Can the Method of Reflections help predict future growth?

G. Ourens

Discussion Paper 2013-8

Institut de Recherches Économiques et Sociales
de l'Université catholique de Louvain

Can the Method of Reflections help predict future growth?

Guzmán Ourens*

Abstract

Building upon an original and fruitful research line, a recent paper by Hidalgo and Hausmann (2009) proposed new indicators of product sophistication and economic complexity constructed solely upon international trade data, in their *Method of Reflections*. The authors find their indicators for economic complexity to be highly related to countries' income and show evidence supporting their use as predictors of future growth in the short and long run. This would make these indicators very appealing to empirical economists and policy-makers. This work tests these properties for the indicators constructing them upon a more disaggregated database and changing some other important methodological decisions. Results show that MR indicators are strongly related to income and they can be considered good predictors of long-term growth under certain conditions. Evidence supporting MR indicators as good predictors of short-term growth could not be found.

Keywords: Method of reflections, specialization, growth, economic complexity.

JEL Classification numbers: O47, O33, F14.

Acknowledgements

I would like to thank Florian Mayneris and Hylke Vandenbussche for their many comments and suggestions. I also appreciate comments received by Mery Ferrando and Yannick Thuy. I gratefully acknowledge financial support from Universidad de la República, Uruguay and from the Belgian French-speaking Community (convention ARC 09/14-018 on "Sustainability"). Remaining errors are my own responsibility.

*IRES-Université catholique de Louvain (Belgium) and dECON-Universidad de la República (Uruguay)

1 Introduction

Should countries make a deliberate effort to change their specialization in order to enhance their growth possibilities? If the answer is yes, then in which direction should the changes be heading? These important questions are a matter of debate and are of key relevance to policy-makers, especially in poor countries.

Classic growth and trade models state that the kind of specialization a country has does not determine its future growth. Benchmark growth models like those in Ramsey (1928) or Solow (1956) are built upon one product economies so no importance is given to the difference in production between countries. In trade literature, models focusing on specialization are mostly based on the Heckscher-Ohlin model, which concludes that a country should specialize in activities that use intensively the resources that it has a relative advantage in (see Heckscher and Ohlin (1991)). But again the model is mute on whether the specialization in one kind of production yields higher growth than specialization in another.

There are some early contributions paying attention to the fact that what a country produces is related to its growth possibilities. For Prebisch (1949), the process through which countries diverge in income levels is explained by their original specialization. While some countries initially specialized in high productivity activities, the rest specialized in a variety of activities with heterogeneous productivity levels which make this second group grow at a lower average rate over the long run. Closer to the mainstream, in Lewis (1954) it is possible to find one of the earliest models showing how growth processes imply structural changes in the long run. In his model of two sectors, capital accumulation in the high productivity sector induces growth but also determines the subtraction of labor from the low productivity sector.

As time passed by, the empirical evidence that emerged strongly supported the idea that rich countries produce different products than poor countries (see for example Sachs and Warner (1995), Lall (2000), Hausmann, et al. (2007) or Ranjan Basu and Das (2011)). But even though this idea has been around for a while and empirical evidence supporting it is strong, there are still many works overlooking that fact.

Over the last few years new contributions have emerged on this debate. One particular research line that has received great attention from economic advisers and policy-makers around the world is that started by Hausmann and Rodrik (2003) and further developed by many works, the most recent being Hausmann and Hidalgo (2011). In Hausmann et al. (2007) there is a proposal to measure the contribution each product makes to the growth process and, building on that, they also presented a synthetic measure of the growth possibilities of nations according to what they are currently producing. The potential these tools have to be used for policy-making recommendation is huge and did not go unnoticed, there are plenty of policy-oriented documents using them.

Hidalgo and Hausmann (2009) further developed these tools presenting the *Method of Reflections* (MR) which provides an original approach to the measure of *product sophistication* and *economic complexity*, based only on product's international trade data. The main argument is that the required capabilities for the production of one good can only be partially substituted by some others and so the set of capabilities in the economy determines what can be potentially produced in it. Sophisticated goods (i.e. those requiring a large set of diverse capabilities) will be produced only by complex economies (i.e. those having a large diversity of capabilities) which implies that the characteristics of a country's current production determine its growth possibilities. The authors claim that, by looking at what a country is exporting with revealed comparative advantages, their indicators are able to extract information about each countries' productive capabilities and provide a synthetic measure of economic complexity that is not only related to countries' current income but can predict future growth in the long and short run as well.

