Inferring Macroeconomic Complexity from Country-Product Network Data

César A. Hidalgo and Ricardo Hausmann

Center for International Development and Harvard Kennedy School, Harvard University

Abstract

We present a simple method to infer the relative number of non-tradable inputs available in a country from trade data connecting countries to the products they export. We show that countries approaches over the long run a level of income that is determined by the diversity of inputs available in the country, as approximated by the measures introduced here.

Introduction

For Adam Smith, wealth was related to the division of labor. As people and firms specialize in different activities, economic efficiency increases, suggesting that development is associated with an increase in the number of individual activities and with the complexity that emerges from the interactions between them (Smith 1776).Despite this, neo-classical growth theory and almost all empirical studies on growth have disregarded Smith's intuition. Theoretical studies on economic growth have instead attempted to explain the economic growth process using aggregate production functions which assume that a country's prosperity can be explained as a combination of a few unspecific factors of production. Empirical studies, on the other hand, have not attempted to include any measures of economic complexity. Moreover, being mostly based on regression analysis, empirical work on economic growth cannot be connected to theoretical studies. This is because the observables of economic theory, being highly stylized and abstract, do not have direct representations in the empirical world, leaving theory disconnected from data and limiting our understanding of the economic development process.

There are crucial aspects of the real world that cannot be properly considered by using theories that use aggregate production functions. In simple, an aggregate production function is a map between a few aggregate inputs and a single output. The most common inputs that go into aggregate production functions are Labor (L) and Capital (K), where L represents the number of people in a population and K represents the amount of capital that is available for people to work with. An easy way to understand the concept of

capital is to think of it as the tools and infrastructure that people need to produce things. Despite the conspicuous diversity of capital, theories aggregate into a single value (K) the total number of hammers, telephones, fork-lifts, computer screens, shovels, fertilizers, bridges, power-plants, industrial warehouses, etc. The aggregation is based on their market prices, so if a computer costs 10,000 times more than a pencil, then in this theory 10,000 pencils would be considered equivalent to a computer. On the other side of an aggregate production function all sorts of outputs are aggregated. Hence apples, oranges, high-speed motorcycles and smartphones are added according to their market prices to calculate the output of a given amount of K and L. These abstractions clearly are not adequately constructed to distinguish between different forms of capital, labor skills or output. From the way in which aggregate production functions are constructed, it cannot matter whether you produce oranges, enzymes or diesel engines, as a large enough number of oranges is equivalent to a diesel engine, in *aggregate* terms. The diversity of capital and labor cannot be considered when the world is abstracted in such a way. We find, however, that this diversity is not a detail, but rather an essential aspect of a country's macroeconomic structure.

Results and Methods

The question is then: How can we quantify economic diversity? or complexity? To do this, first, we need to abstract the world differently. Instead of abstracting the world as aggregates of capital and labor that are mapped into an aggregate output through a functional form, we abstract the world as a network that connects countries, products and inputs. This view can be understood easily by thinking of each type of product as a form of output and by thinking of it as a Lego Model, or toy, in which each of the pieces that goes into it is one of its required inputs. Here, each product is defined as a long string of ones and zeros, where the ones indicate the inputs, or Lego pieces, that the product requires, and the zeros indicate what inputs, or Lego pieces, the product does not require. The set of all products and inputs forms a bipartite network connecting each product to the inputs required for its production. In this case, however, the inputs that we have in mind as those that matter most are not the physical and tradable inputs that go into the production of a product,

but rather the variety of non-tradable, or hardly tradable inputs that need to be available to produce that good. The Lego analogy is therefore conceptual, rather than direct. For instance, if the product is apples, in our abstraction of its production some of the Lego pieces it requires could be adequate customs, a proper port infrastructure, cold storage transport services and facilities, phytosanitary regulations, a climate that is adequate for growing apples and specific human capital skills such as grafting and expertise in apple tree biology. Countries, on the other hand, at any point in time are endowed with a specific set of Lego pieces which determines the mix of products they can make, as countries can only make products for which they have all the necessary pieces available.

