows from using Equation (A.4) in combination with  $y_i^* = y_j^*$ , and the inequality follows from  $\mathbf{y}^{*T} \mathbf{D} \mathbf{y}^* = 1$ . It follows, however, that we can scale  $\tilde{\mathbf{y}}$  by some factor  $\beta \in (0, 1)$  such that  $\beta \tilde{\mathbf{y}}^T \mathbf{D} \beta \tilde{\mathbf{y}} = 1$ . Clearly, the vector  $\beta \tilde{\mathbf{y}} \in \mathcal{Y}$ . Moreover,

$$\beta \tilde{\mathbf{y}}^T \mathbf{L} \beta \tilde{\mathbf{y}} = \beta^2 \tilde{\mathbf{y}}^T \mathbf{L} \tilde{\mathbf{y}} < \tilde{\mathbf{y}}^T \mathbf{L} \tilde{\mathbf{y}} < \mathbf{y}^{*T} \mathbf{L} \mathbf{y}^*, \quad (\text{A.7})$$

where the first inequality follows from Equation (7) in combination with the facts that  $\mathbf{A}$  is positive valued and that  $\tilde{\mathbf{y}}$  is non-constant. Inequality (A.7) is a contradiction to  $\mathbf{y}^{*T} \mathbf{L} \mathbf{y}^*$  being minimal. This concludes the proof of the lemma. □

Lemmata 2 and 3 jointly imply Theorem 1. In particular, according to Lemma 3

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

is in the interior of set  $\mathcal{A} := \{\mathbf{y} \in \mathbb{R}^I : y_i \leq y_j \ \forall i \leq j\}$ . On the one hand, this implies that  $\mathbf{y}^*$  must be strictly monotonic by the definition of set  $\mathcal{A}$ . On the other hand, the fact that  $\mathbf{y}^*$  is in the interior of set  $\mathcal{A}$  implies that it must be the eigenvector corresponding to the second smallest eigenvalue of (10), by Lemma 2.<sup>32</sup> □

## B Details on Numerical Simulations of Section 3.3

In this appendix, we provide details and further results for the Monte Carlo simulation of Section 3.3.

---

<sup>32</sup>Note that Inequality (A.7) 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  $\lambda_2$ , 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.

As discussed in Section 3.3, the figures shown in Table 1 are based on  $80k$  randomly generated matrices  $\mathbf{A}$ — $10k$  for each column. To generate these matrices, we generate symmetric supermodular matrices  $\tilde{\mathbf{A}}$ , add noise to these matrices, and finally exponentiate them elementwise.

To generate the supermodular matrices  $\tilde{\mathbf{A}}$ , we 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  $\mathbf{A}$  with elements in contiguous rows and columns,  $i' > i$  and  $k' > k$ , respectively, we must have

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

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

1. Randomly draw an  $I \times I$  matrix  $\mathbf{R}$  with elements  $R_{ij}$  iid from a uniform distribution on  $[0, 1]$ .
2. Randomly draw an index  $i$  from the discrete uniform distribution with support  $i \in \{1, 2, \dots, I\}$ .
3. Set  $\tilde{A}_{i,:} = R_{i,:} * 100$  and  $\tilde{A}_{:,i} = R_{:,i}^T * 100$ .<sup>33</sup>
4. Fill  $\tilde{\mathbf{A}}$  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  $k = i - 1, i - 2, \dots, 1$  and  $l = i + 1, i + 2, \dots, I$ .
  - (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  $k = i - 1, i - 2, \dots, 1$  and  $l = i - 1, i - 2, \dots, k$ .
  - (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  $k = i + 1, i + 2, \dots, I$  and  $l = k, k + 1, \dots, I$ .

This procedure results in a matrix that is positive, symmetric and supermodular. We add to this matrix another  $I \times I$  symmetric random matrix  $\mathbf{S}$ , drawn iid from a uniform distribution with lower bound 0 and upper bound as shown in the respective column-header in Table 1.

---

<sup>33</sup>We scale these elements to increase the variance of elements in this initial row / column vis-à-vis the variance of shocks that govern the log-supermodularity of the matrix (see next step). We do this to emphasize that log-supermodularity is a ‘diff-in-diff’ condition and not concerned with absolute sizes of elements in different rows and columns. Note that simulation results are virtually the same when using a scaling factor of 1.

Table 6: Robustness with  $50 \times 50$  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 our  $80k$  randomly generated  $50 \times 50$  matrices— $10k$  for each column. All else is the same as described in the footer of Table 1.

Table 7: Robustness with  $10 \times 10$  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 our  $80k$  randomly generated  $10 \times 10$  matrices— $10k$  for each column. All else is the same as described in the footer of Table 1.