This work proposes to test the robustness of these properties by constructing MR indicators over a different dataset and by changing some important methodological decisions. The use of trade data in a six-digit aggregation level (opposed to the four-digit aggregation data used by the authors to support the properties) allows for a greater accuracy in the distinction of different products' capability requirements which makes it more suitable for the construction of the indicators. This work also presents results using different country samples, changing the revealed comparative advantage parameter in the construction of the indicators and including control variables in the analysis to see whether results are depending on these decisions or not. By performing these robustness checks for such a promising set of indicators this work aims to contribute to the debate on the influence specialization has on growth and, more particularly, aims at providing useful information to policy-makers concerned about structural change. Results show that MR complexity indicators are robustly correlated with per capita GDP. These indicators can function as perdictors of long term future growth when the country sample is restricted to those countries exhibiting low complexity variablility. Adding some control variables normally used in growth regressions also increases the indicators capacity of predicting long term growth. Results presented here do not support the conclussion that the indicators can be used as predictors of short term growth.

The organization of this work is as follows. Section 2 overviews the main works exploring the relationship between specialization and growth and identifies among them the main ideas that provide theoretical support for the use of complexity indicators like the ones proposed by Hidalgo and Hausmann (2009) to predict future growth. Section 3 presents the database used and Section 4 present the MR indicators and their most important features. Section 5 introduces the filter applied by this work to select the different country samples used in income and growth regressions and Section 6 presents the main control variables to include. Results for the different exercises performed in this work are shown in Section 7. Finally Section 8 concludes.

2 Related literature

The debate on which may be the driving forces behind the fact that specializing in some products yield higher growth than others can be organized differentiating two broad groups. On one hand, there is a group of works arguing that the main difference comes from the international demand (see for example Thirlwall (1979), Thirlwall and Hussain (1982), Pasinetti (1981) or more recently Araujo and Lima (2007)). On the other hand, there is a second group of works claiming that the reason relies within countries and is supply-based. This latter literature is the one more closely related to the ideas behind the MR indicators. In this part of the literature the argument is mostly based upon the assumption that some sectors can absorb more technological advances than others and this implies that countries where those sectors are relatively more important have larger growth rates in the long term.

One of the earliest examples of the supply-based literature can be found in Baumol (1967). The paper presents a model with two sectors, a progressive sector (that incorporates innovations at a high rate) and a stagnant sector (that does it at a lower rate), and shows that with such a setting the progressive sector will decrease its relative costs and prices as it incorporates technology and the stagnant sector will tend to vanish. Structural change will then take place as the economy develops.

The endogenous growth literature also pointed to the fact that there are some productive processes that contribute more than others to growth. In Romer (1986), Aghion and Howitt (1992) and Grossman and Helpman (1991) the authors present models with technologically advanced sectors where structural change is the main driver of long term growth. In their models, structural change comes through the accumulation of new capital, the increase of labor division or greater quality goods. They all agree that structural change (i.e. changing what you are producing) influences growth through technological externalities, indivisibilities and complementarities in productive processes.

Within the endogenous growth literature Fagerberg (1994) presents a clear explanation of the implications of abandoning the assumption of technology being a *free* good. This assumption is behind the prediction of equal growth rates across countries in the neoclassical setting: if growth is mostly explained by technological progress and this is a shared good across nations, then every integrated economy will eventually grow at the same rate in the long run, and the only possibility for observing differences in growth rates is during transitional dynamics. By allowing technological progress not to be perfectly transferable, the possibility for everlasting heterogeneous growth rates arises.

There are many recent works pointing at the fact that most rich countries undergo a process of diversification while growing. Imbs and Wacziarg (2003), Klinger and Lederman (2004 and 2006) and Cadot et al. (2011) show that diversification is a common feature at early stages of development: poor countries grow by diversifying their production. They also point out that more advanced stages of development bring some degree of specialization.

