WORKING PAPERS

N° 1648

June 2025

“The Theory of Economic Complexity”

César Hidalgo and Viktor Stojkoski

Toulouse School of Economics

The Theory of Economic Complexity

César A. Hidalgo[*1,2,3] and Viktor Stojkoski[1,2,4]

[1]Center for Collective Learning, IAST, Toulouse School of Economics
[2]Center for Collective Learning, CIAS, Corvinus University of Budapest
[3]Alliance Manchester Business School, University of Manchester
[4]Ss. Cyril and Methodius University in Skopje

June 24, 2025

Abstract

Economic complexity estimates rely on eigenvectors derived from matrices of specialization to explain differences in economic growth, inequality, and sustainability. Yet, despite their widespread use, we still lack a principled theory that can deduce these eigenvectors from first principles and place them in the context of a mechanistic model. Here, we calculate these eigenvectors analytically for a model where the output of an economy in an activity increases with the probability the economy is endowed with the factors required by the activity. We show that the eigenvector known as the Economic Complexity Index or ECI is a monotonic function of the probability that an economy is endowed with a factor, and that in a multi-factor model, it is an estimate of the average endowment across all factors. We then generalize this result to other production functions and to a short-run equilibrium framework with prices, wages, and consumption. We find that our main result does not depend on the introduction of prices or wages, and that the derived wage function is consistent with the convergence of economies with a similar level of complexity. Finally, we use this model to explain the shape of networks of related activities, such as the product space and the research space. These findings solve long standing theoretical puzzles in the economic complexity literature and validate the idea that metrics of economic complexity are estimates of an economy being endowed with multiple factors.

1 Introduction

For about two decades, scholars working in economic complexity have focused on two key goals. The first one, is the development of economic complexity metrics to explain international and regional variations in economic growth [1–20], inequality [21–29], and sustainability [30–39].[1] The second one, is mapping networks of related activities, such as the product space [2, 48], industry space [49–51], or research space [52, 53], to explain path dependencies in economic development[54]. Yet, despite several attempts to develop a mathematical theory of economic complexity [1, 2, 12, 13, 55–64], we still lack an analytical connection between the metrics used in the empirical literature and a production function based model that can provide a clear interpretation of these metrics in terms of a model's assumptions.[2]

Here, we connect empirical economic complexity work with a few theoretical models to provide four contributions.

*Corresponding author: cesar.hidalgo@tse-fr.eu

[1]among other outcomes [40–47]

[2]for a review of the field see [65, 66]

First, we calculate the eigenvector known as the economic complexity index (ECI) for a capability model that expands Kremer's O-Ring model model of development [67]. In our model, economies (such as countries or cities) are endowed with a probability of having the capabilities required by each activity (such as products or industries). This means the the output of economies is constrained by the capabilities they have and the geography of activities is limited to the places endowed with the capabilities they require. We solve the one capability instance of this model analytically to show that the economic complexity index, or ECI, divides economies among those with an above and below average probability of having the capability. Interestingly, this property is independent of how capabilities are distributed and can be generalized to other production functions (such as a shifted Cobb-Douglas factor intensity function).

Second, we extend this result numerically to models involving many capabilities assigned idiosyncratically to each economy. We show that in this case ECI is a monotonic function of the average capability endowment of an economy. By exploring models combining correlated and uncorrelated capabilities, we show this result to be robust to substantial levels of noise, holding even when more than 50 percent of an economy's capability endowments are random. This helps show that ECI is indeed a measure of economic complexity, since it is a statistic informing us if an economy is endowed with multiple capabilities.

Third, we extend the single capability model to a short-run equilibrium model and calculate wages, prices, and consumption. We show analytically that under these assumptions ECI still separates economies among those with high and low capability endowments. We also determine an equilibrium wage to help interpret the known empirical relationship between economic complexity and growth, and show that the prices of goods in this model follows a concave function of their capability requirements.

Finally, we use the multi-capability model to explain known variations in the shapes of networks of related activities, such as the product space (based on product co-exports) and the research space (based on co-publication patterns). We show that the core-periphery structure observed in the product space[48], comes from correlated capability endowments and that the ring structures observed in networks of related research fields[52, 68] can be explained by capability endowments following a circulant matrix.

There are a few reasons why these results should interest those working on economic complexity and international development.

First, while the economic complexity index or ECI enjoys wide adoption in policy circles[3], the lack of a theoretical foundation has left them open to critics of being an ad-hoc or uninterpretable measure [13, 62, 80, 81]. Our findings provide a clear interpretation of ECI in the context of multi-factor model of production. We show that ECI provides a monotonic estimate of the factor endowment of an economy derived from a multi-product specialization matrix. This provides an interpretation in terms of a model's parameters that is consistent with previous work exploring the interpretability of the economic complexity index as a clustering method [57, 60, 61] and the connection between ECI and log-supermodularity [56].

Second, these findings dispel the notion that economic complexity is a measure of diversity, as it was originally suggested [1]. The analytical solutions show that economies specialized

[3]For example, it is the number one mission of Malaysia's New Industrial Master Plan [69], it was used in the recent competitiveness report by Mario Draghi [70], and it is a key development target for rich resource intensive economies, such as Saudi Arabia and the United Arab Emirates. It has also motivated the creation of regional reports for Australia[71], Turkey [72], Uruguay [73], Russia[74, 75], Mexico [76, 77], Quebec[78], and Italy[79], among other places.

in the largest number of activities (the more diverse economies) are not necessarily the ones with the highest probability of having a capability[4]. In fact, the model predicts that economic development is a process of diversification only until a certain point, since economies with the highest capability endowments specialize in complex activities–and are therefore–less diverse than slightly less complex economies. This also provides a theoretical foundation for the finding that countries at high-level of development tend to specialize (e.g., Imbs and Warcziag [82]), which is consistent with the notion that ECI is higher for "small" yet sophisticated economies, such as those of Singapore, Switzerland, and Finland, while larger and more diverse economies, like those of Spain and Italy, are not necessarily as complex. Still, the model predicts a positive correlation between capability endowments and diversity, but through a non-monotonic function, explaining why measures of diversity or concentration are non-ideal estimates of the complexity of an economy.

Third, these results also provide a mean to interpret the structure of the networks of related activities, such as the product space [48], industry space [49], or research space [52]. These networks have been used extensively to model path dependencies and generate measures of export or employment potential [49–52, 54, 65, 83–90]. Yet, the structure of these networks differs depending on the data used to generate them. For instance, product space networks derived from exports data, are known to have a core composed of densely interconnected activities with are high in complexity [48]. Research spaces, connecting academic fields based on citations [68] or co-authorships [52], follow a ring structure, with fields connected with a few neighbors and without a clear center[5]. While these differences in structure are self-evident, we hitherto lack a way to explain them based on the mechanics of a model. Here, we show how to generate these structures by changing the shape of the capability endowment matrices.

Finally, we present a short-run equilibrium version of the model showing that our main result is robust to these additional assumptions.

Together, these findings help solve some long-standing puzzles in the economic complexity and international development literature by providing a bridge linking its theoretical and empirical contributions.

1.1 Empirical and Theoretical Work in Economic Complexity

Empirical work in economic complexity usually starts with matrices summarizing the geography of many economic activities (e.g., exports by country and product, payroll by city and industry, patents by city and technology, etc.). These rectangular matrices (or bipartite networks) are then used for two things. The first one is to estimate networks of similar activities [48, 49, 52, 54, 65, 83, 91–93] which are used to create measures of diversification potential or relatedness. These measures are based on the notion that economies are more likely to enter (and less likely to exit) activities that share capabilities with each other.

The second one is to create measures of an economy's structure, or of the value of the portfolio of activities an economy specializes in, known as measures of economic complexity [1, 2, 11–13, 65, 94, 95]. These measures were also motivated as estimates of the capabilities available in

[4]The notion that economic complexity is different from diversity was noted theoretically by [57] and has been in the literature from the beginning, since the work introducing ECI showed that measures of diversity or concentration, such as entropy or the Herfindahl-Hirschmann index (HHI) failed to explain future economic growth as ECI did [1].

[5]In simple, the ring: medicine, biology, chemistry, physics, computer science and math, economics, cognitive science, neuroscience, and back to medicine

an economy [1] and are often based on the assumption that high-complexity economies specialize in high-complexity activities. In fact, the definition of economic complexity introduced by [1], the economic complexity index of ECI, defines the complexity of an economy as the average complexity of the activities it specializes in, and the complexity of an activity as the average complexity of the economies specialized in that activity.[6]. These measures of complexity, in particular ECI enjoy wide adoption in international and regional development circles, as they have been shown to be robust estimators of future economic growth [1–13], and of international variations in inequality [21–25, 29, 36], and emissions [30–35, 37, 38, 96]. This supports the notion that an economy's pattern of specialization matters for subsequent economic development [97, 98], which was a key policy intuition motivating these quantitative efforts[7]. Yet, despite copious empirical work, we still lack an understanding of why the eigenvectors of these matrices are a good metric of an economy's complexity and a good predictor of its subsequent development.

Theoretical work on economic complexity has focused instead on the construction of models of economic development and innovation that follow a combinatorial tradition [1, 55, 67, 108–116]. This tradition builds on the notion that economies are endowed with capabilities [2, 117–121], or factors, which activities may or may not require. Since these capabilities are complementary, the producing an activity requires the simultaneous presence of many of them. That's why these theories have been dubbed as the "Lego" or "Scrabble" theory of development. In these models, the ability of economies to produce a product depend on having the right combination of capabilities, like in a proverbial game of scrabble where products are "words" and economies are endowed with "letters."

Here we build on the combinatorial model introduced by [1], which is a generalization of Kremer's O-Ring model of development [67].

The O-Ring model assumes a multi-step production process where the output of an economy is the product of the probabilities that it succeeds at each step. In other words, producing an item in the O-Ring model requires a sequence of tasks, each of which has a probability of failing. This implies that the output of an economy decays exponentially with the length of the production chain at a rate determined by the probability of succeeding at a task. The key outcome is that economies with higher probabilities of completing a task should specialize in activities requiring multiple steps[8]. Here we focus on a generalized version of this model, where economies are endowed with probabilities of having a capability (similar to the probability of succeeding at a task in the O-Ring model) and where activities also differ in the probability of requiring a capability. This allows us to model matrices involving an arbitrary number of economies, activities, and capabilities, while also making the capabilities specific to activities and economies. The resulting matrices, which can be made equivalent in size to the ones used in the empirical literature, can be used to create theoretical estimates of the economic complexity eigenvector ECI that we can connect to the key parameter in the model: the probability that an economy is endowed with a capability.

[6]A similar definition was proposed over a decade later by [13]. In their words: "If a country is known to be more capable than another, say the United States (US) versus Bangladesh (BG), then one can identify any good k as more complex than another reference good k_0 if, relative to the reference good, it is more likely to be exported by the United States than Bangladesh. [...] Conversely, if a good is known to be more complex than another, say medicines (ME) versus men's underwear (UW), then one can identify any country i_1 as more capable than another reference country i_0 if, relative to the reference country, it is more likely to export medicines than underwear. "

[7]This policy intuition is connected to an old debate in development economics, going back to at least Alexander Hamilton's Report on Manufactures [99], which advocated for the industrialization of the United States, and has been central to the works of scholars such as Rosenstein-Rodan [100, 101], Rostow[102], Hirschman[103], Prebisch[104], Gerschenkron[105], and Balassa[106]. For a discussion on how these different development theories related to economic complexity see [107]

[8]see also [122] for an extension of the O-Ring model to trade.

We find that, for a wide variety of model specifications, ECI recovers the key parameter in the model: the probability that an economy is endowed with a capability. This result is valid for models involving multiple capabilities, even when these are highly idiosyncratic, since ECI recovers the average of the probability that an economy is endowed with a capability. We then generalize this finding to a Cobb-Douglas type factor intensity production function and use this to show that the ability of ECI to separate among better and worse endowed economies can be generalized to any shifted production function of the form $Y_{cp} = B + f_c g_p$ where f_c is a general function characterizing an economy and g_p is a general function characterizing an activity.

As in Kremer's O-Ring model [67], our initial exercise builds on a supply-side model that assumes prices are exogenous and does not provide an explicit model of wages or demand. So, we then embed our model in a short-run equilibrium framework and estimate functions for the implied wages, consumption, and prices. We show that wages increase with the capability endowment of an economy, consumption grows with income, and prices are higher for more demanding products (products having a higher probability of requiring a capability). Surprisingly, our main result (that the economic complexity eigenvector separates among high- and low-capability economies) is unaffected by the introduction of these additional assumptions.

Finally, we use this model to explore the connection between the structure of networks of related activities, such as the product space and research space, and show that it is possible to generate networks with a similar structure than the ones observed in the empirical literature by manipulation the capability endowment and requirements of economies and activities.

The remainder of the paper is organized as follows. The next section (Section 2) introduces the single-capability model and solves it analytically. Section 3 generalizes these results numerically to several versions of a multi-capability model. Section 4 explores additional production functions and Section 5 embeds the model in a short-run equilibrium framework. Section 6 uses the model to explain the structure of networks of related activities, and the Section 7 concludes.

2 The Single Capability Model

We start with the basic model of economic complexity introduced in [1] and we will move later to other production functions.

This model assumes that an economy c is endowed with capability b with probability $r_{c,b}$ and that activity p requires a capability b with probability $q_{p,b}$.[9]

For pedagogical reasons, we start with the case in which there is a single capability or factor, and an arbitrary number of economies and activities (that is $r_{c,b} \to r_c$ and $q_{p,b} \to q_p$). This case will allow us to get a basic intuition that we will then generalize to more complex functional forms. The advantage of starting with the single capability model is that it can be solved analytically, so it can provide us with a baseline understanding that we can extend to the numerical solutions derived for more complex models.

Let the output Y_{cp} of economy c in activity p be given by the matrix:

$$Y_{cp} = A(1 - q_p(1 - r_c)) \tag{1}$$

[9]This deviates from previous work [1, 55] which tended to assume either a single r and q for all economies and activities or a known distribution.

where A is a constant or scale factor and $1 - q_p(1 - r_c)$ is the probability that economy c has the capability that product p requires. This probability is written as a complement. That is, one minus the probability that the activity requires the capability (q_p) and the economy does not have it ($1 - r_c$). In matrix form, the output matrix follows the form:

$$Y_{cp} = A \begin{bmatrix} 1 - q_1(1 - r_1) & 1 - q_2(1 - r_1) & \dots \\ 1 - q_1(1 - r_2) & \dots & \dots \\ \dots & \dots & 1 - q_N(1 - r_N) \end{bmatrix} \tag{2}$$

Going forward, we always sort rows in descending order of r and columns in ascending order of q. That is, the first cell of the matrix (Y_{11}) is the output of the economy with the highest probability of having the capability in the activity with the lowest probability of requiring it. This sorting convention will greatly facilitate the visual inspection of these matrices.

A key difference between this implementation of the model and previous work[1, 55] is that here we use the model to simulate an output matrix, whereas previous work used it to simulate a specialization matrix (which has already been through important manipulations and normalizations). We find that doing these normalization steps explicitly is essential for obtaining the right connection between the economic complexity index and the model parameters. So, next, we perform the manipulations applied to output matrices in the empirical literature to derive a metric of economic complexity. These are:

(i) Estimating the matrix of revealed comparative advantage or RCA R_{cp} according to Balassa's (1965) definition. This matrix normalizes the output matrix by the sum of its rows and columns and it is equivalent to a matrix comparing the observed output (Y_{cp}) with the expected output in a probabilistic model (see eqn. (3)). RCA is also known as the location quotient (LQ) in economic geography and innovation studies.

(ii) Estimating the binary specialization matrix M_{cp}. This is a matrix that is 1 if $R_{cp} \geq 1$ and 0 otherwise. This binary matrix is motivated in the empirical literature as a way to remove the tails of the R_{cp} matrix, since the ratio definition of R_{cp} results in larger variance for economies with low levels of output (small $Y_c = \sum_p Y_{cp}$) and activities with small markets (small $Y_p = \sum_c Y_{cp}$).

(iii) Estimating the $M_{cc'}$ matrix. This is a square matrix connecting economies with similar specialization patterns and is the one used to derive the economic complexity index. This matrix is defined using the reciprocal averaging method known as the method of reflections [1], but it can also be defined as the product of a four matrices related to M_{cp} (we will introduce the exact formula at that point).