We then normalize all elements in this matrix by  $1/5$  of the largest element.<sup>34</sup> Finally, we exponentiate this matrix element-by-element to get the positive and symmetric matrix  $\mathbf{A}$ . For each of these matrices, we then compare the ranking of rows and columns implied by the eigenvector corresponding to the second smallest eigenvalue of (10) to the ‘true’ ranking of the underlying log-supermodular matrix  $\tilde{\mathbf{A}}$ , i.e. to a vector with elements  $[1, 2, \dots, I]$ , where  $I$  is the size of the matrix. To determine the sign of the eigenvector, we require that the sum of its first three elements must be positive, analogous to some outside information that we might use in practical applications, e.g. the requirements that industrialized countries be ranked high in case of our complexity ranking.

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

<sup>34</sup>We choose this normalization to avoid very large values in our final matrix  $\mathbf{A}$  that might cause computational problems. Note that this normalization does not affect the generalized eigenvectors (5) of matrix  $\mathbf{A}$ .

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 6 and 7, respectively. Considering our discussion from Section 3.3 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 6 and 7 confirm that the ranking is very robust to random noise as long as the size of this noise is not too big relative to the size of matrix **A**.

## C Detailed Rankings

### C.1 Full List of Country Rankings

Table 8: Country Rankings of Economic Complexity

Country	ECI	ECIrank	PPML	PPMLrank	OLS	OLSrank
JPN	2.08	1	1.49	1	1.42	1
CHE	1.99	2	1.27	4	1.20	3
DEU	1.84	3	1.32	3	1.19	4
KOR	1.74	4	1.35	2	1.26	2
SGP	1.72	5	1.11	15	0.96	20
GBR	1.59	6	1.14	9	1.03	11
SWE	1.58	7	1.19	7	1.13	8
AUT	1.53	8	1.20	6	1.14	7
CZE	1.52	9	1.21	5	1.16	6
USA	1.48	10	1.11	14	1.01	13
FIN	1.46	11	1.18	8	1.16	5
IRL	1.38	12	0.95	22	1.01	14
ISR	1.37	13	0.94	23	0.92	25
SVN	1.36	14	1.12	13	1.11	9
FRA	1.31	15	1.14	11	1.00	15
HUN	1.25	16	1.04	17	0.97	19
BEL	1.20	17	1.14	10	1.07	10
ITA	1.18	18	1.12	12	0.97	18
SVK	1.17	19	1.03	18	1.00	16
DNK	1.12	20	0.94	24	0.94	22
POL	1.10	21	1.02	19	0.93	23
NLD	1.09	22	0.97	20	0.89	27

Table 8: Country Rankings of Economic Complexity

Country	ECI	ECIrank	PPML	PPMLrank	OLS	OLSrank
EST	1.04	23	0.88	29	0.85	30
CHN	0.99	24	1.09	16	1.02	12
NOR	0.96	25	0.96	21	0.99	17
MYS	0.95	26	0.91	27	0.87	29
MEX	0.91	27	0.73	41	0.72	41
ROU	0.88	28	0.87	31	0.87	28
HRV	0.86	29	0.80	35	0.79	35
HKG	0.85	30	0.93	25	0.95	21
ESP	0.81	31	0.93	26	0.80	34
LTU	0.79	32	0.85	32	0.82	32
CAN	0.78	33	0.88	30	0.93	24
THA	0.77	34	0.85	33	0.78	36
BLR	0.77	35	0.74	37	0.75	38
LVA	0.65	36	0.75	36	0.78	37
SRB	0.62	37	0.66	46	0.69	44
PRT	0.60	38	0.81	34	0.81	33
BGR	0.59	39	0.74	39	0.82	31
RUS	0.57	40	0.89	28	0.90	26
PAN	0.55	41	0.55	49	0.43	54
UKR	0.55	42	0.64	47	0.66	46
PHL	0.55	43	0.54	50	0.63	48
BIH	0.54	44	0.44	55	0.40	56
TUR	0.53	45	0.71	42	0.72	40
ARE	0.50	46	0.74	40	0.70	43
IND	0.46	47	0.74	38	0.72	39
SAU	0.46	48	0.57	48	0.60	50
ZAF	0.38	49	0.66	45	0.62	49
LBN	0.37	50	0.17	64	0.24	63
NZL	0.36	51	0.49	52	0.58	51
GRC	0.31	52	0.52	51	0.66	47
CRI	0.26	53	0.28	60	0.23	64
BRA	0.25	54	0.71	43	0.66	45
MKD	0.22	55	0.18	63	0.21	65
SLV	0.20	56	0.08	69	0.15	68
TUN	0.19	57	0.37	56	0.42	55
JOR	0.19	58	0.05	71	0.09	71
COL	0.14	59	0.31	58	0.31	58
URY	0.14	60	0.04	72	0.04	75
AUS	0.13	61	0.67	44	0.71	42
OMN	0.06	62	0.15	67	0.15	69
KAZ	0.01	63	0.16	66	0.07	74