The MR indicators this work tests are one of the outcomes of a research line that can be considered to begin with Hausmann and Rodrik (2003). The authors underline that specializing on some products can bring higher growth than specializing in some others focusing on the concept of *cost discovery*: to undertake a new production within a country it is required that one pioneer firm takes the first step and discovers what the real costs of production are (i.e. invests in cost discovery). This pioneer has private losses if it fails but generates spills over for the entire economy if it succeeds as new information will be available to all firms. The resulting externality implies that the activity of cost discovery will be under-provided in decentralized economies (compared to the centralized economy solution) unless the state implements a policy to make firms internalize it. This is, according to the authors, behind the different development path between South-East Asian countries and Latin-American countries in the second part of the last century: while the first group had the state aiding the private sector in cost discovery activities the second group did not.

In Hausmann et al. (2007) the authors argue that some products have a higher level of associated productivity than others and therefore specializing in these products will bring higher growth. This constituted a strong argument for state promoted structural change: if rich countries export rich country products then in order to become a rich country an effort should be made to reach production of such goods. In order to evaluate empirically which are those products that are related with higher income levels the authors proposed an indicator ($PRODY$) that assigns to each product the per capita GDP of countries that export it with revealed comparative advantages. Then they built another indicator that approaches an economy's wealth in sophisticated goods ($EXPY$) by computing the average $PRODY$ of each country's exports basket, and showed that this indicator is a good predictor of future growth.

Going one step further, Hausmann and Klinger (2007) and Hidalgo et al. (2007) proposed an index of distance between any two products (*proximity*). To construct this index they use trade data to measure how much exporting product a is contributing to the probability that product b is exported as well by a country. The authors argued that two products with high *proximity* are likely similar in terms of their productive requirements. The matrix that gather a measure of *proximity* for every pair of products constitutes what they called the *Product Space*. Hidalgo et al. (2007) shows that more densely connected products in the *Product Space* were also products having a greater valuation in terms of $PRODY$, so the conclusion was very clear: countries producing these goods are countries that have it easier to grow since not only their current production is correlated with high income levels but their diversifying options are correlated with high income as well.

In addition these works suggested a measure of distance between any non produced product and the current production of the economy which they called *density*. This index, when used along with $PRODY$, has great potential for policy-making. After all, if it is possible to measure how easy it is for a country to produce a new good and also to establish how much each product contributes to per capita income, then it is straightforward to obtain a clear idea of which new products should be stimulated and which should not. The policy-recommendation quality of the indicators did not go unnoticed. Many documents were written using them with that purpose (see for example Hausmann and Klinger (2006), Record and Nghardsaysone (2010), Abdon and Felipe (2011) or Jankowska et al. (2012)).

Although $PRODY$ and $EXPY$ represented original contributions there were possibilities for their improvement. The fact that the indicators used per capita GDP in the valuation of products' sophistication meant that there was some degree of endogeneity embedded in the conclusion that rich countries were exporting rich country products: it could be the case that there is no real valuable attribute that is intrisic to high $PRODY$ products besides being exported by rich countries. Under such an interpretation one it would be interesting to know why, other than possesing valuable capabilities, is it that high $PRODY$ products are only exported by rich countries. Still, critiques were legitimate enough to inhibit the broad use of the proposed estimators. This motivated Hidalgo and Hausmann (2009) to present a new set of indicators, further developed in Hidalgo (2009) and Hidalgo and Hausmann (2010), in their *Method of Reflections*, which drops the use of per capita GDP to evaluate products' sophistication. Instead, the new proposal exploits to the fullest the information inside the global trade matrix: the authors claim to achieve measures of product sophistication and economic complexity by looking only at who is exporting what.

In Hidalgo (2009) the author shows that MR indicators of product sophistication and economic complexity are highly correlated with $PRODY$ and $EXPY$ respectively, and in Hidalgo and Hausmann (2009) the authors present evidence supporting the properties this work tests , i.e. that MR complexity indicators are highly related to countries' income and help predict their future growth both in the long run and in the short run. Serving the same purposes than $PRODY$ and $EXPY$ but with less shortcomings this new set of indicators are even more appealing than those previously suggested by Hausmann et al. (2007) which make these tools very likely to be used for policy making throughout the world. This possibility provides enough justification to the task proposed in this work.