We begin with the standard definition of the RCA matrix or R_{cp} which is:

$$R_{cp} = \frac{Y_{cp} \sum_{c,p} Y_{cp}}{\sum_c Y_{cp} \sum_p Y_{cp}} \tag{3}$$

Also, since it will simplify the math going forward, we use Einstein's notation, where summed indices are "suppressed" or "muted" (e.g. $Y_c = \sum_p Y_{cp}$). In this notation R_{cp} takes the more compact form:

$$R_{cp} = \frac{Y_{cp}Y}{Y_c Y_p} \tag{4}$$

To estimate R_{cp} for the single capability model we need to notice a couple of things. First, since the scale factor A is common to all terms, it cancels out of R_{cp} (so we can ignore it). Second, we should notice that applying the sum operator to the terms in R_{cp} transforms variables into averages. We can illustrate this by using the sum over p as an example (the derivation is analogous for the other terms):

$$\begin{aligned} Y_p &= \sum_p (1 - q_p(1 - r_c)) \\ &= N_p - (1 - r_c)\sum_p q_p \\ &= N_p(1 - (1 - r_c)\langle q \rangle) \end{aligned} \tag{5}$$

where N_p is the number of activities or products and $\langle q \rangle$ is the average of q_p over all activities. Using this property, we can now rewrite R_{cp} as:

$$R_{cp} = \frac{(1 - q_p(1 - r_c))(1 - \langle q \rangle(1 - \langle r \rangle))}{(1 - q_p(1 - \langle r \rangle)(1 - \langle q \rangle(1 - r_c))}. \tag{6}$$

To derive M_{cp} we need to identify when R_{cp} is larger or smaller than one. We can do this by manipulating the inequality.

$$(1 - q_p(1 - r_c))(1 - \langle q \rangle(1 - \langle r \rangle)) \geq (1 - q_p(1 - \langle r \rangle)(1 - \langle q \rangle(1 - r_c)). \tag{7}$$

Which simplifies to:

$$q_p(1 - r_c) + \langle q \rangle(1 - \langle r \rangle) \leq q_p(1 - \langle r \rangle) + \langle q \rangle(1 - r_c) \tag{8}$$

leading to the condition:

$$(r_c - \langle r \rangle)(q_p - \langle q \rangle) \geq 0 \tag{9}$$

Since this is an inequality, we need to be careful about the signs of $(q_p - \langle q \rangle)$ and $(r_c - \langle r \rangle)$. Changes in sign flip the inequality operator. So what this condition means is that $R_{cp} \geq 1$ when $r_c \geq \langle r \rangle$ for $q_p - \langle q \rangle \geq 0$ and when $r_c < \langle r \rangle$ for $q_p - \langle q \rangle < 0$. We can also get this condition intuitively by by considering the case when $q_p = \langle q \rangle$ or $r_c = \langle r \rangle$. In these two cases $R_{cp} = 1$, meaning that these lines divide the matrix into regions where the values of R_{cp} are higher or smaller than one. In sum, from the condition above M_{cp} is a matrix divided into four quadrants:

$$\begin{array}{llll} M_{cp} = 1 & \text{if} \quad r_c \geq \langle r \rangle & \& & q_p \geq \langle q \rangle \\ M_{cp} = 0 & \text{if} \quad r_c < \langle r \rangle & \& & q_p \geq \langle q \rangle \\ M_{cp} = 0 & \text{if} \quad r_c \geq \langle r \rangle & \& & q_p < \langle q \rangle \\ M_{cp} = 1 & \text{if} \quad r_c < \langle r \rangle & \& & q_p < \langle q \rangle \end{array} \tag{10}$$

This matrix represents a world in which countries with a high probability of having the capability (r_c higher than average), specialize in products with high probability of requiring the capability (q_p higher than average), and countries with low probability of having the capability specialize in products with low probability of requiring it. This is related to the idea of log super-modularity in trade theory [56].

As an example, consider a world with four countries and six products, where two countries have above average r_c and three products have above average q_p. In this example, the binary specialization matrix M_{cp} takes the form:

$$M_{cp} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{11}$$

Finally, we use M_{cp} to derive $M_{cc'}$. Here we use the standard reciprocal average method or method reflections. This method proposes that the complexity of an economy is the average complexity of the activities that economy is specialized in, and that the complexity of an activity is the average complexity of the economies specialized in that activity. Using the economic complexity index (ECI) and the product complexity index (PCI) to indicate the complexity of economies and activities we obtain:

$$\begin{aligned} ECI_c &= \frac{1}{M_c} \sum_p M_{cp} PCI_p \\ PCI_p &= \frac{1}{M_p} \sum_c M_{cp} ECI_c \end{aligned} \tag{12}$$

putting the second equation into the first one can show that ECI_c is the solution to the following self-consistent equation:

$$ECI_c = \sum_{c'} M_{cc'} ECI_{c'} \tag{13}$$

with

$$M_{cc'} = \frac{1}{M_c} \sum_p \frac{M_{cp} M_{c'p}}{M_p} \tag{14}$$

Meaning that the economic complexity vector ECI_c must be an eigenvector of the $M_{cc'}$ matrix representing the steady state of the mapping defined by the system in eqns. (12) (the same derivation can be used to define the $M_{pp'}$ matrix used to estimate PCI).[10]

Estimating the first eigenvector of $M_{cc'}$ is trivial because $M_{cc'}$ is a stochastic matrix (each row adds to one). That means its first eigenvector will always be **1**. This is easy to prove by summing $M_{cc'}$ over c'.

[10] $M_{cc'}$ can also be defined as the product of four matrices $M_{cc'} = D_c M_{cp} D_p M_{pc'}$ where D_c is a diagonal matrix of $1/M_c$ and D_p is a diagonal matrix of $1/M_p$.

$$
\begin{aligned}
M_{cc'}\mathbf{1} &= \sum_{c'} \frac{1}{M_c} \sum_p \frac{M_{cp}M_{c'p}}{M_p} \\
M_{cc'}\mathbf{1} &= \frac{1}{M_c} \sum_p \frac{M_{cp}M_p}{M_p} \\
M_{cc'}\mathbf{1} &= \frac{1}{M_c} \sum_p M_{cp} = \mathbf{1}
\end{aligned} \tag{15}
$$

Since the first eigenvector is $\mathbf{1}$, the steady state of the system represented by eqns (12) is given by the second eigenvector. To estimate that eigenvector, we need to calculate $M_{cc'}$. Here, we consider three cases. When the number of economies and activities is even, when the number of economies is odd and the number of activities is even, and when both the number of economies and activities are odd. The need to consider these cases separately will become self-evident once they are introduced.

We begin with the simplest case, which is an even number of economies and activities. We let also $\langle r \rangle$ and $\langle q \rangle$ be the medians of their distributions. In that case, $M_{cc'}$ reduces to a block diagonal matrix with two blocks with values of $1/M_p$ (all economies have the same diversity and all activities the same ubiquity). That is:

$$
\begin{aligned}
&M_{cc'} = \frac{1}{M_p} \quad \text{if} \quad r_c \,\&\, r_{c'} \geq \langle r \rangle \quad \text{or} \quad r_c \,\&\, r_{c'} < \langle r \rangle \\
&M_{cc'} = 0 \quad \text{otherwise}
\end{aligned} \tag{16}
$$

For the example above, with four economies and six activities, $M_{cc'}$ takes the form:

$$
M_{cc'} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \tag{17}
$$

Since we know the first eigenvector of this matrix is the vector $e_c^1 = \mathbf{1}$, and since this matrix is symmetric, and has therefore orthogonal eigenvectors, we can use these properties to find the second eigenvector, which is:

$$
e_c^2 = ECI = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \tag{18}
$$

In this case, this eigenvector is also associated with the eigenvalue of one (this matrix is degenerate, meaning that it has more than one eigenvector associated with the same eigenvalue).[11] This eigenvector is easy to verify through multiplication.

What it is important for us is that this eigenvector separates economies with above and below average r, that is:

[11] In this case, all linear combinations of these eigenvectors are eigenvectors themselves. For example the vector $[a, a, b, b]$ is also an eigenvector, since we can construct it as a linear combination of $[1, 1, 1, 1]$ and $[1, 1, -1, -1]$

$$
\begin{aligned}
e_c^2 = ECI_c = 1 \qquad & \text{if} \quad r_c \geq \langle r \rangle \\
e_c^2 = ECI_c = -1 \qquad & \text{if} \quad r_c < \langle r \rangle
\end{aligned}
\tag{19}
$$

showing that in this example the second eigenvector of the $M_{cc'}$ matrix or ECI separate economies that are above or below average in their probability of having the only capability in the model.[12]

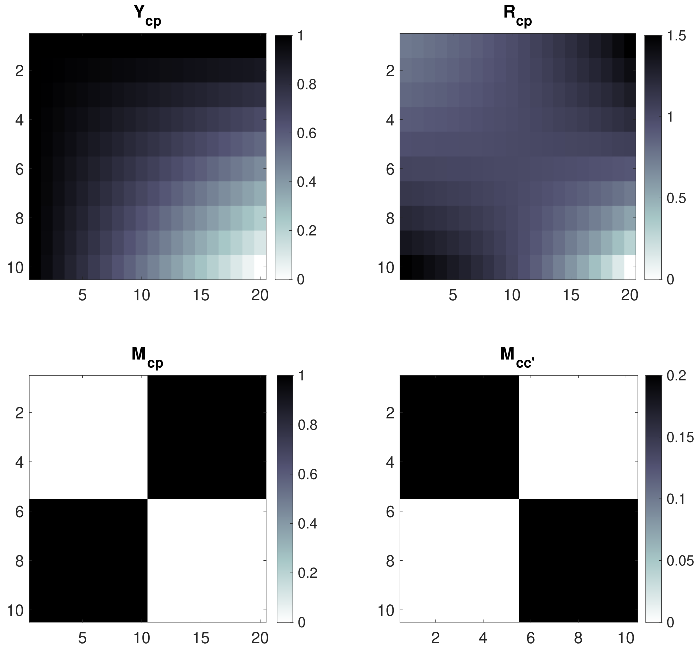

Figure 1: Graphical description of the four matrices involved in the single capability model for 10 countries and 20 products. In *cp* matrices row represents economies (countries) and columns represent activities (products). Rows are sorted from highest r_c to lowest r_c and columns are sorted from lowest q_p to highest q_p. That is, cell $(1, 1)$ is the output of the country with the highest probability of having the capability on the product with the lowest probability of requiring it, and cell $(10, 20)$ is the output of the country with the lowest probability of having a capability in the product with the highest probability of requiring it.

Figure 1 visualizes the matrices in the single capability model for a case involving an even number of economies and activities (10 economies and 20 activities). These graphical representations will help us develop our intuition when interpreting more complex models later.

From top left to bottom right, we start with the Y_{cp} matrix, which the output matrix, the RCA matrix R_{cp} which estimates the specialization of each economy in each activity, the binary specialization matrix M_{cp}, and the stochastic square matrix $M_{cc'}$ from which we derive ECI. The output matrix Y_{cp} shows a nested pattern. This is the tendency for the rows that are less

[12]At this point, it is worth noting a key property of eigenvectors, which is that they have a freedom of sign. That is, if e_c is an eigenvector of a matrix M so is $-e_c$. This is trivial from the fact that if $Me_c = \lambda e_c$ then $M(-e_c) = \lambda(-e_c)$. This means that the eigenvector derivation of ECI separates among economies based on their capability endowments, but is agnostic about which of the two clusters is the high-capability cluster. In the empirical literature, this is solved by iterating the system in eqs. 12 to estimate ECI starting from an initial condition that is correlated with the high-capability cluster (e.g. initializing the system with diversity M_c) and stopping at an even iteration. Other methods to estimate complexity empirical (e.g. [13] also rely on an initial guess).

filled to be subsets of the rows that are more filled. Nestedness is a well-known feature of matrices summarizing the geography of fine-grained economic activities, such as exports by country and product, employment by city and industry, or patents by city and technology [123]. This example shows how the matrix transformations simplify the complexity of the Y_{cp} matrix, reducing it to a couple of clusters with above and below average probability of having a capability. But the high level of symmetry of this case limits our ability to explore more interesting properties of the method, such as the ability to separate capability endowments from simple measures of diversity. For that, we need to consider other cases.

Next, we focus on the case where the number of economies is odd and the number of activities is even (and where the averages of r and q are still their medians). For example, $N_c = 5$ and $N_p = 6$. This example is interesting, because unlike in the previous case where the diversity of economies and the ubiquity of activities was constant, in this case only the ubiquity of activities remains fixed. This example is important because it will teach us about the ability of ECI to recover r_c, even when the most diverse economy is the one that has a probability of having a capability equal to the average ($r_c = \langle r \rangle$) (it is actually specialized in all activities).

In this *odd-even* case, M_{cp} is given by:

$$\begin{aligned} M_{cp} &= 1 \quad \text{if} \quad r_c > \langle r \rangle \quad \& \quad q_p > \langle q \rangle \\ M_{cp} &= 1 \quad \text{if} \quad r_c < \langle r \rangle \quad \& \quad q_p < \langle q \rangle \\ M_{cp} &= 1 \quad \text{if} \quad r_c = \langle r \rangle \\ M_{cp} &= 0 \quad \text{otherwise} \end{aligned} \tag{20}$$

which for $N_c = 5$ and $N_p = 6$ results in the binary specialization matrix that is completely filled on the third row (so the matrix is no longer symmetric):

$$M_{cp} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

Clearly the most diverse economy is the one in the third row, which is specialized in all activities.

Moving to $M_{cc'}$ gives us:

$$\begin{aligned} M_{cc'} &= \frac{1}{M_p} \quad \text{if} \quad c = c' \\ M_{cc'} &= \frac{1}{M_p} \quad \text{if} \quad r_c \,\&\, r_{c'} > \langle r \rangle \quad \text{or} \quad r_c \,\&\, r_{c'} < \langle r \rangle \\ M_{cc'} &= \frac{1}{M_c} \sum_p \frac{M_{cp} M_{c'p}}{M_p} \quad \text{if} \quad r_c \,\&\, r_{c'} = \langle r \rangle \quad \& \quad c \neq c' \\ M_{cc'} &= 0 \quad \text{otherwise} \end{aligned} \tag{22}$$

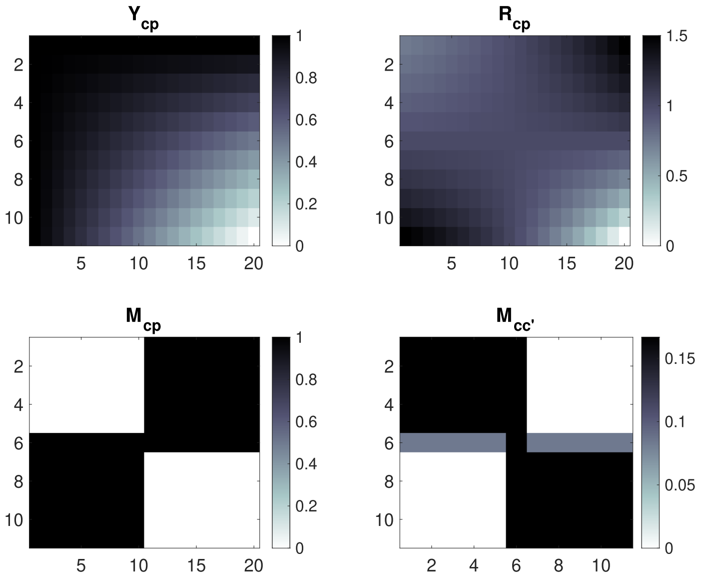

Figure 2: Graphical description of the four matrices involved in the single capability model for 11 economies (e.g. countries) and 20 activities (e.g products). In cp matrices row represents economies (countries) and columns represent activities (products). Rows are sorted from highest r_c to lowest r_c and columns are sorted from lowest q_p to highest q_p. That is, cell $(1, 1)$ is the output of the country with the highest probability of having the capability on the product with the lowest probability of requiring it, and cell $(11, 20)$ is the output of the country with the lowest probability of having a capability in the product with the highest probability of requiring it.

which for five economies and six activities results in the matrix:

$$M_{cc'} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/6 & 1/6 & 1/3 & 1/6 & 1/6 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} \tag{23}$$

This matrix is also quite regular, and has the following second eigenvector which can be verified simply using matrix multiplication:

$$e_c^2 = ECI_c = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix} \tag{24}$$

In more general terms it is given by:

$$\begin{aligned} e_c^2 = ECI_c = 1 \qquad & \text{if} \quad r_c > \langle r \rangle \\ e_c^2 = ECI_c = -1 \qquad & \text{if} \quad r_c < \langle r \rangle \\ e_c^2 = ECI_c = 0 \qquad & \text{if} \quad r_c = \langle r \rangle \end{aligned} \tag{25}$$

This is an interesting result, since it shows that the second eigenvector or ECI is not "fooled by diversity." On the contrary, it is able to recover the fact that the economy that is specialized in all activities has a probability of having a capability that is in between that of the high probability and low probability clusters.

Figure 2 summarizes the matrices in the single capability model for a case involving an odd number of economies and an even number of activities (11 economies and 20 activities). In this case, the key difference is that center row of M_{cp} which extends through all columns of the matrix and results in a small overlap between the two clusters in $M_{cc'}$.