Table 8: Country Rankings of Economic Complexity

Country	ECI	ECIrank	PPML	PPMLrank	OLS	OLSrank
EGY	0.00	64	0.13	68	0.30	60
ARG	-0.04	65	0.34	57	0.32	57
CHL	-0.07	66	0.27	61	0.29	61
VNM	-0.10	67	0.47	53	0.58	52
DOM	-0.11	68	-0.10	76	0.00	76
IDN	-0.18	69	0.47	54	0.51	53
KGZ	-0.18	70	-0.66	95	-0.66	95
MUS	-0.19	71	-0.06	74	-0.06	78
GTM	-0.20	72	-0.12	77	-0.00	77
MDA	-0.25	73	-0.13	78	-0.10	81
QAT	-0.26	74	0.21	62	0.30	59
KWT	-0.28	75	0.17	65	0.15	70
BWA	-0.31	76	0.31	59	0.27	62
GEO	-0.35	77	-0.17	81	-0.10	80
LKA	-0.38	78	-0.14	79	0.08	72
JAM	-0.40	79	-0.73	97	-0.81	99
IRN	-0.43	80	0.08	70	0.17	67
HND	-0.45	81	-0.39	87	-0.24	84
PER	-0.45	82	-0.23	83	-0.10	79
PAK	-0.48	83	-0.02	73	0.18	66
ALB	-0.51	84	-0.38	86	-0.34	86
MNG	-0.55	85	-0.45	91	-0.58	94
NAM	-0.56	86	-0.20	82	-0.14	82
TTO	-0.56	87	-0.62	94	-0.50	90
MAR	-0.61	88	-0.07	75	0.08	73
SEN	-0.61	89	-0.74	98	-0.74	98
PRY	-0.62	90	-0.97	101	-0.87	100
CUB	-0.62	91	-1.31	111	-1.29	114
AZE	-0.65	92	-0.43	90	-0.51	91
KEN	-0.66	93	-0.70	96	-0.66	96
KHM	-0.75	94	-0.35	85	-0.35	87
UZB	-0.85	95	-0.96	100	-0.97	103
BOL	-0.85	96	-1.50	119	-1.55	120
MLI	-0.86	97	-0.96	99	-0.98	104
VEN	-0.88	98	-0.33	84	-0.42	88
ZMB	-0.95	99	-1.18	108	-1.25	110
LBY	-0.95	100	-1.07	104	-1.35	115
DZA	-0.97	101	-0.41	88	-0.52	92
ECU	-0.97	102	-0.54	93	-0.42	89
BGD	-1.02	103	-0.43	89	-0.27	85
NIC	-1.02	104	-1.37	115	-1.23	108

Table 8: Country Rankings of Economic Complexity

Country	ECI	ECIrank	PPML	PPMLrank	OLS	OLSrank
UGA	-1.05	105	-1.46	116	-1.41	117
ETH	-1.08	106	-0.99	102	-0.88	101
TKM	-1.08	107	-1.33	114	-1.24	109
ZWE	-1.14	108	-1.49	118	-1.52	119
AGO	-1.22	109	-0.14	80	-0.17	83
GAB	-1.23	110	-0.52	92	-0.57	93
MOZ	-1.27	111	-1.22	109	-1.28	113
GHA	-1.27	112	-1.12	106	-1.05	105
TZA	-1.30	113	-1.32	112	-1.25	111
MDG	-1.31	114	-1.24	110	-1.22	107
LAO	-1.35	115	-1.59	120	-1.67	121
LBR	-1.38	116	-1.10	105	-0.73	97
COG	-1.38	117	-1.33	113	-1.51	118
TJK	-1.44	118	-1.87	122	-2.23	122
MWI	-1.45	119	-2.39	124	-2.49	125
MRT	-1.51	120	-1.67	121	-1.27	112
CMR	-1.51	121	-1.02	103	-0.96	102
YEM	-1.60	122	-2.90	127	-3.18	127
SDN	-1.66	123	-2.83	126	-3.08	126
CIV	-1.67	124	-1.47	117	-1.38	116
NGA	-1.78	125	-1.14	107	-1.13	106
PNG	-1.92	126	-2.20	123	-2.30	124
GIN	-2.10	127	-2.42	125	-2.28	123

This table shows country rankings of economic complexity for the year 2016 using trade data at the HS4d classification level. ECI refers to the original Economic Complexity Index, PPML (OLS) to the normalized eigenvector using PPML (OLS) in the first-step regression.