While in practice, we do not know what Lego pieces

Figure 1: **a-c**, GDP as a function of our first three measures of diversification ($k_{c,0}$, $k_{c,2}$, $k_{c,4}$), normalized by subtracting their respective means ($\langle k_{c,N} \rangle$) and dividing them by their standard deviations (stdev($k_{c,N}$)). **a**, $k_{c,0}$, **b**, $k_{c,2}$. **c**, $k_{c,4}$. **d**, Absolute value of the Pearson correlation between the log(GDP) of countries and its local network structure characterized by $k_{c,N}$. (Figure from Hidalgo and Hausmann 2009)

countries have, and what Lego pieces products require, in principle it should be possible to say something about the relative number of Lego pieces in a country, or required by a product, by looking at the network that connect countries to the products they make. This is because the production of a product by a country carries information on the Lego pieces available in that country, and on the Lego pieces required by that product. Hence, the structure of the bipartite network connecting countries to products has information on the structure of the tripartite network connecting countries to the Lego pieces they have and products to the Lego pieces they require. Formally, we represent the network connecting countries to products using the adjacency matrix M_{cp} and will show how some of its statistical properties carry information on the networks connecting countries to the Lego pieces they have (C_{ca}) and products to the Lego pieces they require (P_{pa}).
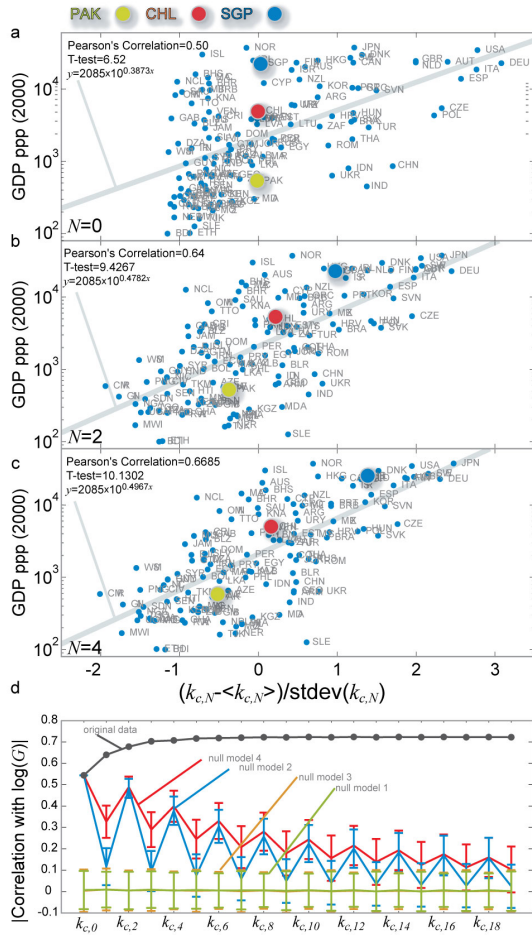
We observe M_{cp} empirically using trade data compiled by Feenstra et al. connecting 132 countries to 1006 product categories disaggregated according to the SITC-4 rev2 classification (Feenstra et al 2005). We consider a country as the exporter of a product if its Revealed Comparative Advantage (R) (Balassa 1986) is larger or equal than 1. The Revealed Comparative Advantage that a country has on a product is defined as the ratio between: (i) the share of the market of a country that a product has and (ii) the share of the world market that a product represents. For instance, in the year 2000, copper represented 26% of Chile's exports, but accounted for 0.2% of total world trade. Hence, Chile had an RCA on copper for the year 2000 of $R=26/0.2=130$, indicating that Chile is an extremely competitive copper exporter. In general, it is customary to say that country c has RCA in product p if $R_{cp} \geq 1$. By definition if a country does not export a product it will have no revealed comparative advantage in it ($R=0$). RCA is a measure of the importance of a product in a country's export basket that controls for both the size of the country's economy and the size of the products market. This simplification allows us to define M_{cp} a simple graph that connects countries to the products they export if the RCA of a country in a product is larger or equal than 1.

Inferring the relative number of Lego pieces in a country can be done relatively simply if we assume that countries can only do products if they have all the Lego pieces that those products require. In such a case, the diversification, or number of products that a country makes, will be related to the number od Lego pieces that that country has available. This is because countries that have a larger variety of Lego pieces will have, even by chance, the combinations required to produce a larger variety of products than countries with few Lego pieces. From a network perspective the diversification of a country can be calculated simply as the degree, or number of links, that that country has in the country-product network. We denote the degree, or diversification, of a country in the country product network as $k_{c,0}$, which can be cal-
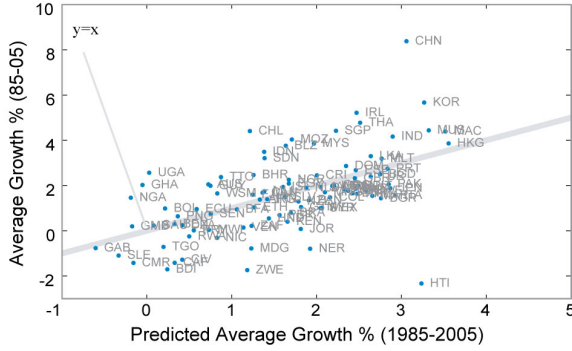
Figure 2: Growth in GDP observed between 1985-2005 as a function of growth predicted from kc,18 and kc,19 measured in 1985 and controlling for GDP in 1985. (Figure from Hidalgo and Hausmann 2009)

culated as:

$$k_{c,0} = \sum_p M_{cp}. \tag{1}$$

Diversification is related to the number of Lego pieces available in a country, albeit imperfectly. This is because countries producing the same number of products could be making goods that require different numbers of Lego pieces. In such cases, the diversification of countries would not be the most accurate estimator of the number of Lego pieces available in those countries, and we would need a measure of the number of Lego pieces required by a product to correct for this.