Next, we consider the case in which the number of economies and activities are odd. In this case, the diversity of economies and the ubiquity of activities is no longer constant. Now the M_{cp} matrix has both, one row and one column that are completely filled, which correspond respectively to the economy and activity with $r_c = \langle r\rangle$ and $q_c = \langle q\rangle$. That is:

$$\begin{aligned}
M_{cp} &= 1 \quad \text{if} \quad r_c > \langle r\rangle \quad \& \quad q_p > \langle q\rangle \\
M_{cp} &= 1 \quad \text{if} \quad r_c < \langle r\rangle \quad \& \quad q_p < \langle q\rangle \\
M_{cp} &= 1 \quad \text{if} \quad r_c = \langle r\rangle \\
M_{cp} &= 1 \quad \text{if} \quad q_c = \langle q\rangle \\
M_{cp} &= 0 \quad \text{otherwise}
\end{aligned} \tag{26}$$

Which we can bring to an example with five economies and seven activities:

$$M_{cp} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{27}$$

In this case $M_{cc'}$ will have a more complex form which we can express by noticing that the diversity and ubiquity of the economy and activity in the middle row and column of M_{cp} is the number of economies N_c and the number of activities N_p. Since all other economies and activities have the same diversity and ubiquity, which we will denote by M_c and M_p, we obtain:

$$\begin{aligned}
M_{cc'} &= \frac{1}{M_c}(1 + \frac{1}{N_p}) \quad \text{if} \quad r_c \,\&\, r_{c'} > \langle r\rangle \quad \text{or} \quad r_c \,\&\, r_{c'} < \langle r\rangle \\
M_{cc'} &= \frac{1}{M_c}(\frac{1}{N_p}) \quad \text{if} \quad r_c > \langle r\rangle \quad \& \quad r_{c'} < \langle r\rangle \text{ and vice versa} \\
M_{cc'} &= \frac{1}{N_c}(\frac{1}{N_p} + \frac{N_c - 1}{M_p}) \quad \text{if} \quad c = c' \quad \& \quad r_c = \langle r\rangle \\
M_{cc'} &= \frac{1}{N_c}(1 + \frac{1}{N_p}) \quad \text{if} \quad c \neq c' \quad \& \quad r_c = \langle r\rangle \\
M_{cc'} &= \frac{1}{M_c}(1 + \frac{1}{N_p}) \quad \text{if} \quad r_{c'} = \langle r\rangle
\end{aligned} \tag{28}$$

Which might be easier to parse when presented in matrix form:

$$M_{cc'} = \begin{bmatrix} \frac{1}{M_c}(\frac{N_p+1}{N_p}) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) & \dots & \frac{1}{M_c}(\frac{1}{N_p}) \\ \frac{1}{M_c}(\frac{N_p+1}{N_p})) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) & \dots & \frac{1}{M_c}(\frac{1}{N_p}) \\ \frac{1}{N_c}(1+\frac{1}{N_p}) & \dots & \frac{1}{N_c}(\frac{1}{N_p}+\frac{N_c-1}{M_p}) & \dots & \frac{1}{N_c}(1+\frac{1}{N_p}) \\ \frac{1}{M_c}(\frac{1}{N_p}) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) \\ \frac{1}{M_c}(\frac{1}{N_p}) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) & \dots & \frac{1}{M_c}(\frac{N_p+1}{N_p}) \end{bmatrix} \tag{29}$$

Bringing this the five economies and seven activities example gives us:

$$M_{cc'} = \begin{bmatrix} 3/10 & 3/10 & 3/10 & 1/20 & 1/20 \\ 3/10 & 3/10 & 3/10 & 1/20 & 1/20 \\ 6/35 & 6/35 & 11/35 & 6/35 & 6/35 \\ 1/20 & 1/20 & 3/10 & 3/10 & 3/10 \\ 1/20 & 1/20 & 3/10 & 3/10 & 3/10 \end{bmatrix} \tag{30}$$

Which again has a second eigenvector of the form:

$$e_c^2 = ECI_c = \begin{bmatrix} a \\ a \\ 0 \\ -a \\ -a \end{bmatrix} \tag{31}$$

This is easy to verify through multiplication. Since the vector adds all of the elements up to the center column and then subtracts all of the elements after the central column, and since the number of elements before and after the central column are the same, we can simply subtract the first and last element of the first row of matrix (eqn. 29) to obtain:

$$M_{cp}v_c = \frac{a}{M_c}(1+\frac{1}{M_p}) - \frac{a}{M_c}(\frac{1}{M_p}) = \frac{a}{M_c} \tag{32}$$

Doing the same operation on the last row we get:

$$M_{cp}v_c = \frac{a}{M_c}(\frac{1}{M_p}) - \frac{a}{M_c}(1+\frac{1}{M_p}) = -\frac{a}{M_c} \tag{33}$$

Since in the central row of the matrix all elements, except the one in the diagonal, are the same, this vector sends that row to zero. Thus, up to a normalization constant, the second eigenvector of $M_{cc'}$ is given by:

$$\begin{aligned} e_c^2 = ECI_c = a \quad & \text{if} \quad r_c > \langle r \rangle \\ e_c^2 = ECI_c = -a \quad & \text{if} \quad r_c < \langle r \rangle \\ e_c^2 = ECI_c = 0 \quad & \text{if} \quad r_c = \langle r \rangle \end{aligned} \tag{34}$$

Figure 3 presents these matrices in graphical form. We would like to notice two things about this version of the single capability model. The first one is that in this case $M_{cc'}$ no longer has blocks of 0s. The second one is that this is also an example in which the highest diversity economy (the 8^{th} row in M_{cp} is correctly identified as not being the economy with the highest

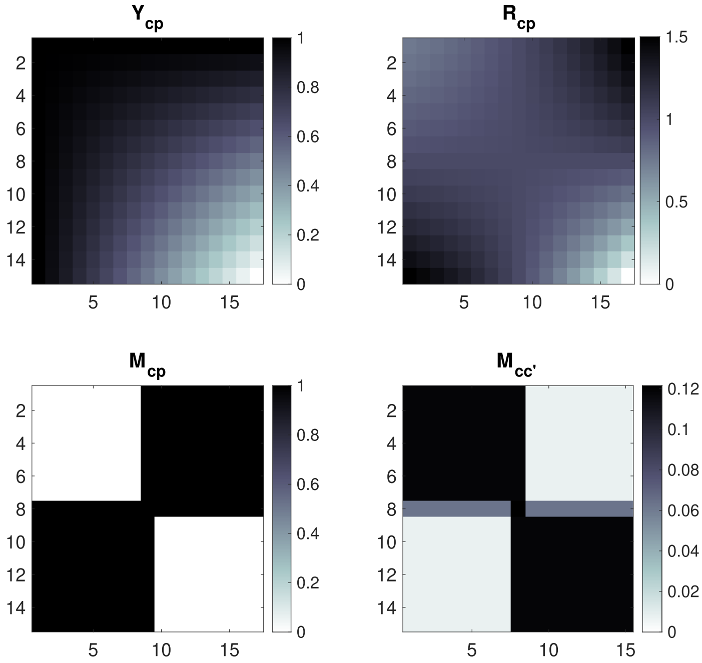

Figure 3: Graphical description of the four matrices involved in the single capability model for 15 economies (e.g. countries) and 17 activities (e.g products). In cp matrices row represents economies (countries) and columns represent activities (products). Rows are sorted from highest r_c to lowest r_c and columns are sorted from lowest q_p to highest q_p. That is, cell $(1,1)$ is the output of the economy or country with the highest probability of having the capability on the product or activity with the lowest probability of requiring it, and cell $(11,20)$ is the output of the economy with the lowest probability of having a capability in the activity or product with the highest probability of requiring it.

probability of having the capability.

Thus, we have shown that, in the context of the single capability or single factor model, the second eigenvector of the $M_{cc'}$ matrix, known as the economic complexity index or ECI, separates economies among those that have a higher and lower than average probability of having the single capability in the model.

In the next section we will use these figures to explore more complex forms of these model, involving multiple capabilities. We will then mode to different production functions to explore the generalizability of this result.

3 The Multi Capability Model

The multi capability version of the combinatorial model can be defined by letting the probability that a country has capability b be $r_{c,b}$ and the probability that a product requires a capability b be $q_{p,b}$. For a country to produce a product it needs to have all of the capabilities that the product requires. That is, the product of these probabilities for all of the capabilities in the model. Mathematically, that translates into an output matrix of the form:

$$Y_{cp} = A \prod_{b=1}^{N_b} (1 - q_{p,b}(1 - r_{c,b})) \tag{35}$$

To avoid over-parameterizing the model too early, and to simplify our exploration, we will begin by discussing the case in which these probabilities are independent of the capability and of each other. That is:

$$Y_{cp} = A \prod_{b=1}^{N_b} (1 - q_p(1 - r_c)) \tag{36}$$

Which reduces to a well-known binomial form[13]:

$$Y_{cp} = (1 - q_p(1 - r_c))^{N_b} \tag{38}$$

This form assumes that a country has the same probability of having each of the different capabilities required by a product. The need for multiple capabilities, therefore, enters only in the probability of missing one of them, making this similar in sprit to Kremer's O-Ring model [67]. In fact, Kremer's O-Ring production function can be recovered from eqn.(35) by setting $q_{p,b} = 1$ for all activities and $r_{c,b} = r_b$ for all capabilities (called tasks in the O-Ring model).[14]

We will explore this model is by using the same matrices we derived analytically for the single capability model. Figure 4 shows these matrices for a model involving ten capabilities, one hundred economies, and one thousand activities. The number of economies and activities gets to a scale and granularity that is similar to the one used in empirical economic complexity studies.

In this example, economies and activities are modeled using evenly spaced probabilities in the $[0, 1]$ interval. That is, for an eleven economy model the probabilities would be given by $0, 0.1, 0.2, \ldots, 0.9, 1$. The result is a highly nested output matrix Y_{cp} and strongly off-diagonal specialization matrices (R_{cp} and M_{cp}).

It is worth noting that the more diverse economies in this model are not the ones with the highest r_c, but the ones with an r_c that is below the largest (around 0.8). This is because the reduced output of these economies in the most demanding activities (the ones with highest q_p) means they are relatively more specialized in products with lower q_ps compared to the economies with the highest r_cs. This effect is analogous to what we saw in the one capability model when we considered an odd number of economies.

Figure 4 also shows that $M_{cc'}$ follows a similar block diagonal structure than before, but much smoother than in the single capability model.

While it would certainly be substantially more difficult to estimate the eigenvectors of this model analytically, we can still explore them numerically. Figure 5 compares the r_c of each economy with its second eigenvector of the $M_{cc'}$ matrix (the non-normalized ECI), diversity

[13] While this form looks relatively simple, even the solution for $N_b = 2$ can result in a mathematical form that is substantially more complicated than the one for the single-capability model. In fact, after some algebra one can show that the condition for $R_{cp} \geq 1$ in the $N_b = 2$ case is:

$$\begin{aligned} &[q_p^2 - \langle q^2\rangle][r_c^2 - \langle r^2\rangle + 2(r_c - \langle r\rangle)] + 2[q_p - \langle q\rangle][r_c - \langle r\rangle] + \\ &2[q_p\langle q^2\rangle - \langle q\rangle q_p^2][r_c - \langle r\rangle r_c^2\langle r\rangle - \langle r^2\rangle r_c + \langle r^2\rangle - r_c^2] \geq 0. \end{aligned} \tag{37}$$

[14] In that case, the production function reduces to $Y = A \prod_{b=1}^{N_b} r_b$.

(M_c), and the ranking of economies according to ECI. Unlike in the single capability model, where ECI told us only if an economy was above or below average, in this example we get a less discrete second eigenvector that increases monotonically with r. This results in a perfect correlation between the ranked values of r and ECI. Diversity, however, peaks for economies with r_c less than the maximum, meaning that it is a non-ideal way to estimate the capability endowment of economies in this model. That is, we recover the fact that the second eigenvector of $M_{cc'}$–the economic complexity index (ECI)–is a good method to estimate the key parameter for the economies in the model (r_c). This validates the idea that ECI is a good way to recover the relative value of r for a country in a multi-capability model, and that is is therefore, and estimate of the complexity of an economy (an estimate of the economy being endowed with multiple complementary capabilities).

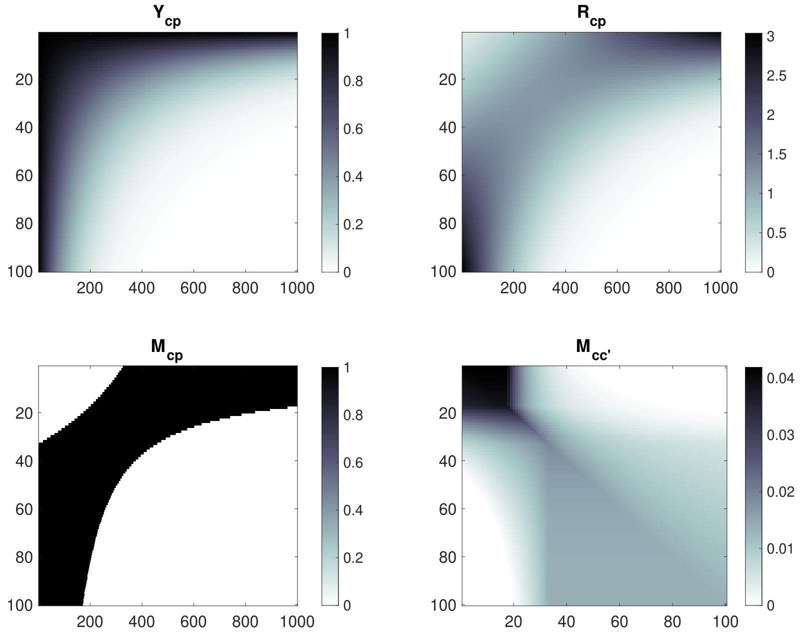

Figure 4: The four matrices involved in economic complexity calculations using a ten capability model for 100 countries and 1,000 products. In *cp* matrices row represents economies (countries) and columns represent activities (products). Rows are sorted from highest r_c to lowest r_c and columns are sorted from lowest q_p to highest q_p.

Figure 6 illustrates the behavior of this model for different number of capabilities (from 2 to 60). Overall, the behavior observed is consistent with the one observed for the ten capability example. Across the board, ECI behaves as a perfect estimator of the probability that a country is endowed with a capability. We can observe, however, that diversity improves as an indicator for the models with the highest number of capabilities (60), becoming almost perfectly monotonic in that case.

To continue our exploration we relax our assumptions about the distributions of r_c and q_p. So far, our simulations have involved evenly spaced r_cs and q_ps in the $[0, 1]$ interval, which means we have been using an idealized uniform distribution. So we replace these uniform distributions for Gaussians by drawing a random numbers from a normal distribution for each r_c and q_p and min-max normalizing these random numbers to ensure they fall in the $[0, 1]$ interval.

Figures 7 and 8 show the results of this exercise. Unlike in the previous example, the specialization matrix M_{cp} exhibits a bit more "roughness", with non-perfectly smooth edges. That said, the behavior of this model is otherwise quite similar to the previous one. $M_{cc'}$ is

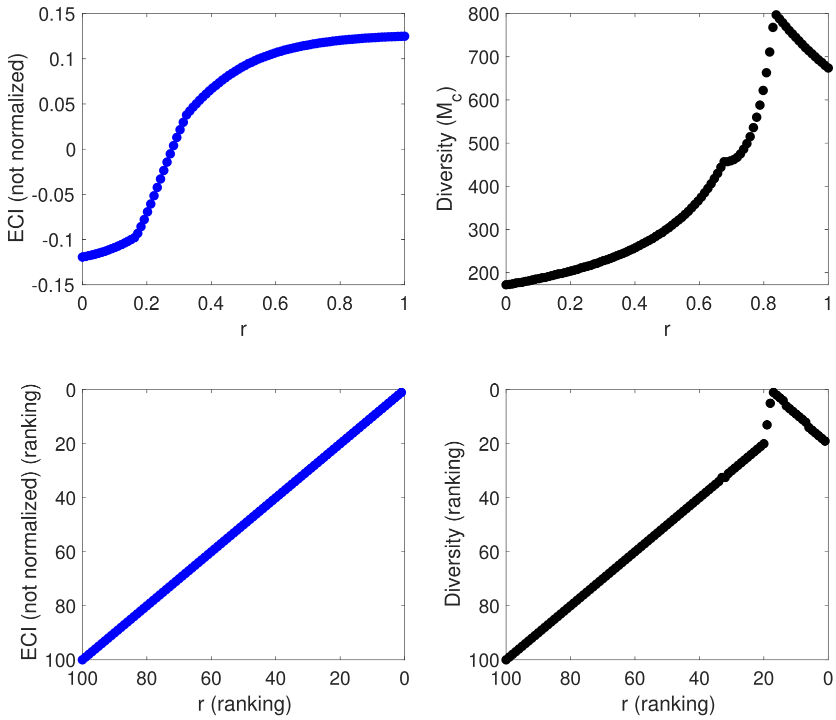

Figure 5: Comparison between the key parameters representing economies in the model (r), the second eigenvector of the $M_{cc'}$ matrix (ECI), and the diversity (M_c) of economies in the model. The top two panels show the raw relationship between the variables while the bottom two compare their rankings.

roughly block diagonal and the second eigenvector of $M_{cc'}$ or ECI almost perfectly captures r_c, as seen in its monotonic relationship with r and in their rank correlation (Figure 8), whereas diversity peaks for economies with an r_c of around 3/4, making it a non-ideal estimator of r_c.

Now that we have developed our intuition around these versions of the multi-capability version of the model (equation 35), we consider the case in which the probability that an economy is endowed with a capability, and that an activity requires one, is not equal across all capabilities. That is, we consider the case where:

$$r_c \to r_{c,b} \tag{39}$$
$$q_p \to q_{p,b} \tag{40}$$

We consider this case using the following parametrization:

$$r_{c,b} = \alpha r_c + (1 - \alpha)\text{random}(0,1) \tag{41}$$
$$q_{p,b} = \alpha q_p + (1 - \alpha)\text{random}(0,1) \tag{42}$$

That is, we set a baseline level for the probability that an economy has a capability or an activity requires one, and mix that with a random number according to the proportions α and $1 - \alpha$. When $\alpha = 1$ the probability that an economy is endowed with a capability is the same for all economies and we recover our previous case. When $\alpha = 0$ the capability endowments are fully random.