## C.2 Full List of Product Rankings

Table 9: Product Rankings of Complexity

Product	PCI	PCIranks	PPML	PPMLrank	OLS	OLSrank
37	2.37	1	1.83	2	1.53	3
82	2.01	2	0.66	29	1.01	17
84	1.89	3	0.64	30	1.75	1
92	1.72	4	0.62	32	0.53	35
59	1.46	5	1.64	4	1.33	7
87	1.41	6	0.78	23	1.42	5

Table 9: Product Rankings of Complexity

Product	PCI	PCIranks	PPML	PPMLranks	OLS	OLSranks
95	1.37	7	1.11	10	1.13	13
90	1.31	8	0.76	24	1.50	4
30	1.23	9	0.79	22	0.97	19
86	1.19	10	0.59	34	0.42	38
83	1.18	11	0.96	15	1.21	10
73	1.13	12	0.20	48	1.22	9
94	1.06	13	0.82	20	1.10	15
38	1.05	14	0.32	43	0.92	21
32	1.01	15	0.64	31	0.64	30
85	0.96	16	0.79	21	1.71	2
68	0.95	17	0.49	37	0.55	34
70	0.93	18	0.31	44	0.71	27
93	0.86	19	1.54	6	1.39	6
66	0.84	20	1.94	1	1.31	8
49	0.83	21	0.17	50	0.44	36
91	0.83	22	1.01	13	1.05	16
54	0.82	23	1.22	9	0.90	22
88	0.78	24	0.15	51	0.30	47
35	0.77	25	0.90	16	0.62	31
39	0.77	26	0.11	54	1.11	14
96	0.70	27	0.90	17	0.73	26
48	0.63	28	0.38	42	1.01	18
29	0.62	29	0.57	36	0.96	20
55	0.58	30	1.10	11	0.69	28
56	0.55	31	0.58	35	0.40	41
33	0.55	32	0.13	53	0.41	40
42	0.53	33	1.01	12	0.60	33
69	0.51	34	0.68	28	0.75	25
40	0.50	35	-0.05	59	0.84	23
34	0.46	36	-0.01	57	0.41	39
57	0.41	37	0.68	27	0.33	45
60	0.38	38	1.24	8	0.42	37
58	0.37	39	0.88	19	0.34	44
97	0.33	40	-1.17	81	-0.86	76
45	0.32	41	1.57	5	1.20	11
72	0.32	42	-0.50	67	0.19	49
4	0.31	43	-0.33	63	-0.58	68
21	0.21	44	-0.26	62	0.00	55
64	0.19	45	0.73	25	0.37	42
19	0.18	46	-0.14	61	-0.15	58
43	0.14	47	0.89	18	-0.04	56

Table 9: Product Rankings of Complexity

Product	PCI	PCIranks	PPML	PPMLranks	OLS	OLSranks
89	0.08	48	0.39	41	0.10	52
47	0.04	49	0.22	47	-0.27	61
22	0.02	50	-0.61	68	-0.20	60
36	-0.02	51	0.31	45	0.32	46
50	-0.08	52	1.73	3	0.83	24
99	-0.10	53	-0.91	76	0.60	32
63	-0.12	54	0.06	55	0.24	48
51	-0.14	55	0.03	56	-0.37	65
2	-0.16	56	-0.06	60	-0.63	69
79	-0.18	57	0.40	40	-0.06	57
1	-0.20	58	-0.70	70	-1.20	81
80	-0.20	59	1.28	7	0.69	29
62	-0.29	60	0.42	39	0.34	43
76	-0.32	61	-0.82	74	0.02	54
20	-0.32	62	-0.76	73	-0.84	75
61	-0.34	63	0.29	46	0.12	51
11	-0.40	64	-0.37	65	-0.73	70
81	-0.44	65	0.20	49	0.08	53
53	-0.49	66	0.59	33	-0.32	63
6	-0.49	67	-0.85	75	-1.49	86
5	-0.55	68	-0.74	71	-1.35	84
65	-0.55	69	0.70	26	0.16	50
16	-0.60	70	-0.49	66	-0.76	72
23	-0.60	71	-1.49	87	-1.51	87
10	-0.61	72	-0.02	58	-0.77	73
24	-0.63	73	-1.11	80	-1.51	88
17	-0.79	74	-1.24	83	-1.52	89
41	-0.93	75	-1.72	92	-1.68	92
14	-0.95	76	-1.02	77	-1.70	93
44	-1.01	77	-1.43	85	-0.36	64
78	-1.01	78	-1.55	89	-1.79	94
13	-1.04	79	-0.75	72	-1.06	78
3	-1.08	80	-1.61	90	-1.40	85
7	-1.11	81	-1.65	91	-1.63	90
18	-1.11	82	-1.45	86	-1.05	77
28	-1.14	83	-0.64	69	-0.16	59
67	-1.16	84	0.43	38	-0.47	67
52	-1.17	85	-1.21	82	-1.31	83
46	-1.25	86	0.15	52	-0.75	71
25	-1.29	87	-1.04	78	-0.81	74
8	-1.30	88	-1.84	95	-1.66	91