Using the symmetry of the bipartite network we can estimate the number of capabilities required by a product by looking at that product's ubiquity, which is given by the number of countries that produce a product and can be calculated as:

$$k_{p,0} = \sum_c M_{cp}. \tag{2}$$

The ubiquity of a product will be related to the number of Lego pieces that product requires because products that require few Lego pieces will be more likely to be produced in many countries. This is because countries with both, many and few Lego pieces, will have the necessary mix required by products needing only a small number of Lego pieces. Therefore, we can improve our estimate of the number of Lego pieces available in a country, or of the complexity of its economy, by looking at the average ubiquity of the products that a country exports.

Yet, the ubiquity of a product is also an imperfect measure of the number of Lego pieces it requires and it needs to be corrected by a measure of the number of Lego pieces available in the countries producing that product. From the above we know that we can approximate the number of Lego pieces in a country by that country's level of diversification. Hence, diversification and ubiquity are both measures of the

number of Lego pieces available in a country, or required by a product, that have complementary biases and can be used to create estimates of the relative number of Lego pieces available in a country and required by products by iteratively correcting for one another. This is the idea behind the Method of Reflections (Hidalgo and Hausmann 2009) which results in a family of variables that can be used to estimate the complexity of countries' productive structures and the sophistication of products. The variables are given by:

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_p M_{cp} k_{p,N-1}, \tag{3}$$

$$k_{p,N} = \frac{1}{k_{p,0}} \sum_c M_{cp} k_{c,N-1}, \tag{4}$$

The variables produced by the method of reflections can be seen as estimators of the number of Lego pieces available in a country and required by products and will be used in the rest of the paper as indicators of the complexity of country's productive structure and product sophistication. As we increase N, the values taken by these variables converge to the mean, so we need to worry only about the relative values of these variables, rather than their absolute magnitudes.

Economic Complexity Income and Growth

To better understand the value of the macroeconomic complexity measures developed above we compare it with the GDP per capita of countries adjusted by purchasing power parity (G), which is a well established measure of a country's level of economic development. When doing this we find that Pearson's correlation between G and $k_{c,n}$ grows as a function of n (Figure 1), indicating that there is a relationship between a country's level of income and the number of Lego Pieces we infer that country to have. More interestingly, however, is the fact that the residual of this relationship predicts economic growth (Figure 2). This suggest that countries tend to approach a level of income which is determined by the complexity of their economies.

Conclusion

By abstracting the economy using Networks, instead of aggregate production functions, it is possible to quantify the diversity of economic inputs. This allows us to differentiate easily between countries that, while having the same aggregate amount of capital, have their capital endowment structured differently. This distinction allowed us to show that the diversity of capital, or more generally, the diversity in the different number of inputs available in a country determines the ability of countries to generate income and prosperity.

The diversity of capital implies that coordination problems should be among the most common obstacles of economic development. This is because when diversity is taken into account it becomes clear that economic growth is not about accumulating more capital, but it is rather about

accumulating complementary forms of capital. The returns to each one of this specific forms of capital, or Lego pieces, however, depends on the existence of other pieces. This lies at the root of the coordination problems that policy makers and entrepreneurs face when trying to develop a new industrial sector. Countries partially overcome this coordination problems by moving towards products that require a similar set of Lego pieces than those required by the products they are currently producing. This defines a network of relatedness of products, or Product Space (Hidalgo et al 2007) that can be approximated by projecting the country-network into the space of products. Ultimately, the Product Space can be used to help guide industrial policy and development in places where the few options available can be hard to find, if any.

References

Smith, A. 1776 The Wealth of Nations

Feenstra, R.C.; Lipsey, R.E.; Deng, H.; Ma, A.C.; Ma, H. 2005 *NBER Working Paper* 11040.

Balassa, B. 1986. Comparative advantage in manufactured goods: A reappraisal. *Rev. Econ. Stat.* 68:315.

Hidalgo, C.A.; Hausmann, R. 2009. The building blocks of economic complexity. *Proc. Natl. Acad. Sci.* 106(26):10570-10575.

Hidalgo, C.A.; Klinger, B.; Barabási, A.-L.; Hausmann, R. 2007. The product space conditions the development of nations. *Science* 317:482-487.