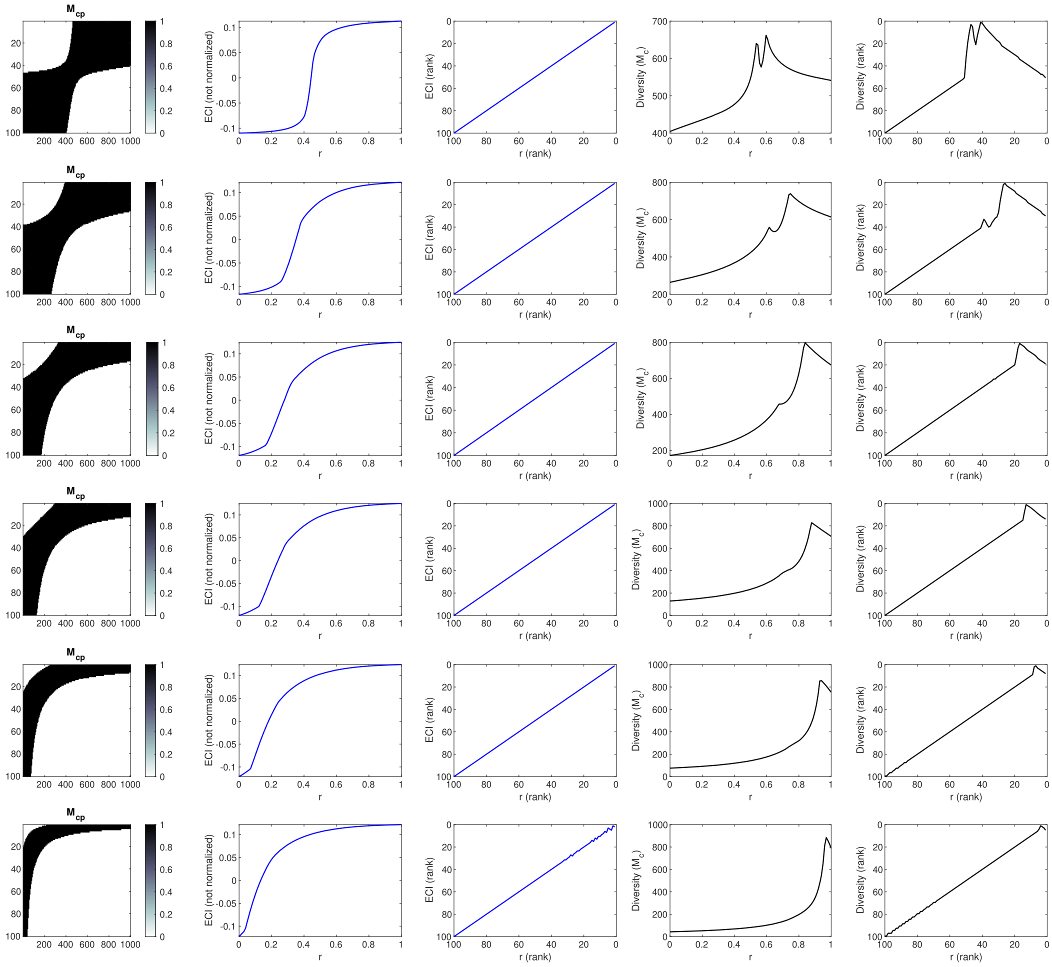

Figure 6: Comparison for the binary specialization matrix and the correlation between complexity, diversity, and the probability that a country is endowed with a capability r for models using $2, 5, 10, 15, 30$, and 60 capabilities

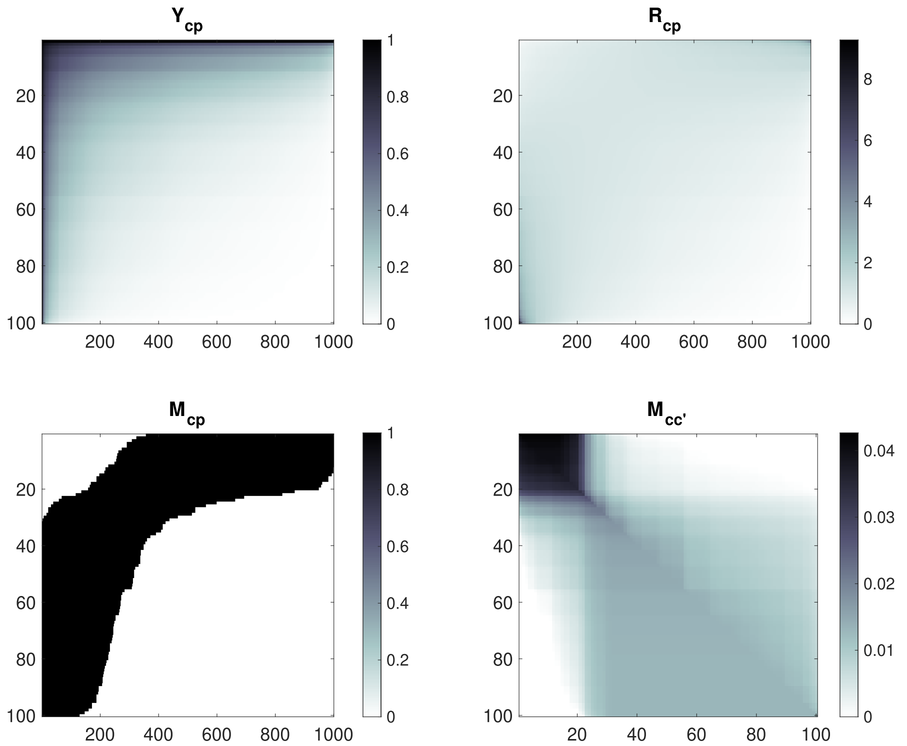

Figure 7: The four matrices involved in economic complexity calculations using a ten capability model for 100 economies and 1000 activities using randomly assigned r_cs and q_ps according to a normal distribution. In *cp* matrices row represents economies (countries) and columns represent activities (products). Rows are sorted from highest r_c to lowest r_c and columns are sorted from lowest q_p to highest q_p.

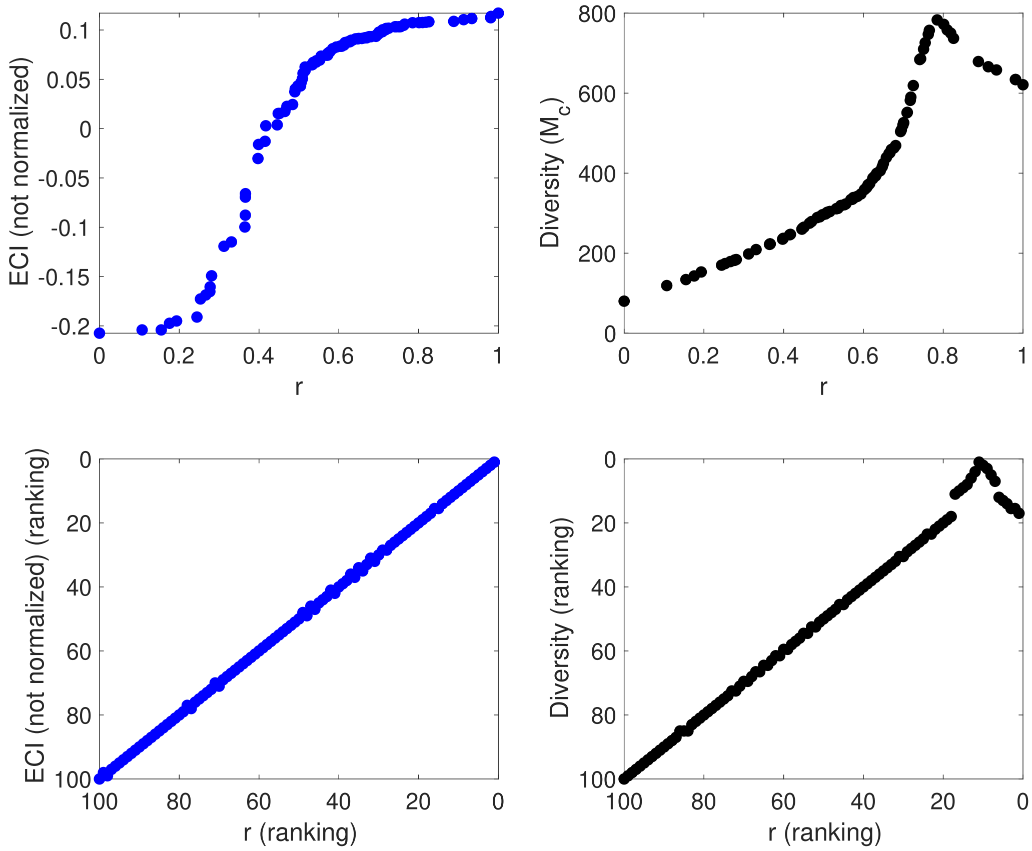

Figure 8: Comparison between the key parameter representing economies in the model (r), the second eigenvector of the $M_{cc'}$ matrix (ECI), and the diversity (M_c) of economies in a model involving 100 economies, 1000 activities, and 10 capabilities. The top two figures show the relationship between the raw variables and the bottom two show that relationship in rankings.

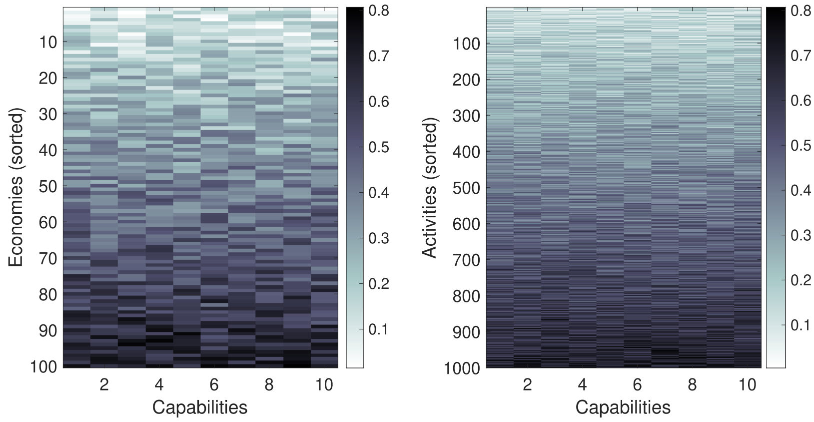

Figure 9: Parametrization of $r_{c,b}$ and $q_{p,b}$ for ten capabilities in a model where the probability that an economy is endowed with a capability, or that an activity requires it, is 3/4 of a linearly spaced baseline in the $[0, 1]$ interval and 1/4 random.

In this case, our goal is to explore whether ECI is able to recover the average capability endowment of an economy. That is, we will compare ECI with $\langle r \rangle_b$, where:

$$\langle r \rangle_c = \frac{1}{N_b} \sum_b r_{c,b} \tag{43}$$

Figure 9 provides an illustration of this parametrization for the case when the probability that an economy is endowed with a capability, or that an activity requires it, is 3/4 of a linearly spaced baseline in the $[0, 1]$ interval and 1/4 random. The matrices resulting from this model are shown in Figure 10.

We can see that despite introducing substantial variation into the capability endowment of economies the matrices retain a similar shape. In fact, we find that in this case ECI still does a good job as an estimator of the average capability endowment of an economy $\langle r \rangle_c$ as we can see in Figure 11. This means that in the context of a model with multiple capabilities we can interpret ECI as an estimate of the average capability endowment of an economy.

But how far can we take this intuition? Does this method work for completely random capability endowments? Or does it require an adequate level of correlation between the different capabilities?

We can explore this question by using the parametrization introduced in equation 42 to vary the level of randomness in capability endowments. Figure 12 performs this exploration, by showing the capability endowment matrices, M_{cp}, and the correlation between the ranks of ECI and the average capability endowment of an economy $\langle r \rangle_c$ for $\alpha = [0.9, 0.75, 0.6, 0.45, 0.3, 0.15]$. This exercise reveals that the method is rather robust, and is able to capture the average capability endowment for an economy even when the endowment is 60 percent random and 40 percent based on a baseline. This exercise also shows that the relationship between ECI and $\langle r \rangle_c$ breaks somewhere between $\alpha = 0.45$ and $\alpha = 0.3$, suggesting a potential phase transition in this behavior.

Figure 13 explores this phase transition by presenting the average correlation between ECI and $\langle r \rangle_c$ observed after sweeping through the parametrization parameter α 250 times using a linearly space grid of 50 points for the interval $\alpha = [0.01, 1]$. We can see that there is a phase transition around 0.35, meaning that the ability of ECI to recover the average capability endowment of an economy in this model ($\langle r \rangle_c$) is valid as long as there is a strong

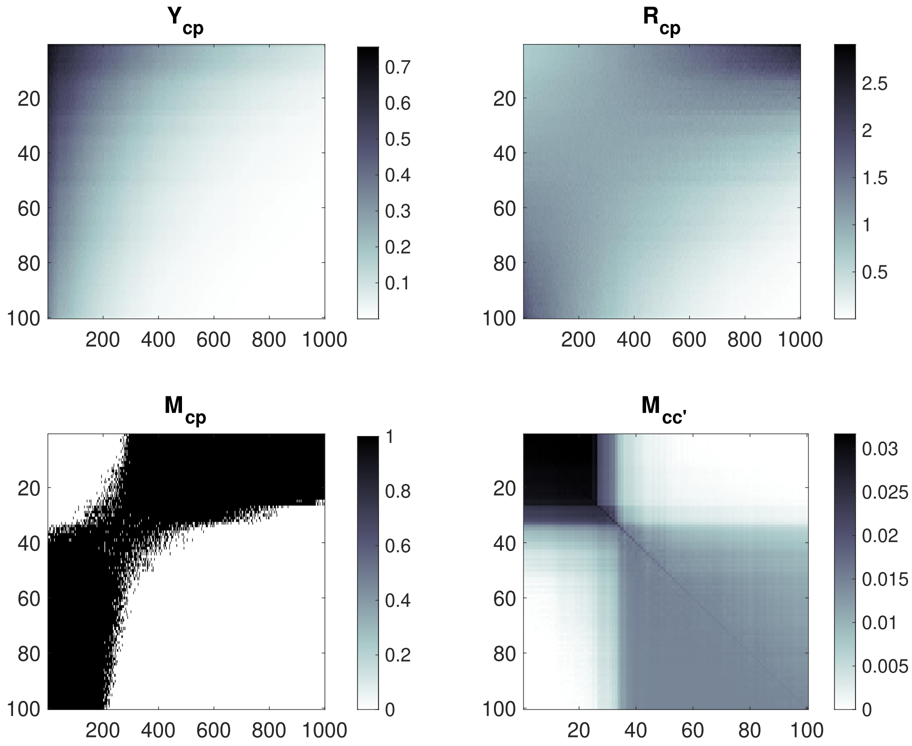

Figure 10: Matrices for a 10 capability, 100 economies, and 1000 activities model, where the probability that an economy is endowed with a capability, or that an activity requires it, is 3/4 of a linearly spaced baseline in the $[0, 1]$ interval and 1/4 random.

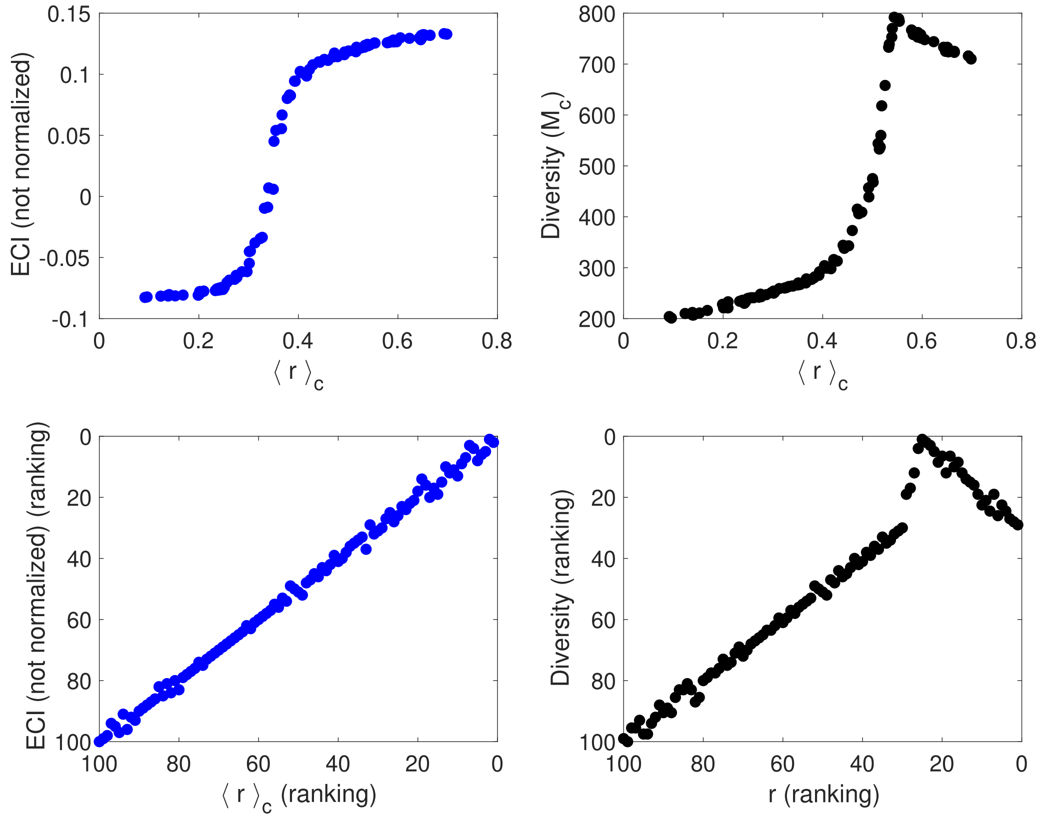

Figure 11: Relationship between the the average capability endowment of an economy $\langle r \rangle_c$, ECI, diversity, and the rankings of ECI and diversity for a model where the probability that an economy is endowed with a capability, or that an activity requires it, is 3/4 of a linearly spaced baseline in the $[0, 1]$ interval and 1/4 random.