Table 9: Product Rankings of Complexity

Product	PCI	PCIranks	PPML	PPMLranks	OLS	OLSranks
74	-1.32	89	-1.42	84	-1.08	79
15	-1.34	90	-1.08	79	-1.22	82
75	-1.35	91	0.97	14	1.13	12
9	-1.36	92	-1.73	93	-1.80	95
12	-1.38	93	-1.76	94	-1.88	96
31	-1.55	94	-0.36	64	-0.41	66
71	-1.87	95	-1.50	88	-0.32	62
26	-2.40	96	-2.27	97	-2.14	97
27	-2.94	97	-1.87	96	-1.19	80

This table shows rankings of product complexity for the year 2016 using trade data at the HS2d classification level. PCI refers to the original Product Complexity Index, PPML (OLS) to the normalized eigenvector using PPML (OLS) in the first-step regression.

## References

- Acemoglu, D., García-Jimeno, C., and Robinson, J. A. (2015). State capacity and economic development: A network approach. *American Economic Review*, 105(8):2364–2409.
- Albeaik, S., Kaltenberg, M., Alsaleh, M., and Hidalgo, C. A. (2017). 729 new measures of economic complexity (addendum to improving the economic complexity index). arxiv:1708.04107.
- Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers - International Monetary Fund*, 16(1):159.
- Balassa, B. (1965). Trade liberalisation and 'revealed' comparative advantage. *The Manchester School*, 33(2):99–123.
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):363–363.
- Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396.
- Bloch, F., Jackson, M. O., and Tebaldi, P. (2019). Centrality measures in networks. SSRN Scholarly Paper ID 2749124, Social Science Research Network, Rochester, NY.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117.
- 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.
- Chung, F. R. K. (1997). *Spectral Graph Theory*. American Mathematical Society, Providence, RI.

- Correia, S. (2017). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Mimeo.
- Correia, S., Guimarães, P., and Zylkin, T. (2019a). ppmlhdf: 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. (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.
- Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.
- Eaton, J., Kortum, S. S., and Sotelo, S. (2012). International Trade: Linking Micro and Macro. NBER Working Paper 17864, National Bureau of Economic Research, Inc.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.
- Golub, G. H. and van Loan, C. F. (2013). *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 4th edition.
- 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.
- 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.
- Hidalgo, C. A. and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton, NJ.
- 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.
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43.
- Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics; London*, 6(11):888–893.
- König, M. D., Liu, X., and Zenou, Y. (2018). R&D networks: Theory, empirics, and policy implications. *The Review of Economics and Statistics*, 101(3):476–491.
- König, M. D., Rohner, D., Thoenig, M., and Zilibotti, F. (2017). Networks in conflict: Theory and evidence from the Great War of Africa. *Econometrica*, 85(4):1093–1132.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70(5):950–959.

- 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.
- Liao, H., Mariani, M. S., Medo, M., Zhang, Y.-C., and Zhou, M.-Y. (2017). Ranking in evolving complex networks. *Physics Reports*, 689:1–54.
- Mealy, P., Farmer, J. D., and Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, 5(1):eaau1705.
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.
- Morrison, G., Buldyrev, S. V., Imbruno, M., Arrieta, O. A. D., Rungi, A., Riccaboni, M., and Pammolli, F. (2017). On economic complexity and the fitness of nations. *Scientific Reports*, 7(15332).
- Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600.
- Perry, M. and Reny, P. J. (2016). How to count citations if you must. *American Economic Review*, 106(9):2722–2741.
- Petralia, S., Balland, P.-A., and Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, 46(5):956–969.
- Poncet, S. and Starosta de Waldemar, F. (2013). Export upgrading and growth: The prerequisite of domestic embeddedness. *World Development*, 51:104–118.
- 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](http://www.salaimartin.com/media/pdf/1.3_The_Global_Comp_Index.pdf) in August 2019.
- Schetter, U. (2019). Quality differentiation, comparative advantage, and international specialization across products. Technical report, SSRN. <https://ssrn.com/abstract=3091581>.
- Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*. Springer Science+Business Media.

- Servedio, V. D. P., Buttà, P., Mazzilli, D., Tacchella, A., and Pietronero, L. (2018). A new and stable estimation method of country economic fitness and product complexity. arxiv:1807.10276.
- Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.
- Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.
- Sotelo, S. (2019). Practical aspects of implementing the multinomial pml estimator. Mimeo.
- 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.

# Online Appendix

## D Robustness of Rankings

In this part of the appendix, we provide robustness checks for the country rankings of Section 6.3 and for the product rankings of Section 7. We will in turn vary our choices of thresholds for data-cleaning, the censoring threshold for outliers, our normalization of exporter-product fixed effects, and the year. For each of these robustness checks, we provide rank correlations for the implied country and product rankings, respectively, across the different specifications. Apart from the respective robustness check under consideration, data and data cleaning choices are the same as in our baseline specification of Sections 6.2 and 6.3. That is, we use bilateral trade data for 127 exporters and importers at the 4-digit HS level (2-digit HS level for the case of product rankings) for the year 2016. We drop export values of less than USD 1000 at the bilateral-product level, 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. We normalize the estimated exporter-product fixed effects such that for every country  $i$  and every product  $s$  it holds

$$\begin{aligned}\sum_{\hat{s} \in \mathcal{S}_i} \delta_i^{\hat{s}} &= 0 \\ \sum_{\hat{i} \in \mathcal{I}^s} \delta_i^s &= 0 ,\end{aligned}$$

and take the square root of the exponentiated fixed effects to account for the fact that our objective is a quadratic form. We finally censor outliers by setting exporter-product fixed-effects in the top 5% equal to the value at the 95<sup>th</sup> percentile.

Further details on the various robustness checks are provided in the footnotes to the respective table.

## D.1 Robustness of Country Rankings

Table 10: Robustness of Country Rankings to Minimum Threshold for Number of Export Destinations by Exporter-Product

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

This table shows rank correlations between different rankings of economic complexity. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘cut\_x’ indicates that prior to our first-step regression, we dropped all exporter-product observations 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 11: Robustness of Country Rankings to Minimum Threshold for Trade flows at the Bilateral Product Level

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

This table shows rank correlations between different rankings of economic complexity. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘minx\_x’ indicates that prior to our first-step regression, we dropped export values of less than USD x at the bilateral-product level. All other specifications are as described at the onset of this appendix.

Table 12: Robustness of Country Rankings to Normalization of Exporter-Product Fixed Effects

	ECI	norm_lsum		norm_cossim		norm_vECI		norm_nsqrt	
		PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS
ECI	1.00	0.97	0.96	0.97	0.96	0.97	0.97	0.97	0.95
PPML_norm_lsum		1.00	1.00	0.99	0.98	1.00	0.99	0.99	0.98
OLS_norm_lsum			1.00	0.99	0.99	0.99	1.00	0.98	0.98
PPML_norm_cossim				1.00	0.99	0.99	0.98	0.97	0.96
OLS_norm_cossim					1.00	0.98	0.99	0.96	0.95
PPML_norm_vECI						1.00	0.99	0.98	0.97
OLS_norm_vECI							1.00	0.97	0.96
PPML_norm_nsqrt								1.00	0.99
OLS_norm_nsqrt									1.00

This table shows rank correlations between different rankings of economic complexity. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘norm\_x’ indicates which normalization has been used. ‘norm\_lsum’ denotes our baseline normalization. All other normalizations start from this baseline. ‘norm\_cossim’ denotes the cosine similarity, that is, the country-country similarity matrix  $\hat{\mathbf{A}}$  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  $\hat{\mathbf{A}}$  has elements

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

Finally, in rows and columns ‘norm\_nsqrt’, the exponentiated exporter-product fixed effects have been used directly, instead of the square-root thereof. All other specifications are as described at the onset of this appendix.

Table 13: Robustness of Country Rankings to Censoring Threshold for Outliers of Exporter-Product Fixed Effects

		cens_90		cens_92.5		cens_95		cens_97.5		cens_99	
	ECI	PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS
ECI	1.00	0.96	0.96	0.97	0.96	0.97	0.96	0.97	0.97	0.98	0.97
PPML_cens_90		1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99	0.98	0.98
OLS_cens_90			1.00	0.99	1.00	0.99	1.00	0.98	0.99	0.97	0.98
PPML_cens_92.5				1.00	0.99	1.00	0.99	0.99	0.99	0.98	0.99
OLS_cens_92.5					1.00	0.99	1.00	0.99	0.99	0.98	0.99
PPML_cens_95						1.00	1.00	1.00	0.99	0.99	0.99
OLS_cens_95							1.00	0.99	1.00	0.99	0.99
PPML_cens_97.5								1.00	1.00	1.00	0.99
OLS_cens_97.5									1.00	0.99	1.00
PPML_cens_99										1.00	0.99
OLS_cens_99											1.00

This table shows rank correlations between different rankings of economic complexity. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘cens\_x’ indicates that normalized exporter-product fixed effects have been censored if they were above the  $x^{th}$  percentile. All other specifications are as described at the onset of this appendix.