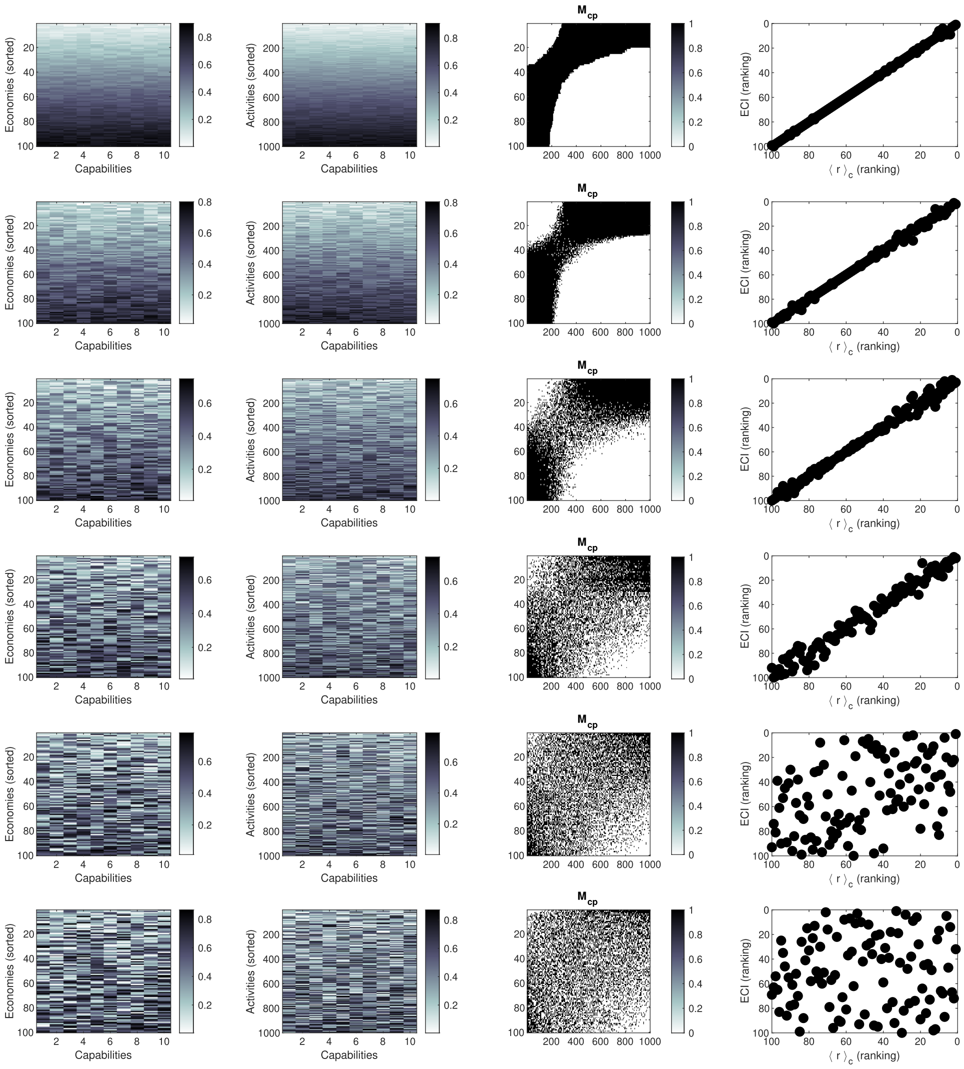

Figure 12: Numerical implementation of the multi-capability model for 100 economies, 1000 activities, and 10 capabilities. The probabilities that economies are endowed with a capability, or that activities require them, follows the parametrization in equation (42). Each row of this figure represents a different level of mixing between a baseline probability and a uniform random number. From top to bottom, the weight of the baseline α are 0.9, 0.75, 0.6, 0.45, 0.3, and 0.15.

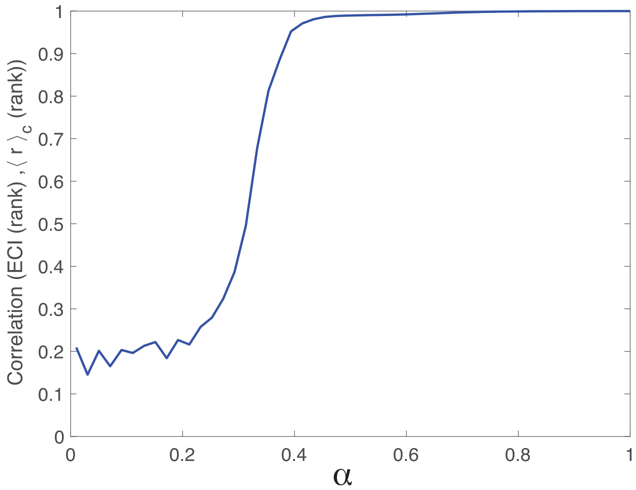

Figure 13: Average correlation between ECI and the average capability endowment of an economy $\langle r \rangle_c$ as a function of the mixing probability α. We can see that the correlation between ECI and $\langle r \rangle_c$ remains close to one for mixing probabilities above 0.4.

enough correlation among the probabilities that an economy is endowed with different capabilities.

Overall, despite the added complexity of the multi-capability model, and the added variation of using randomly drawn probabilities for r_c and q_p, the behavior of the second eigenvector or ECI, and the shapes of the matrices leading to its calculation, are largely consistent with the intuition we developed in the single capability model. That extends the robustness of this idea to models with a wide range of capabilities, including models with substantial levels of noise on how those capabilities are assigned to economies.

But are these observations particular to models based on capabilities and probabilities? Or can we use the second eigenvector method of economic complexity to recover factors in models based on other production functions?

In the next section we explore extensions of this method to other production functions to delineate the effective boundaries of this theory.

4 More Production Functions

You may now be wondering if the ability of the second eigenvector method to recover the key parameters characterizing an economy are a more general property that applies to a wide family of production functions. Is the second eigenvector or ECI method something that works only for stochastic models of capabilities? Or is it a more general idea that works also for a wide range of production functions? If so, what are the characteristics that a production function needs to satisfy to fall within the scope of this theory?

We begin by considering a production function that won't work and that can teach us a valuable lesson about those that do. This is a relative factor intensity Cobb-Douglas type production function of the form:

$$Y_{cp} = A(K_c/K_p)^\gamma \tag{44}$$

The problem with eqn. (44) is that in this model all economies have a comparative advantage equal to one in all activities. In fact, the idea that the output of an economy is perfectly proportional to a power of its factor endowment means that there cannot be any visible specialization (at least not visible using Balassa's (1965) revealed comparative advantage indicator). This is easy to prove using the formula for R_{cp}.

$$R_{cp} = \frac{(K_c/K_p)^\gamma \sum_{c,p}(K_c/K_p)^\gamma}{\sum_c (K_c/K_p)^\gamma \sum_p (K_c/K_p)^\gamma} \tag{45}$$

which after some manipulation becomes

$$R_{cp} = \frac{(K_c/K_p)^\gamma \sum_c K_c^\gamma \sum_p (1/K_p)^\gamma}{(K_c/K_p)^\gamma \sum_c K_c^\gamma \sum_p (1/K_p)^\gamma} = 1. \tag{46}$$

In fact, we can extend this property to to all separable functions of the form:

$$Y_{cp} = Af(K_c)g(K_p) \tag{47}$$

using this same exact calculation.

This gives us a hint of what was special of about the capability model. What made the capability model work was not that we were working with probabilities and the concept of capabilities, but that we were working with a non-multiplicative-separable function (something of the form $A + f_c g_p$). So next, we explore a shifted version of the Cobb-Douglas factor intensity production function. This involves breaking the symmetry of the separability by including an additive term B which we can interpret as a baseline cost when it is negative and a baseline level of production when positive. We can describe this function as:

$$Y_{cp} = B + f_c g_p \tag{48}$$

Where f_c is a function describing the factor endowment of an economy and g_p is a function describing the factor intensity requirements of an activity. Applying the revealed comparative advantage formula to this shifted production function we get:

$$R_{cp} = \frac{(B + f_c g_p) \sum_{c,p}(B + f_c g_p)}{\sum_c (B + f_c g_p) \sum_p (B + f_c g_p)} \tag{49}$$

Bringing this to an inequality in which $R_{cp} \geq 1$ and doing some algebra will lead us to the condition:

$$(f_c - \langle f \rangle)(g_p - \langle g \rangle) \geq 0 \tag{50}$$

which means:

$$\begin{aligned} M_{cp} &= 1 \quad \text{if} \quad f_c \geq \langle f \rangle \quad \& \quad g_p \geq \langle g \rangle \\ M_{cp} &= 1 \quad \text{if} \quad f_c < \langle f \rangle \quad \& \quad g_p < \langle g \rangle \\ M_{cp} &= 0 \quad \text{otherwise} \end{aligned} \tag{51}$$

Meaning that we have recovered the binary specialization matrix of the single capability model.

At this point, it is important to note one more peculiarity of the Cobb-Douglas factor intensity function that can teach us a lesson. Note that equation (50) is expressed in terms of the functions, not the factors. This is important because it means that the slopes of these function come into play. The Cobb-Douglas factor intensity model in equation (44) has opposite derivatives for the factor related to economies and the factor related to activities. That is:

$$\frac{dY_{cp}}{dK_c} > 0 \quad \& \quad \frac{dY_{cp}}{dK_p} < 0 \tag{52}$$

Assuming $\gamma > 0$. This means that when K_p is large g_p will be smaller than average $g_p < \langle g \rangle$. That means economies with a high factor endowment will specialize in activities with low factor intensity requirements. That will make M_{cp} block-diagonal. Yet, even though this makes this model economically unreasonable, it does not change the ability of the second eigenvector to separate among these two clusters.

Zooming out, there are three reasons that make the condition in equation (50) interesting. First, it tells us that the single capability model results are valid for any production functions of the form $Y_{cp} = B + f_c g_p$. Second, working through the algebra tells us that this comes from the symmetry break introduced by adding the shifting term (B in this case), which makes the function non multiplicative-separable, and hence, the specialization of economies in activities not perfectly proportional to their factor endowments. And third, since the single capability model divides the world into two clusters, the more continuous eigenvectors we observe in the empirical literature, as well as the specialization matrices (e.g. M_{cp}) can be taken as evidence of a more complex model, or at least, a model with multiple factors.

5 Prices, Wages, and Consumption

We conclude our theoretical exploration by considering an extension of the single-capability model to a short-run equilibrium framework, with variable prices, wages, and consumption. We let the output of an economy in an activity depend explicitly on the price of each activity π_p by generalizing our output function to:

$$Y_{cp} = \pi_p(1 - q_p(1 - r_c)) = \pi_p y_{cp} \tag{53}$$

We use this function to explore a few things. First, we derive a simple relationship between capability endowments and wages. Then, we derive a new condition from the specialization matrix R_{cp}, which is the key condition connecting the empirical economic complexity estimate ECI with the model's capability endowment. Finally, we estimate product prices by exploring an extension of the model where economies maximize their utility of consumption constrained by their income and the global supply of goods.

First, we focus on wages.

In a perfectly competitive market where labor is the only factor, and all income goes into wages, then the total income of an economy Y_c must equal the wages w_c it pays times the amount of labor L_c it employs. That is:

$$Y_c = w_c L_c \tag{54}$$

Which means:

$$w_c = \frac{\sum_p \pi_p (1 - q_p(1 - r_c))}{L_c} \tag{55}$$

dividing the numerator and denominator by $1/N_p$ (one over the total number of activities) we can transform the sums into averages to obtain an equilibrium wage w_c^*:

$$w_c^* = \frac{N_p(\langle \pi \rangle + \langle q\pi \rangle (r_c - 1))}{L_c} \tag{56}$$

which means that wages are proportional to the probability an economy is endowed with a capability r_c–which we can interpret as a measure of human capital, knowledge, or skill in that capability. In fact, wages grow in proportion to the product of prices times the probability an activity requires a capability and are inversely proportional to population:

$$\frac{dw_c^*}{dr_c} = \frac{N_p \langle q\pi \rangle}{L_c} \tag{57}$$

This finding is consistent with the notion that economic complexity, which we now understand as an estimate of r_c, implies an equilibrium level of wages for an economy, and thus, explains future economic growth. In this model, economies must have a wage given by eqn. 56 in equilibrium. When out of equilibrium, economies should adjust (to first order) according to:

$$\frac{dw_c}{dt} \propto -\eta(w_c - w_c^*) \tag{58}$$

where η is some proportionality constant (e.g. a speed or rate of adjustment). Economies with wages larger than equilibrium experience a downward pressure, whereas those with wages lower than equilibrium experience an upward pressure on their incomes.

Next, we calculate R_{cp} to determine the condition separating the two specialization clusters that are key to determining economic complexity. Going back to the definition of R_{cp} implies the condition :

$$R_{cp} = \frac{\pi_p(1 - q_p(1 - r_c)) \sum_{cp} \pi_p(1 - q_p(1 - r_c))}{\sum_c \pi_p(1 - q_p(1 - r_c)) \sum_p \pi_p(1 - q_p(1 - r_c))} \geq 1 \tag{59}$$

which after some algebra results in the inequality:

$$(r_c - \langle r \rangle)(q_p \langle \pi \rangle - \langle q\pi \rangle) \geq 0 \tag{60}$$

That brings us again to a specialization condition based on two clusters where economies with an above average probability of being endowed with the capability ($r_c > \langle r \rangle$) are specialized in products with a higher probability of requiring the capability, and where those with a below average probability of being endowed with the capability (($r_c < \langle r \rangle$))specialize in less demanding products. Yet, the threshold for activities is now:

$$q_p \geq \frac{\langle q\pi \rangle}{\langle \pi \rangle} \tag{61}$$

which using the standard covariance identity

$$\langle q\pi \rangle = \langle q \rangle \langle \pi \rangle + \text{cov}(q, \pi) \tag{62}$$

yields:

$$q_p \geq \langle q \rangle + \frac{\text{cov}(q, \pi)}{\langle \pi \rangle}, \tag{63}$$

this means that we recover the naked single-capability model when prices are uncorrelated with the probability that an activity requires a capability (when $\text{cov}(q, \pi) = 0$). This equation also tells us that the specialization of high complexity economies in demanding activities is more pronounced when there is a positive correlation between the price of an activity and the probability it requires the capability in the model (which is a reasonable assumption). That is, in a world where prices are higher for more demanding activities, high complexity economies will specialize in a more narrow set of complex activities. Yet, for the purposes of this paper, what is important is that the specialization matrix is still divided into two clusters, just like in the single-capability model with no prices, and that these clusters separate among economies with high and low capability endowments.

Finally, we explore an extension of this model including a demand side, by assuming a logarithmic utility function. That is, we let the utility of economy c be given by:

$$U_c = \sum_p B_{cp} \log(C_{cp}) \tag{64}$$

We also assume that consumption is limited by the budget constraint:

$$\sum_p \pi_p C_{cp} \leq Y_c \tag{65}$$

which means that economies consumption is limited by the revenue generated by their total output. We also assume that the global production of goods is limited by the availability of capabilities, thus:

$$\sum_p C_{cp} = y_c \tag{66}$$

This means that in this model production capacity is fixed, and what adjusts is the price of an activity based on how demanding it is and how preferences for that activity are distributed across economies.

We start by maximizing utility following the Lagrangian:

$$\mathcal{L} = \sum_p B_{cp} \log(C_{cp}) - \lambda(\sum_p \pi_p C_{cp} - Y_c) \tag{67}$$

differentiating against consumption C_{cp} and equating to zero we obtain the condition:

$$C_{cp} = \frac{B_{cp}}{\lambda \pi_p} \tag{68}$$

And using the budget constraint equation (which we use here as an equality) we can solve for λ:

$$\sum_p \pi_p \frac{B_{cp}}{\lambda \pi_p} = Y_c \quad \rightarrow \lambda = \frac{\sum_p B_{cp}}{Y_c} \tag{69}$$

meaning that consumption is given by:

$$C_{cp} = \frac{B_{cp} Y_c}{\pi_p \sum_{p'} B_{cp'}} \tag{70}$$

Finally, since:

$$Y_c = N_p(\langle \pi \rangle - (1 - r_c)\langle q\pi \rangle) \tag{71}$$

then, consumption is given by:

$$C_{cp} = \frac{B_{cp} N_p(\langle \pi \rangle - (1 - r_c)\langle q\pi \rangle)}{\pi_p \sum_p B_{cp}} \tag{72}$$

moving the N_p to the denominator allows us to transform the remaining sum into an average:

$$C_{cp} = \frac{B_{cp}(\langle \pi \rangle - (1 - r_c)\langle q\pi \rangle)}{\pi_p \langle B_c \rangle} \tag{73}$$

which means that consumption is downward slopping with the price of a good (π_p appears only in the denominator [15]) and grows with an economy's preference for a specific activity (B_{cp}) and its capability endowment (r_c).

Now, we estimate prices by using the market clearing condition:

$$\sum_c C_{cp} = y_p \tag{74}$$

and since:

$$y_p = N_c(1 - q_p(1 - \langle r \rangle)) \tag{75}$$

then:

$$\sum_c \frac{B_{cp}(\langle \pi \rangle - (1 - r_c)\langle q\pi \rangle)}{\pi_p \langle B_c \rangle} = N_c(1 - q_p(1 - \langle r \rangle)) \tag{76}$$

which after some algebra can be brought to the form[16]:

[15] π_p also appears implicitly in the average $\langle \pi \rangle$ and $\langle q\pi \rangle$ but its contribution is much smaller (divided by $1/N_p$). Also, the average can be thought of as a common price level, since it is the same for all products p

[16] here we used the notion that averages are constants to arrive to an expression where π_p is expressed as a function of its ensemble averages.

$$\pi_p = \frac{\sum_c \frac{B_{cp}}{\langle B_c \rangle}(\langle \pi \rangle - \langle q\pi \rangle(1 - r_c))}{N_c(1 - q_p(1 - \langle r \rangle))} \tag{77}$$

which means the price of activity p grows with the probability it requires the capability (q_p), since the denominator is the smallest it can be when $q_p = 1$ and it is the maximized for $q_p = 0$. Prices also grow when high capability economies (high r_c and hence high-income Y_c and high-wage economies) have a stronger preference (B_{cp}) for an activity.