Table 14: Rank Correlations of Country Rankings Across Different Years

	year_2012			year_2013			year_2014			year_2015			year_2016		
	ECI	PPML	OLS												
ECI_year_2012	1.00	0.95	0.95	0.99	0.95	0.94	0.98	0.95	0.95	0.98	0.96	0.95	0.98	0.95	0.95
PPML_year_2012		1.00	0.99	0.95	1.00	0.99	0.95	0.99	0.98	0.95	0.98	0.97	0.95	0.98	0.98
OLS_year_2012			1.00	0.95	0.99	1.00	0.95	0.99	0.99	0.95	0.98	0.98	0.95	0.98	0.98
ECI_year_2013				1.00	0.95	0.95	0.99	0.95	0.95	0.98	0.96	0.95	0.98	0.95	0.94
PPML_year_2013					1.00	0.99	0.96	0.99	0.99	0.95	0.99	0.98	0.96	0.99	0.98
OLS_year_2013						1.00	0.95	0.99	0.99	0.95	0.99	0.98	0.95	0.99	0.99
ECI_year_2014							1.00	0.96	0.96	0.99	0.96	0.96	0.97	0.95	0.95
PPML_year_2014								1.00	0.99	0.96	0.99	0.99	0.96	0.99	0.98
OLS_year_2014									1.00	0.96	0.99	0.99	0.96	0.98	0.99
ECI_year_2015										1.00	0.97	0.96	0.98	0.96	0.95
PPML_year_2015											1.00	0.99	0.96	0.99	0.99
OLS_year_2015												1.00	0.96	0.98	0.99
ECI_year_2016													1.00	0.97	0.96
PPML_year_2016														1.00	1.00
OLS_year_2016															1.00

This table shows rank correlations between different rankings of economic complexity. ECI refers to the original Economic Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘year- $x$ ’ indicates that trade data for year  $x$  has been used. All other specifications are as described at the onset of this appendix.

## D.2 Robustness of Product Rankings

Table 15: Robustness of Product Rankings to Minimum Threshold for Number of Export Destinations by Exporter-Product

	cut_1			cut_3			cut_5		
	PCI	PPML	OLS	PCI	PPML	OLS	PCI	PPML	OLS
PCI_cut_1	1.00	0.72	0.85	1.00	0.69	0.83	0.99	0.67	0.79
PPML_cut_1		1.00	0.81	0.74	0.98	0.84	0.74	0.95	0.86
OLS_cut_1			1.00	0.86	0.79	0.99	0.86	0.76	0.95
PCI_cut_3				1.00	0.72	0.84	0.99	0.70	0.81
PPML_cut_3					1.00	0.84	0.73	0.99	0.89
OLS_cut_3						1.00	0.84	0.82	0.98
PCI_cut_5							1.00	0.72	0.82
PPML_cut_5								1.00	0.89
OLS_cut_5									1.00

This table shows rank correlations between different rankings of product complexity. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘cut\_x’ indicates that prior to our first-step regression, we dropped all exporter-product observations 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 16: Robustness of Product Rankings to Minimum Threshold for Tradeflows at the Bilateral Product Level

	minx_0			minx_1000			minx_10000		
	PCI	PPML	OLS	PCI	PPML	OLS	PCI	PPML	OLS
PCI_minx_0	1.00	0.70	0.84	1.00	0.71	0.84	0.99	0.75	0.79
PPML_minx_0		1.00	0.77	0.71	0.99	0.84	0.67	0.97	0.91
OLS_minx_0			1.00	0.84	0.78	0.99	0.82	0.80	0.92
PCI_minx_1000				1.00	0.72	0.84	0.99	0.75	0.79
PPML_minx_1000					1.00	0.84	0.68	0.98	0.92
OLS_minx_1000						1.00	0.81	0.86	0.96
PCI_minx_10000							1.00	0.73	0.77
PPML_minx_10000								1.00	0.94
OLS_minx_10000									1.00

This table shows rank correlations between different rankings of product complexity. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘minx\_x’ indicates that prior to our first-step regression, we dropped export values of less than USD x at the bilateral-product level. All other specifications are as described at the onset of this appendix.

Table 17: Robustness of Product Rankings to Normalization of Exporter-Product Fixed Effects

	PCI	norm_lsum		norm_cossim		norm_vECI		norm_nsqrt	
		PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS
PCI	1.00	0.72	0.84	0.71	0.83	0.67	0.83	0.74	0.86
PPML_norm_lsum		1.00	0.84	1.00	0.85	0.99	0.89	0.96	0.74
OLS_norm_lsum			1.00	0.84	1.00	0.78	0.99	0.84	0.96
PPML_norm_cossim				1.00	0.85	0.99	0.89	0.96	0.74
OLS_norm_cossim					1.00	0.78	0.99	0.84	0.96
PPML_norm_vECI						1.00	0.84	0.95	0.67
OLS_norm_vECI							1.00	0.88	0.93
PPML_norm_nsqrt								1.00	0.78
OLS_norm_nsqrt									1.00

This table shows rank correlations between different rankings of product complexity. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘norm\_x’ indicates which normalization has been used. ‘norm\_lsum’ denotes our baseline normalization. All other normalizations start from this baseline. ‘norm\_cossim’ denotes the cosine similarity, that is, the product-product similarity matrix  $\hat{\mathbf{B}}$  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  $\hat{\mathbf{B}}$  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 have been used directly, instead of the square-root thereof. All other specifications are as described at the onset of this appendix.

Table 18: Robustness of Product Rankings to Censoring Threshold for Outliers of Exporter-Product Fixed Effects

		cens_90		cens_92.5		cens_95		cens_97.5		cens_99	
	PCI	PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS
PCI	1.00	0.70	0.82	0.71	0.83	0.72	0.84	0.72	0.84	0.69	0.85
PPML_cens_90		1.00	0.85	1.00	0.84	0.99	0.82	0.98	0.81	0.93	0.80
OLS_cens_90			1.00	0.86	1.00	0.86	0.99	0.85	0.99	0.78	0.97
PPML_cens_92.5				1.00	0.85	1.00	0.84	0.98	0.83	0.93	0.82
OLS_cens_92.5					1.00	0.85	1.00	0.84	0.99	0.78	0.98
PPML_cens_95						1.00	0.84	0.99	0.84	0.95	0.83
OLS_cens_95							1.00	0.83	1.00	0.77	0.99
PPML_cens_97.5								1.00	0.83	0.98	0.83
OLS_cens_97.5									1.00	0.78	1.00
PPML_cens_99										1.00	0.78
OLS_cens_99											1.00

This table shows rank correlations between different rankings of product complexity. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘cens\_x’ indicates that normalized exporter-product fixed effects have been censored if they were above the  $x^{th}$  percentile. All other specifications are as described at the onset of this appendix.

Table 19: Rank Correlations of Product Rankings Across Different Years

	year_2012			year_2013			year_2014			year_2015			year_2016		
	PCI	PPML	OLS												
PCI_year_2012	1.00	0.67	0.75	0.95	0.66	0.77	0.94	0.69	0.78	0.90	0.67	0.79	0.94	0.65	0.77
PPML_year_2012		1.00	0.86	0.67	0.99	0.85	0.68	0.99	0.86	0.66	0.99	0.86	0.73	0.98	0.86
OLS_year_2012			1.00	0.75	0.85	0.99	0.80	0.84	0.99	0.78	0.84	0.99	0.82	0.82	0.99
PCI_year_2013				1.00	0.67	0.77	0.96	0.68	0.77	0.92	0.67	0.78	0.93	0.64	0.76
PPML_year_2013					1.00	0.85	0.68	0.98	0.85	0.65	0.98	0.85	0.72	0.98	0.86
OLS_year_2013						1.00	0.81	0.85	0.99	0.78	0.84	0.99	0.82	0.82	0.99
PCI_year_2014							1.00	0.70	0.81	0.95	0.68	0.82	0.96	0.66	0.81
PPML_year_2014								1.00	0.86	0.66	0.99	0.85	0.76	0.98	0.86
OLS_year_2014									1.00	0.79	0.85	0.99	0.84	0.84	0.99
PCI_year_2015										1.00	0.65	0.80	0.93	0.62	0.79
PPML_year_2015											1.00	0.85	0.74	0.99	0.86
OLS_year_2015												1.00	0.85	0.83	0.99
PCI_year_2016													1.00	0.72	0.84
PPML_year_2016														1.00	0.84
OLS_year_2016															1.00

This table shows rank correlations between different rankings of product complexity. PCI refers to the original Product Complexity Index, PPML (OLS) to the alternative ranking using PPML (OLS) in the first-step regression. ‘year\_ $x$ ’ indicates that trade data for year  $x$  has been used. All other specifications are as described at the onset of this appendix.