6 Relatedness and The Product Space

The other key observable used frequently in the economic complexity literature is a network connecting related activities [2, 48, 49, 51–54, 87, 89, 90, 124–132]. When these activities are products, this network goes by the name of "product space." From an application perspective, the product space is used to estimate the potential of economy in an activity (e.g. the probability that a city specializes in an industry [49–51, 89], a country starts exporting a product [2, 48, 54], or a university starts producing papers in a given field [52, 53]. These estimates of potential are known as measures of relatedness, and are akin to traditional recommender system methods in computer science [133]. Yet, in the economic complexity literature, they are used to explain economic development trajectories (e.g. countries entering new products) instead of individual consumption patterns (e.g. customers choosing to purchase a products at an online retailer) or to explore strategies to optimize industrial promotion efforts[134–136][17]

Product space type networks are important in empirical work since they help capture information about an economy's productive structure that is specific to an economy and activity. Thus, they can be used to either model path dependencies, or to control for them in work looking at the impact of other factors in economic diversification[145–148].

Here we begin by focusing on a particular characteristics of the product space that was emphasized when it was introduced as a network nearly twenty years ago: the fact that the core of the product space, its most densely connected part, is composed of high-complexity activities[48].

This is a characteristic that is true for networks derived from trade data, since networks derived from other data can have different forms. For example, networks connecting research fields based on citation patterns or co-authorships tend to follow a "ring" structure [52, 68]. Networks connecting skills based on the occupations that require them tend to follow a "dumbell" structure (two big clusters connected by a bridge) [124].

We begin our exploration by of the structure of the product space implied by the single and multi-capability theory by estimating a measure of proximity, which is an estimate of the similarity between products. Unlike in the case of ECI, where we have a more strict definition based on a second eigenvector, measures of proximity, in both the economic complexity and recommender systems literature, tend to be more ad-hoc, since there are many ways to estimate similarity among pairs of activities. In [48], proximity was introduced using the minimum of the conditional probability that two products are exported in tandem. In our notation, this translates to:

[17] In recent years there have also been multiple efforts to look at relatedness in the context of sustainability, starting from the idea of a green product space, [137–144]

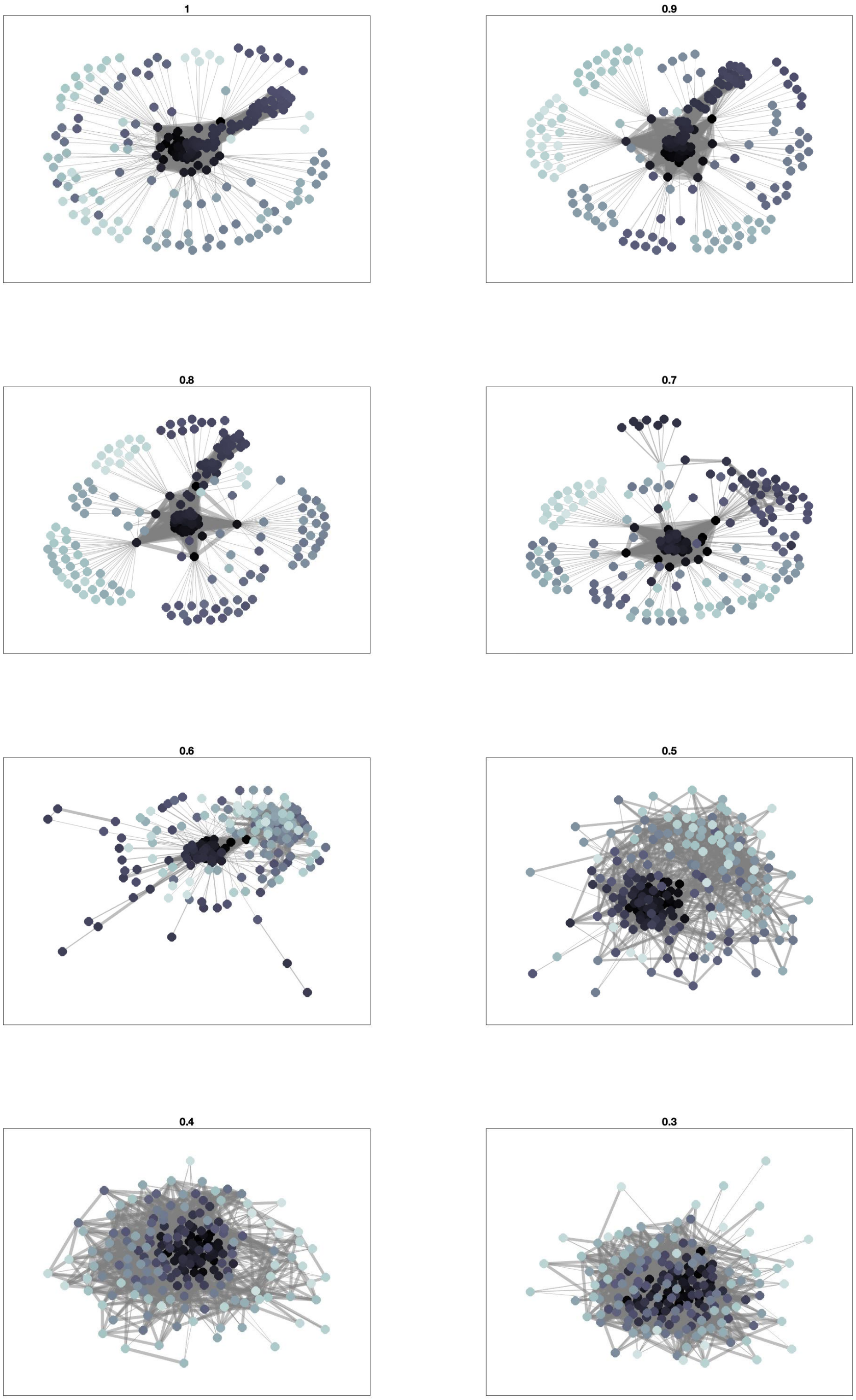

Figure 14: Average correlation between ECI and the average capability endowment of an economy $\langle r \rangle_c$ as a function of the mixing probability π (number indicated on the top of each chart). We can see that the correlation between ECI and $\langle r \rangle_c$ remains close to one for mixing probabilities above 0.4.

$$\phi_{pp'} = \frac{\sum_c M_{cp} M_{cp'}}{\max(Mp, M_{p'})} \tag{78}$$

In [49] they use simply the number of activities that are common to two economies.

$$\phi_{pp'} = \sum_c M_{cp} M_{cp'} \tag{79}$$

In general, it is not uncommon to find proximity matrices and recommender systems based on variations of $\sum_c M_{cp} M_{cp'}$ (usually with a normalization), so we will begin by exploring this basic form.

The product space implied by the single-capability model can be derived easily for the case in which the number of economies and activities is even. In that case, the proximity matrices are:

$$\phi_{pp'} = \sum_c M_{cp} M_{cp'} = M_p \quad if \quad q_p > \langle q \rangle \quad \& \quad q_{p'} > \langle q \rangle \tag{80}$$

$$\phi_{pp'} = \frac{\sum_c M_{cp} M_{cp'}}{\max(M_p, M_{p'})} = 1 \quad if \quad q_p > \langle q \rangle \quad \& \quad q_{p'} > \langle q \rangle \tag{81}$$

Which means a network composed of two clusters, one connecting the activities that are produced in high-complexity economies, and one connecting the activities produced in low complexity economies.

A more interesting exercise is to consider the networks implied by the multi-capability model. Here we present three examples in which we estimate networks for different model parameters that we visualize by estimating their minimum spanning tree and adding on top of that all of the links that are one standard deviation above the mean. This is a similar visualization exercise than the one used in the paper that introduced the product space network.

Figure 14 presents this exercise for a model involving 200 activities, 100 economies, and 10 capabilities. The color of the nodes indicates the complexity of the activity (with darker nodes being higher complexity). The number on top of each network visualization shows the mixing parameter π used to combine random and non-random capabilities.

We can see clearly in this example that all of the networks that are above the phase transition threshold are centered around a core of high-complexity activities, with lower complexity activities being peripheral. This reproduces the empirical observation presented in the original product space paper, which claimed that the core of the product space is composed of more sophisticated activities.

But can we use this model to generate the network observed for research activities, which follows a ring instead of a core-periphery structure? Or do we need to radically change our assumptions to obtain that shape?

To generate a ring type network we can use Toeplitz-like matrices for the capability endowments. A Toeplitz matrix is constant along each diagonal. By setting diagonals with decreasing values or $r_{c,b}$ and $q_{p,b}$ we can define correlations among subsets of related activities.

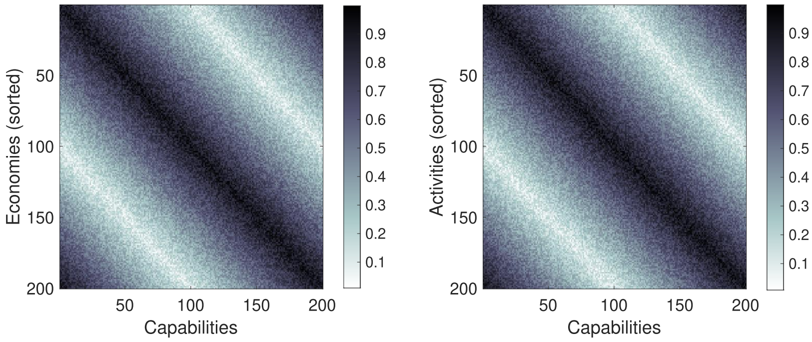

Figure 15: Parametrization of capability endowments and requirements using a symmetric Toeplitz circulant matrix combined with a random matrix in 80 percent and 20 percent proportions.

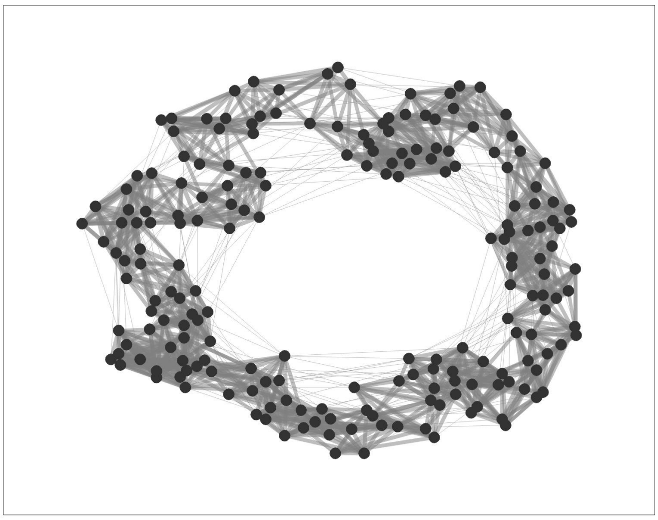

Figure 16: Network of related activities derived from the parametrization presented in figure 15.

Here, we use a parametrization where we combine a symmetric Toeplitz circulant matrix and a random matrix by using proportions of (α) and $(1-\alpha)$. A circulant matrix is a particular type of Toeplitz matrix that has periodic boundary conditions. A symmetric circulant matrix can be constructed by starting from a row that is symmetric with respect to the center. Here, we generate symmetric circulant matrices using linearly spaced probabilities for r_c and q_p that grow symmetrically from the center column of the first row. Figure 15 shows an example of this parametrization for a model with 200 economies, 200 activities, and 200 capabilities. We note that in this model talking about higher and lower complexity economies is not a useful construct, since economies do not differ on their average capability endowment (they are all equal on average), but in which subset of capabilities they are specialized in.

Figure 16 shows the network derived form this parametrization, visualized using the same method than before (minimum spanning tree, plus links that are one standard deviation above the average weight). The visualization shows a clear ring structure mimicking the one observed in networks involving research fields. The connectivity pattern of this network can be interpreted as research fields having a few related activities that share capabilities among them (e.g. capabilities are more re-deployable between molecular biology and biochemistry, than between polymer sciences and experimental psychology). This results in a network structure where each field is connected to a few neighbors.

Finally, we use the same approach to model a "dumbbell" network, which is a network with two well-defined yet clusters, such as the one observed when connecting skills and occupations [124]. Figures 17 and 18 show an example with 100 economies, 500 activities, and 20 capabilities. We note that obtaining this dumbbell structure requires a good level of mixing between the clusters, which can be achieved by setting the noise levels to be high enough so that some of the between cluster links are comparable in strength to the withing cluster links.

What is exciting about this general idea, is that it provides us with an intuitive way to map capability endowments to network structures. For instance, the core periphery-structure of the product space suggest that the capabilities associated with exporting products are correlated among economies, with high complexity economies like that of Singapore, Japan, or the United States, having high-values across a wide set of capabilities. The ring structure of the research space tells a different story. It is a story of specialization in a world of fine grained capabilities. Similarly, we can use this intuition to think about dumbbell structures, which can be modeled by assuming capability endowments made of slightly overlapping blocks.

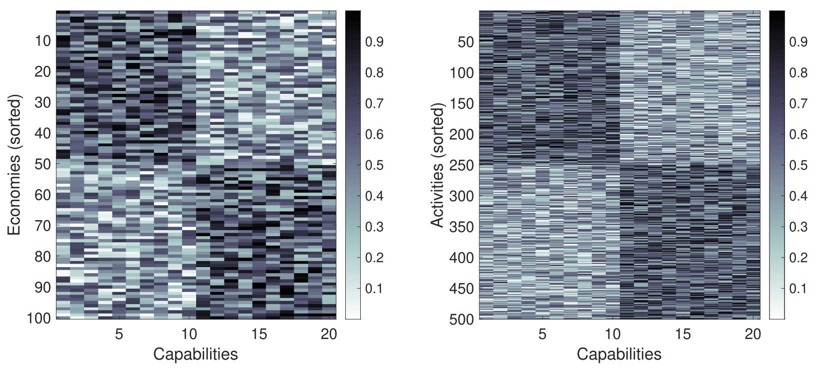

Figure 17: Parametrization of capability endowments and requirements using a 25 percent of a block diagonal matrix and 75 percent of a random matrix.

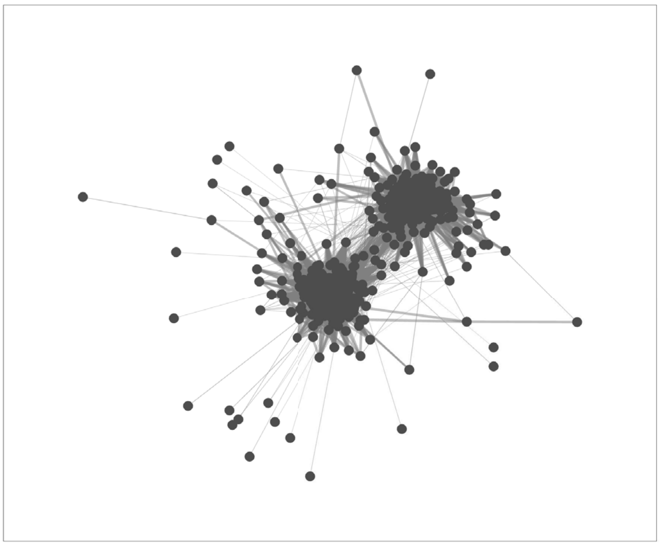

Figure 18: Network of related activities derived from the parametrization presented in figure 17.

7 Conclusion

In this paper, we provided an analytical foundation for the economic complexity index (ECI) starting from a production function. For the single-capability model, we were able to derive its key eigenvectors analytically and show that ECI is a statistic separating economies among those with an above- and below-average probability of having a capability. We then extended this result numerically to a multi-capability setting to show that ECI is as a monotonic estimator of an economy's average capability endowment—even when a substantial share of the capabilities are randomly assigned. In the multi-capability model, ECI is no longer a discrete measure separating low from high capability economies, but a monotonic transformation of the average capability endowment of an economy. These findings differentiate ECI from measures of diversity, which peak for capability endowments below the maximum (they are non-monotonic functions of r_c), and thus are non-ideal estimates of the complexity of an economy. These results validate ECI as a measure of composition or complexity, since they show the eigenvector captures information about an economy being endowed with multiple capabilities, regardless of how these capabilities are defined.

Interestingly, our main result does not depend on assuming an stochastic model or a theory based on capabilities, since the basic idea can be easily generalized to models including factors that are specific to economies and activities. The key condition for the measure of complexity to work is for output to not be perfectly proportional to factor endowments. This condition can be achieved by simply shifting the production function by a constant to make it non-multiplicatively separable.[18]

What is also interesting is that the condition needed for our main result to hold comes from calculating the matrix of specialization R_{cp}. This is a key difference with previous attempts to connect economic complexity theory and empirics [1, 55] which jumped directly to the binary spe-

[18]This mechanism is akin to the idea of symmetry-breaking in physics, since the shift removes symmetries of the function. For example K_c^γ satisfies the scale-invariance symmetry $f(\lambda K) = \lambda^\gamma f(K)$, whereas $B + K_c^\gamma$ does not have this symmetry.

cialization matrix M_{cp}. That assumption results in a monotonic relationship between the number of activities an economy is specialized (its diversity) and its capability endowments[19], which is an uncomfortable result since the empirical work had shown that measures of diversity fail to explain future economic growth like measures of complexity do [1]. We now understand that a key step in the calculation is estimating the specialization matrix, and that skipping this step in theoretical work results in a flawed connection. This change only helps provide a tight connection between the capability model and the economic complexity index, but also explains other findings, like that of Imbs and Warcziag [82], which says that economies diversify only until a certain point. It also opens questions about alternative measures of complexity. During the last fifteen years, many alternatives to the economic complexity index have been proposed, such as the Fitness index [11], the Ability index [12], and several others[13, 62, 80, 81, 149–152]. Since these indexes tend to exhibit strong correlations with ECI, our results provide a way to theoretically explore whether they are also monotonic functions of an economy's capability endowment (r_c and $\langle r_c \rangle$), and if they are, open the question about the importance of the functional form connecting this two quantities.

Our work also embedded this model in a short-run equilibrium framework including wages, consumption, and prices. Overall, we obtain reasonable results for all of them. Wages increase with capability endowments and prices are higher for more demanding products. Interestingly, prices do not affect the specialization condition, meaning that they leave the connection between capability endowments and economic complexity unchanged.[20]

Finally, we showed that the model can explain structural differences in networks of related activities, such as the product space and research space. By controlling the shape of the capability endowment matrices, we were able to reproduce the core-periphery structure observed in the product space [48], the ring structure observed for scientific publications [52, 68], and the dumbbell structure observed for networks of occupations and skills [124].

Together, these findings help resolve a few long-standing tensions in the economic complexity literature. First, and most importantly, the disconnect between its empirical metrics and their theoretical underpinnings. Our findings show that ECI is not an arbitrary or ad-hoc measure, but can be thought of as an estimator of an economy's capability endowments derived from its pattern of specialization. This is an interesting finding, since it provides a mean to estimate the combined presence of factor or capabilities even when these cannot be identified[21]

Second, we use standard macroeconomic assumptions to estimate the wages and prices associated with this model, which help support the well known empirical fact that economies tend to converge to a level of income that is related to their economic complexity [1–19].

And third, we provide theoretical underpinnings for the structure of the networks of related activities. Our findings show that the structure of these networks reflect how capabilities are distributed across economies and activities.

More broadly, our work helps clarify a field that had grown rapidly in its empirical scope while lacking a shared theoretical core. By grounding complexity metrics in production functions, and explaining the structure of networks of relatedness using a capability-based model, we offer

[19]That equation is provided in [55].

[20]We assume prices depend only on the activities and are the same across economies.

[21]This has for long been a tenet of this literature [1, 48, 107]. For example, in [48] relatedness is introduced as an "agnostic approach [...] based on the idea that, if two goods are related because they require similar institutions, infrastructure, physical factors, technology, or some combination thereof, they will tend to be produced in tandem," and in [107] complexity metrics are described as "an estimate of the overall potential of an economic structure, based on information on the geographic distribution of economic activities. As such, they help anticipate economic growth or emissions, not by identifying a specific factor, but by estimating their combined presence."

a framework that not only explains the empirical robustness of ECI, but should also open new paths for integrating economic complexity ideas further into development economics and trade theory.

Acknowledgments

This work owes a very special acknowledgment to Cristian Jara-Figueroa. In 2014, Cristian joined my (César's) group at the MIT Media Lab. During the first year of his Master's he worked on the mathematical theory of economic complexity producing an impressive internal manuscript with many results. Those results were never published, but they stayed with my group. Eleven years later, in 2025, while looking at Cristian's work, I realized we had made an important and simple mistake at the very beginning, which was to assume that the capability model was a model of M_{cp} instead of a model of the output matrix $Y_{c,p}$. This changed everything and motivated me to go back to square one to start estimating the intermediate matrices in the model (such as R_{cp}). In my mind this work owes enormously to that effort by Cristian many years ago. We would also like to acknowledge comments by Johanness Wachs and other members of the Center for Collective Learning. The section on prices and wages was motivated by a very useful conversation with Jean Tirole.

We acknowledge the support of the European Union LearnData, GA no. 101086712 a.k.a. 101086712-LearnDataHORIZON-WIDERA-2022-TALENTS-01 (`https://cordis.europa.eu/project/id/101086712`), IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and the European Lighthouse of AI for Sustainability grant number 101120237-HORIZON-CL4-2022-HUMAN-02.

References

[1] César A. Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575, June 2009.

[2] Ricardo Hausmann, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yildirim. *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press, 2014.

[3] Juan Carlos Chávez, Marco T. Mosqueda, and Manuel Gómez-Zaldívar. Economic complexity and regional growth performance: Evidence from the Mexican Economy. *Review of Regional Studies*, 47(2):201–219, 2017.

[4] Giacomo Domini. Patterns of specialization and economic complexity through the lens of universal exhibitions, 1855-1900. *Explorations in Economic History*, 83:101421, 2022. ISBN: 0014-4983 Publisher: Elsevier.

[5] Philipp Koch. Economic Complexity and Growth: Can value-added exports better explain the link? *Economics Letters*, 198:109682, 2021. ISBN: 0165-1765 Publisher: Elsevier.

[6] Viktor Stojkoski, Zoran Utkovski, and Ljupco Kocarev. The Impact of Services on Economic Complexity: Service Sophistication as Route for Economic Growth. *PLOS ONE*, 11(8):e0161633, August 2016.

[7] Viktor Stojkoski and Ljupco Kocarev. The relationship between growth and economic complexity: evidence from Southeastern and Central Europe. 2017.

[8] Guzmán Ourens. Can the Method of Reflections help predict future growth? *Documento de Trabajo/FCS-DE; 17/12*, 2012. Publisher: UR. FCS-DE.

[9] Sandra Poncet and Felipe Starosta de Waldemar. Economic Complexity and Growth. *Revue économique*, 64(3):495–503, 2013. ISBN: 0035-2764 Publisher: Presses de Sciences Po.

[10] Viktor Stojkoski, Philipp Koch, and César A. Hidalgo. Multidimensional Economic Complexity: How the Geography of Trade, Technology, and Research Explain Inclusive Green Growth, September 2022. arXiv:2209.08382 [cond-mat, q-fin].

[11] Andrea Tacchella, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli, and Luciano Pietronero. A New Metrics for Countries' Fitness and Products' Complexity. *Scientific Reports*, 2:srep00723, October 2012.

[12] Sebastian Bustos and Muhammed A. Yıldırım. Production ability and economic growth. *Research Policy*, 51(8):104153, 2022. ISBN: 0048-7333 Publisher: Elsevier.

[13] David Atkin, Arnaud Costinot, and Masao Fukui. Globalization and the ladder of development: Pushed to the top or held at the bottom? Technical report, National Bureau of Economic Research, 2021.

[14] Felipe Orsolin Teixeira, Fabricio José Missio, and Ricardo Dathein. Economic complexity, structural transformation and economic growth in a regional context: Evidence for brazil. *PSL Quarterly Review*, 75(300):63–79, 2022.

[15] Lilis Hoeriyah, Nunung Nuryartono, and Syamsul Hidayat Pasaribu. Economic complexity and sustainable growth in developing countries. *Economics Development Analysis Journal*, 11(1):23–33, 2022.

[16] Zhuqing Mao and Qinrui An. Economic complexity index and economic development level under globalization: An empirical study. *Journal of Korea Trade*, 25(7):41–55, 2021.

[17] Roberto Basile and Gloria Cicerone. Economic complexity and productivity polarization: Evidence from Italian provinces. *German Economic Review*, March 2022. Publisher: De Gruyter.

[18] Ben-Hur Francisco Cardoso, Eva Yamila da Silva Catela, Guilherme Viegas, Flávio L Pinheiro, and Dominik Hartmann. Export complexity, industrial complexity and regional economic growth in brazil. *arXiv preprint arXiv:2312.07469*, 2023.

[19] J Romero, E Freitas, F Silveira, G Britto, F Cimini, and G Jayme. Economic complexity and regional economic development: evidence from brazil. *EAEPE, Online Proceedings*, pages 1–22, 2021.

[20] Santiago Pérez-Balsalobre, Carlos Llano Verduras, and Jorge Díaz-Lanchas. Measuring subnational economic complexity: An application with spanish data. Technical report, JRC Working Papers on Territorial Modelling and Analysis, 2019.

[21] Dominik Hartmann, Miguel R. Guevara, Cristian Jara-Figueroa, Manuel Aristarán, and César A. Hidalgo. Linking Economic Complexity, Institutions, and Income Inequality. *World Development*, 93:75–93, May 2017.

[22] Margarida Bandeira Morais, J. Swart, and J. A. Jordaan. Economic Complexity and Inequality: Does Productive Structure Affect Regional Wage Differentials in Brazil? *USE Working Paper series*, 18(11), 2018. Publisher: USE Research Institute.

[23] Angelica Sbardella, Emanuele Pugliese, and Luciano Pietronero. Economic development and wage inequality: A complex system analysis. *PloS one*, 12(9), 2017. Publisher: Public Library of Science.

[24] Emilie Le Caous and Fenghueih Huarng. Economic Complexity and the Mediating Effects of Income Inequality: Reaching Sustainable Development in Developing Countries. *Sustainability*, 12(5):2089, January 2020. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[25] Myriam Ben Saâd and Giscard Assoumou-Ella. Economic Complexity and Gender Inequality in Education: An Empirical Study. *Economics Bulletin*, 39(1):321–334, 2019.

[26] Fadi Fawaz and Masha Rahnama-Moghadamm. Spatial dependence of global income inequality: The role of economic complexity. *The International Trade Journal*, 33(6):542–554, 2019. ISBN: 0885-3908 Publisher: Taylor & Francis.

[27] Shengjun Zhu, Changda Yu, and Canfei He. Export structures, income inequality and urban-rural divide in China. *Applied Geography*, 115:102150, February 2020.

[28] Radu Barza, Cristian Jara-Figueroa, CÃ Hidalgo, and Martina Viarengo. Knowledge Intensity and Gender Wage Gaps: Evidence from Linked Employer-Employee Data. 2020. Publisher: CESifo Working Paper.

[29] Chien-Chiang Lee and En-Ze Wang. Economic complexity and income inequality: Does country risk matter? *Social Indicators Research*, 154(1):35–60, 2021.

[30] Muhlis Can and Buhari Doğan. The effects of economic structural transformation on employment: an evaluation in the context of economic complexity and product space theory. In *Handbook of research on unemployment and labor market sustainability in the era of globalization*, pages 275–306. IGI Global, 2017.

[31] Olimpia Neagu. The Link between Economic Complexity and Carbon Emissions in the European Union Countries: A Model Based on the Environmental Kuznets Curve (EKC) Approach. *Sustainability*, 11(17):4753, 2019.

[32] Olimpia Neagu and Mircea Constantin Teodoru. The relationship between economic complexity, energy consumption structure and greenhouse gas emission: Heterogeneous panel evidence from the eu countries. *Sustainability*, 11(2):497, 2019.

[33] João P. Romero and Camila Gramkow. Economic complexity and greenhouse gas emissions. *World Development*, 139:105317, March 2021.

[34] Athanasios Lapatinas, Antonios Garas, Eirini Boleti, and Alexandra Kyriakou. Economic complexity and environmental performance: Evidence from a world sample, March 2019. Library Catalog: mpra.ub.uni-muenchen.de.

[35] Penny Mealy and Alexander Teytelboym. Economic complexity and the green economy. *Research Policy*, page 103948, 2020. ISBN: 0048-7333 Publisher: Elsevier.

[36] Manuel Gómez-Zaldívar, María Isabel Osorio-Caballero, and Edgar Juan Saucedo-Acosta. Income inequality and economic complexity: Evidence from mexican states. *Regional Science Policy & Practice*, 14(6):344–364, 2022.

[37] Daniel Balsalobre-Lorente, Clara Contente dos Santos Parente, Nuno Carlos Leitão, and José María Cantos-Cantos. The influence of economic complexity processes and renewable energy on co2 emissions of brics. what about industry 4.0? *Resources Policy*, 82:103547, 2023.

[38] Fabricio Silveira, João P Romero, Arthur Queiroz, Elton Freitas, and Alexandre Stein. Economic complexity and deforestation in the brazilian amazon. *World Development*, 185:106804, 2025.

[39] Gertjan Dordmond, Heder Carlos de Oliveira, Ivair Ramos Silva, and Julia Swart. The complexity of green job creation: An analysis of green job development in Brazil. *Environment, Development and Sustainability*, pages 1–24, 2020. ISBN: 1573-2975 Publisher: Springer.

[40] Barbaros Güneri and A. Yasemin Yalta. Does economic complexity reduce output volatility in developing countries? *Bulletin of Economic Research*. ISBN: 0307-3378 Publisher: Wiley Online Library.

[41] Trung V. Vu. Does LGBT Inclusion Promote National Innovative Capacity? SSRN Scholarly Paper ID 3523553, Social Science Research Network, Rochester, NY, January 2020.

[42] Diogo Ferraz, Herick Fernando Moralles, Jessica Suárez Campoli, Fabíola Cristina Ribeiro de Oliveira, and Daisy Aparecida do Nascimento Rebelatto. Economic complexity and human development: Dea performance measurement in asia and latin america. *Gestão & Produção*, 25(4):839–853, 2018.

[43] Athanasios Lapatinas and Marina-Selini Katsaiti. EU MECI: A Network-Structured Indicator for a Union of Equality. *Social Indicators Research*, February 2023.

[44] Radu Barza, Edward L. Glaeser, César A. Hidalgo, and Martina Viarengo. Cities as Engines of Opportunities: Evidence from Brazil, May 2024.

[45] Taylan Yenilmez. Understanding complexity in the author-journal space. *Scientometrics*, pages 1–28, 2025.

[46] Ronald Djeunankan, Sosson Tadadjeu, Henri Njangang, and Ummad Mazhar. The hidden cost of sophistication: economic complexity and obesity. *The European Journal of Health Economics*, 26(2):243–265, 2025.

[47] Omar Lizardo. The mutual specification of genres and audiences: Reflective two-mode centralities in person-to-culture data. *Poetics*, 68:52–71, 2018. ISBN: 0304-422X Publisher: Elsevier.

[48] C.A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann. The Product Space Conditions the Development of Nations. *Science*, 317(5837):482–487, July 2007.

[49] Frank Neffke, Martin Henning, and Ron Boschma. How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87:237–265, 2011.

[50] Frank Neffke and Martin Henning. Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297–316, 2013.

[51] C. Jara-Figueroa, Bogang Jun, Edward L. Glaeser, and Cesar A. Hidalgo. The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences*, 115(50):12646–12653, December 2018.

[52] Miguel R. Guevara, Dominik Hartmann, Manuel Aristarán, Marcelo Mendoza, and César A. Hidalgo. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics*, 109(3):1695–1709, December 2016.

[53] Matteo Chinazzi, Bruno Gonçalves, Qian Zhang, and Alessandro Vespignani. Mapping the physics research space: a machine learning approach. *EPJ Data Science*, 8(1):33, 2019. ISBN: 2193-1127 Publisher: Springer Berlin Heidelberg.

[54] Hidalgo, Pierre-Alexandre Balland, Ron Boschma, Mercedes Delgado, Maryann Feldman, Koen Frenken, Edward Glaeser, Canfei He, Dieter F. Kogler, Andrea Morrison, Frank Neffke, David Rigby, Scott Stern, Siqi Zheng, and Shengjun Zhu. The Principle of Relatedness. In Alfredo J. Morales, Carlos Gershenson, Dan Braha, Ali A. Minai, and Yaneer Bar-Yam, editors, *Unifying Themes in Complex Systems IX*, Springer Proceedings in Complexity, pages 451–457. Springer International Publishing, 2018.

[55] Ricardo Hausmann and César A. Hidalgo. The network structure of economic output. *Journal of Economic Growth*, pages 1–34, 2011.

[56] Ulrich Schetter. Comparative advantages with product complexity and product quality. 2016. Publisher: Kiel und Hamburg: ZBW-Deutsche Zentralbibliothek für ....

[57] Penny Mealy, J. Doyne Farmer, and Alexander Teytelboym. Interpreting economic complexity. *Science Advances*, 5(1):eaau1705, January 2019.

[58] Muhammed A. Yildirim. Sorting, Matching and Economic Complexity. *CID Working Paper Series*, 2021. Publisher: Center for International Development at Harvard University.

[59] Benjamin Cakir, Isabelle Schluep, Philipp Aerni, and Isa Cakir. Amalgamation of export with import information: The economic complexity index as a coherent driver of sustainability. *Sustainability*, 13(4):2049, 2021.

[60] Vito DP Servedio, Alessandro Bellina, Emanuele Calò, and Giordano De Marzo. Economic Complexity in Mono-Partite Networks. *arXiv preprint arXiv:2405.04158*, 2024.

[61] Carlo Bottai, Jacopo Di Iorio, and Martina Iori. Reinterpreting Economic Complexity: A co-clustering approach, June 2024. arXiv:2406.16199 [econ, q-fin, stat].

[62] James McNerney, Yang Li, Andres Gomez-Lievano, and Frank Neffke. Bridging the short-term and long-term dynamics of economic structural change, March 2023. arXiv:2110.09673 [physics, q-fin].

[63] Zoran Utkovski, Melanie F. Pradier, Viktor Stojkoski, Fernando Perez-Cruz, and Ljupco Kocarev. Economic complexity unfolded: Interpretable model for the productive structure of economies. *PloS one*, 13(8):e0200822, 2018.

[64] Benoît Desmarchelier, Paulo José Regis, and Nimesh Salike. Product space and the development of nations: A model of product diversification. *Journal of Economic Behavior & Organization*, 145:34–51, 2018.

[65] César A. Hidalgo. Economic complexity theory and applications. *Nature Reviews Physics*, pages 1–22, 2021. ISBN: 2522-5820 Publisher: Nature Publishing Group.

[66] Pierre-Alexandre Balland, Tom Broekel, Dario Diodato, Elisa Giuliani, Ricardo Hausmann, Neave O'Clery, and David Rigby. The new paradigm of economic complexity. *Research Policy*, 51(3):104450, 2022. ISBN: 0048-7333 Publisher: Elsevier.

[67] Michael Kremer. The O-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–575, 1993. ISBN: 1531-4650 Publisher: MIT Press.

[68] Katy Börner, Richard Klavans, Michael Patek, Angela M. Zoss, Joseph R. Biberstine, Robert P. Light, Vincent Larivière, and Kevin W. Boyack. Design and update of a classification system: The UCSD map of science. *PloS one*, 7(7):e39464, 2012. ISBN: 1932-6203 Publisher: Public Library of Science.

[69] Ministry of Investment Trade and Industry of Malaysia. New industrial master plan. `https://www.nimp2030.gov.my/`. [Accessed 16-06-2025].

[70] Mario Draghi. The future of european competitiveness part a: A competitiveness strategy for europe. 2024.

[71] Christian Reynolds, Manju Agrawal, Ivan Lee, Chen Zhan, Jiuyong Li, Phillip Taylor, Tim Mares, Julian Morison, Nicholas Angelakis, and Göran Roos. A sub-national economic complexity analysis of Australia's states and territories. *Regional Studies*, 52(5):715–726, 2018. ISBN: 0034-3404 Publisher: Taylor & Francis.

[72] Birol Erkan and Elif Yildirimci. Economic Complexity and Export Competitiveness: The Case of Turkey. *Procedia-Social and Behavioral Sciences*, 195:524–533, 2015. ISBN: 1877-0428 Publisher: Elsevier.

[73] Natalia Ferreira-Coimbra and Marcel Vaillant. Evolución del espacio de productos exportados:¿ está Uruguay en el lugar equivocado? *Revista de economía*, 16(2):97–146, 2009. ISBN: 0797-5546 Publisher: Banco Central del Uruguay.

[74] Ivan L. Lyubimov, Maria V. Lysyuk, and Margarita A. Gvozdeva. Atlas of economic complexity, Russian regional pages. *VOPROSY ECONOMIKI*, 6, 2018.

[75] I. Lyubimov, M. Gvozdeva, M. Kazakova, and K. Nesterova. Economic Complexity of Russian Regions and their Potential to Diversify. *Journal of the New Economic Association*, 34(2):94–122, 2017.

[76] Fernando Gómez Zaldívar, Edmundo Molina, Miguel Flores, and Manuel de Jesús Gómez Zaldívar. Economic Complexity of the Special Economic Zones in Mexico: Opportunities for Diversification and Industrial Sophistication. *Ensayos Revista de Economía (Ensayos Journal of Economics)*, 38(1):1–40, 2019. ISBN: 2448-8402.

[77] Carla Carolina Pérez Hernández, Blanca Cecilia Salazar Hernández, and Jessica Mendoza Moheno. Diagnóstico de la complejidad económica del estado de Hidalgo: de las capacidades a las oportunidades. *Revista mexicana de economía y finanzas*, 14(2):261–277, 2019. ISBN: 1665-5346 Publisher: Instituto Mexicano de Ejecutivos de Finanzas, AC.

[78] Yihan Wang and Ekaterina Turkina. Economic complexity, product space network and Quebec's global competitiveness. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, 37(3):334–349, 2020. ISBN: 0825-0383 Publisher: Wiley Online Library.

[79] Roberto Basile, Gloria Cicerone, and Lelio Iapadre. Economic complexity and regional labor productivity distribution: evidence from Italy. *Economic complexity and regional labor productivity distribution: evidence from Italy*, 2019.

[80] Jorge Valverde-Carbonell. Rethinking the Literature on Economic Complexity Indexes. *Economic Analysis and Policy*, May 2025.

[81] Carla Sciarra, Guido Chiarotti, Luca Ridolfi, and Francesco Laio. Reconciling contrasting views on economic complexity. *Nature communications*, 11(1):1–10, 2020. ISBN: 2041-1723 Publisher: Nature Publishing Group.

[82] Jean Imbs and Romain Wacziarg. Stages of diversification. *American economic review*, 93(1):63–86, 2003. ISBN: 0002-8282.

[83] Dieter F. Kogler, David L. Rigby, and Isaac Tucker. Mapping knowledge space and technological relatedness in US cities. In *Global and Regional Dynamics in Knowledge Flows and Innovation*, pages 58–75. Routledge, 2015.

[84] Bogang Jun, Aamena Alshamsi, Jian Gao, and César A. Hidalgo. Bilateral relatedness: knowledge diffusion and the evolution of bilateral trade. *Journal of Evolutionary Economics*, pages 1–31, 2019. ISBN: 0936-9937 Publisher: Springer.

[85] Teresa Farinha, Pierre-Alexandre Balland, Andrea Morrison, and Ron Boschma. What drives the geography of jobs in the us? unpacking relatedness. *Industry and Innovation*, 26(9):988–1022, 2019. ISBN: 1366-2716 Publisher: Taylor & Francis.

[86] Pierre-Alexandre Balland, José Antonio Belso-Martínez, and Andrea Morrison. The dynamics of technical and business knowledge networks in industrial clusters: Embeddedness, status, or proximity? *Economic geography*, 92(1):35–60, 2016.

[87] Ron Boschma, Asier Minondo, and Mikel Navarro. The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness. *Economic Geography*, 89(1):29–51, January 2013.

[88] Teresa Farinha, Pierre-Alexandre Balland, Andrea Morrison, and Ron Boschma. What drives the geography of jobs in the us? unpacking relatedness. *Industry and Innovation*, 26(9):988–1022, 2019. ISBN: 1366-2716 Publisher: Taylor & Francis.

[89] Zhao Chen, Sandra Poncet, and Ruixiang Xiong. Inter-industry relatedness and industrial-policy efficiency: Evidence from China's export processing zones. *Journal of Comparative Economics*, 45(4):809–826, December 2017.

[90] Benno Ferrarini and Pasquale Scaramozzino. The product space revisited: China's trade profile. *The World Economy*, 38(9):1368–1386, 2015. ISBN: 0378-5920 Publisher: Wiley Online Library.

[91] Dieter F. Kogler, David L. Rigby, and Isaac Tucker. Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, 21(9):1374–1391, September 2013.

[92] Fengmei Ma, Heming Wang, Asaf Tzachor, César A Hidalgo, Heinz Schandl, Yue Zhang, Jingling Zhang, Wei-Qiang Chen, Yanzhi Zhao, Yong-Guan Zhu, et al. The disparities and development trajectories of nations in achieving the sustainable development goals. *Nature Communications*, 16(1):1107, 2025.

[93] Rachata Muneepeerakul, José Lobo, Shade T Shutters, Andrés Goméz-Liévano, and Murad R Qubbaj. Urban economies and occupation space: Can they get "there" from "here"? *PloS one*, 8(9):e73676, 2013.

[94] Matthieu Cristelli, Andrea Gabrielli, Andrea Tacchella, Guido Caldarelli, and Luciano Pietronero. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one*, 8(8):e70726, 2013.

[95] Carla Sciarra, Guido Chiarotti, Luca Ridolfi, and Francesco Laio. Reconciling contrasting views on economic complexity. *Nature communications*, 11(1):3352, 2020. ISBN: 2041-1723 Publisher: Nature Publishing Group UK London.

[96] Bernardo Caldarola, Dario Mazzilli, Lorenzo Napolitano, Aurelio Patelli, and Angelica Sbardella. Economic complexity and the sustainability transition: A review of data, methods, and literature. *Journal of Physics: Complexity*, 2024.

[97] Ricardo Hausmann, Jason Hwang, and Dani Rodrik. What you export matters. *Journal of Economic Growth*, 12(1):1–25, March 2007.

[98] Dani Rodrik. What's so special about China's exports? *China & World Economy*, 14(5):1–19, 2006. ISBN: 1671-2234 Publisher: Wiley Online Library.

[99] Alexander Hamilton. Report on manufactures. 1791. Publisher: Washington, DC: United States.

[100] Paul N. Rosenstein-Rodan. Notes on the theory of the 'big push'. In *Economic Development for Latin America*, pages 57–81. Springer, 1961.

[101] Paul N. Rosenstein-Rodan. Problems of industrialisation of eastern and south-eastern Europe. *The economic journal*, 53(210/211):202–211, 1943. ISBN: 0013-0133 Publisher: JSTOR.

[102] Walt Whitman Rostow. The stages of economic growth. *The economic history review*, 12(1):1–16, 1959. ISBN: 0013-0117 Publisher: JSTOR.

[103] Albert O. Hirschman. A generalized linkage approach to development, with special reference to staples. *Economic development and cultural change*, 25:67, 1977. ISBN: 0013-0079 Publisher: University of Chicago Press.

[104] Raul Prebisch. The economic development of Latin America and its principal problems. *Economic Bulletin for Latin America*, 1962.

[105] Alexander Gerschenkron. The early phases of industrialization in Russia: afterthoughts and counterthoughts. In *The economics of take-off into sustained growth*, pages 151–169. Springer, 1963.

[106] Bela Balassa. Exports, policy choices, and economic growth in developing countries after the 1973 oil shock. *Journal of Development Economics*, 18(1):23–35, May 1985.

[107] César A. Hidalgo. The policy implications of economic complexity. *Research Policy*, 52(9):104863, 2023. ISBN: 0048-7333 Publisher: Elsevier.

[108] Martin L. Weitzman. Recombinant Growth. *The Quarterly Journal of Economics*, 113(2):331–360, May 1998.

[109] Stuart A. Kauffman. *The origins of order: Self-organization and selection in evolution.* Oxford University Press, USA, 1993.

[110] Francesca Tria, Vittorio Loreto, Vito Domenico Pietro Servedio, and Steven H. Strogatz. The dynamics of correlated novelties. *Scientific reports*, 4:5890, 2014. ISBN: 2045-2322 Publisher: Nature Publishing Group.

[111] T. M. A. Fink, M. Reeves, R. Palma, and R. S. Farr. Serendipity and strategy in rapid innovation. *Nature Communications*, 8(1):2002, December 2017.

[112] T. M. A. Fink and M. Reeves. How much can we influence the rate of innovation? *Science Advances*, 5(1):eaat6107, January 2019.

[113] Anton Pichler, François Lafond, and J Doyne Farmer. Technological interdependencies predict innovation dynamics. *arXiv preprint arXiv:2003.00580*, 2020.

[114] James McNerney, J Doyne Farmer, Sidney Redner, and Jessika E Trancik. Role of design complexity in technology improvement. *Proceedings of the National Academy of Sciences*, 108(22):9008–9013, 2011.

[115] Lee Fleming. Recombinant uncertainty in technological search. *Management science*, 47(1):117–132, 2001.

[116] Alje Van Dam and Koen Frenken. Variety, complexity and economic development. *Research Policy*, 51(8):103949, 2022.

[117] Giovanni Dosi. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162, 1982.

[118] David J Teece, Gary Pisano, and Amy Shuen. Dynamic capabilities and strategic management. *Strategic management journal*, 18(7):509–533, 1997.

[119] Sanjaya Lall. Technological capabilities and industrialization. *World development*, 20(2):165–186, 1992.

[120] Richard R. Nelson and Sidney G. Winter. *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press, 1982. Google-Books-ID: uRm5AAAAIAAJ.

[121] Jorge M Uribe. Investment in intangible assets and economic complexity. *Research Policy*, 54(1):105133, 2025.

[122] Marc J Melitz and Stephen J Redding. Missing gains from trade? *American Economic Review*, 104(5):317–321, 2014.

[123] Sebastián Bustos, Charles Gomez, Ricardo Hausmann, and César A. Hidalgo. The Dynamics of Nestedness Predicts the Evolution of Industrial Ecosystems. *PLOS ONE*, 7(11):e49393, November 2012.

[124] Ahmad Alabdulkareem, Morgan R. Frank, Lijun Sun, Bedoor AlShebli, César Hidalgo, and Iyad Rahwan. Unpacking the polarization of workplace skills. *Science Advances*, 4(7):eaao6030, July 2018.

[125] Louis Knuepling and Tom Broekel. Does relatedness drive the diversification of countries' success in sports? *European Sport Management Quarterly*, 22(2):182–204, 2022.

[126] Benjamin Klement and Simone Strambach. Innovation in creative industries: Does (related) variety matter for the creativity of urban music scenes? *Economic Geography*, 95(4):385–417, 2019.

[127] Fabian Stephany and Ole Teutloff. What is the price of a skill? the value of complementarity. *Research Policy*, 53(1):104898, 2024.

[128] Gloria Cicerone, Philip McCann, and Viktor A. Venhorst. Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space. *Journal of Economic Geography*, 20(1):293–316, 2020. ISBN: 1468-2702 Publisher: Oxford University Press.

[129] Sándor Juhász, Tom Broekel, and Ron Boschma. Explaining the dynamics of relatedness: the role of co-location and complexity. *Papers in Regional Science*, 2020.

[130] César A. Hidalgo, Elisa Castañer, and Andres Sevtsuk. The amenity mix of urban neighborhoods. *Habitat International*, page 102205, 2020.

[131] Jonathan Borggren, Rikard H. Eriksson, and Urban Lindgren. Knowledge flows in high-impact firms: How does relatedness influence survival, acquisition and exit? *Journal of Economic Geography*, 16(3):637–665, 2016. ISBN: 1468-2710 Publisher: Oxford University Press.

[132] Rachata Muneepeerakul, José Lobo, Shade T. Shutters, Andrés Goméz-Liévano, and Murad R. Qubbaj. Urban Economies and Occupation Space: Can They Get "There" from "Here"? *PloS one*, 8(9):e73676, 2013.

[133] Paul Resnick and Hal R. Varian. Recommender systems, March 1997.

[134] Aamena Alshamsi, Flávio L. Pinheiro, and Cesar A. Hidalgo. Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications*, 9(1):1328, April 2018.

[135] Marcin Waniek, Khaled Elbassioni, Flávio L. Pinheiro, César A. Hidalgo, and Aamena Alshamsi. Computational aspects of optimal strategic network diffusion. *Theoretical Computer Science*, 2020. ISBN: 0304-3975 Publisher: Elsevier.

[136] Viktor Stojkoski and César A Hidalgo. Optimizing economic complexity. *arXiv preprint arXiv:2503.04476*, 2025.

[137] Robert Hamwey, Henrique Pacini, and Lucas Assunção. Mapping green product spaces of nations. *The Journal of Environment & Development*, 22(2):155–168, 2013.

[138] Nicola Daniele Coniglio, Raffaele Lagravinese, and Davide Vurchio. Production sophisticatedness and growth: evidence from Italian provinces before and during the crisis, 1997–2013. *Cambridge Journal of Regions, Economy and Society*, 9(2):423–442, July 2016. Publisher: Oxford Academic.

[139] Sandro Montresor and Francesco Quatraro. Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies. *Regional Studies*, pages 1–12, 2019.

[140] Penny Mealy and Alexander Teytelboym. Economic complexity and the green economy. *Available at SSRN 3111644*, 2017.

[141] Mark Huberty and Georg Zachmann. Green exports and the global product space: prospects for EU industrial policy. Technical report, Bruegel working paper, 2011.

[142] Luca Fraccascia, Ilaria Giannoccaro, and Vito Albino. Green product development: What does the country product space imply? *Journal of cleaner production*, 170:1076–1088, 2018.

[143] François Perruchas, Davide Consoli, and Nicolò Barbieri. Specialisation, diversification and the ladder of green technology development. *Research Policy*, 49(3):103922, 2020.

[144] Artur Santoalha and Ron Boschma. Diversifying in green technologies in European regions: does political support matter? *Regional Studies*, pages 1–14, 2020. ISBN: 0034-3404 Publisher: Routledge.

[145] Ron Boschma and Gianluca Capone. Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy*, 44(10):1902–1914, December 2015.

[146] Shengjun Zhu, Canfei He, and Yi Zhou. How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography*, 17(3):521–545, May 2017.

[147] Yongyuan Huang and Shengjun Zhu. Regional industrial dynamics under the environmental pressures in China. *Journal of Cleaner Production*, page 121917, 2020. ISBN: 0959-6526 Publisher: Elsevier.

[148] Nicola Cortinovis, Jing Xiao, Ron Boschma, and Frank G. van Oort. Quality of government and social capital as drivers of regional diversification in Europe. *Journal of Economic Geography*, 17(6):1179–1208, 2017. ISBN: 1468-2702 Publisher: Oxford University Press.

[149] Giorgio Gnecco, Federico Nutarelli, and Massimo Riccaboni. A machine learning approach to economic complexity based on matrix completion. *Scientific Reports*, 12(1):9639, 2022. ISBN: 2045-2322 Publisher: Nature Publishing Group UK London.

[150] Abdulrahman M. AlQurtas. A New Indicator of Economic Complexity to Guide Industrial Policies, 2018.

[151] Inga Ivanova, Øivind Strand, Duncan Kushnir, and Loet Leydesdorff. Economic and technological complexity: A model study of indicators of knowledge-based innovation systems. *Technological Forecasting and Social Change*, 120:77–89, 2017.

[152] Inga Ivanova, Nataliya Smorodinskaya, and Loet Leydesdorff. On measuring complexity in a post-industrial economy: The ecosystem's approach. *Quality & Quantity*, 54(1):197–212, 2020. ISBN: 1573-7845 Publisher: Springer.