Finally Hausmann and Hidalgo (2011) presents a model to formalize the theoretical ideas accompanying the indicators developments. They conclude that countries with fewer capabilities have lower incentives to accumulate new ones. This is because the pay-off they get from an extra capability is lower compared to the one that a country with many capabilities gets, as it will enable the production of a smaller number of new products. This is called by the authors the *quiescence trap* and implies a sort of increasing returns to diversification that helps explain the divergence in growth across countries.

2.1 A discussion on the main concepts stemming from the literature

The concept of economic complexity is in Hidalgo and Hausmann (2009) related to the amount of technological capabilities a country has. Having many diverse capabilities implies having what it is required to produce many different products. Similarly, a product is considered sophisticated when it requires a great amount of different capabilities in its production process. In Hausmann and Hidalgo (2011) the authors show evidence suggesting that poor countries export a small quantity of products that many other countries export while rich countries export those products plus some others that are less frequently exported. This suggests, as the authors point out, that poor countries have accumulated fewer and more commonly spread capabilities than richer countries.

The authors do not provide a precise definition of *capabilities*, but if it includes anything that is fundamental for the production of at least one good, then it can be concluded that the concept is very broad. Tangible things like having certain natural resource or machine are necessary for the production of some products, but also non-tangible things like having an innovative environment or solid institutions might be necessary for the development of some others. It can therefore be seen how this line of research is easily connected with about any of the different branches within the growth literature: geography, demography, institutions, learning by doing processes and a large etcetera.

Hidalgo and Hausmann (2009) explain that non-tradable capabilities are the ones responsible for a country's productivity. But, as will be clear in section 4, the capabilities measured by the MR need not be non-tradable. It is actually more suitable to include tradables into the concept as well since this would help explaining how some countries have acquired so many capabilities over time.

It is also important to notice that the MR approach implicitly points at the fact that some capabilities are more valuable than others. If the value of a given capability is the quantity of production processes in which it has a vital role, then a multi-purpose capability is going to be much more valuable than a very specific capability that only plays a role in a limited number of production processes.

The main theoretical ideas behind the MR seem very much related to some of those previously underlined by the neo-schumpeterian or evolutionist literature on technology and innovation processes. In some of the most renown works related to this literature it is possible to find descriptions of the main characteristics behind the innovation process that are very close to what the authors this work follows are using. In Dosi (1982) for example, the author defines technology as the accumulated pieces of knowledge a country has and explains that these pieces of knowledge can be the result of a physical innovation or simply the outcome of learning. The author emphasises that the pieces of knowledge that form an economy's technology can be something already applied to production or not, so they determine current and also future production. It is easy to see the similarity between the concept of capability and these pieces of knowledge.

Dosi (1982) also provides a characterization of the technological development process that resembles the approach MR authors give to the process of capabilities' accumulation and is also very similar to what endogenous growth authors have in mind. First, Dosi (1982, p. 154) explains that there are strong complementarities between different pieces of knowledge, which means that the accumulation or depletion of one of them can foster or hinder the accumulation of some other. Another characteristic is that the accumulation of knowledge is cumulative to some extent (i.e. a region incorporates knowledge upon what it already has) and this implies that technological trajectories with some degree of path dependence will emerge. Finally the author states that it is not possible to evaluate *ex ante* how fruitful any chosen technological trajectory will be, so any technological choice has some degree of uncertainty. It is noticeable how these ideas resemble those already mentioned by Hidalgo and Hausmann (2009), Hausmann et al. (2007) and Hausmann and Rodrik (2003).

3 Data

This work's main source of information is export data from the *Base pour l'Analyse du Commerce International* (International Trade Database at the Product Level, BACI from here on), as reported by Gaulier and Zignago (2010) from CEPII. The BACI reports values and quantities of product exports from country i to country j in the first version of the Harmonized Commodity Description and Coding System (HS0) at a six-digit aggregation level for the period 1995-2007. This database uses UNCOMTRADE data and applies to it an harmonization method to match records declared by the exporter with those made by the importer as detailed in Gaulier and Zignago (2010).

UNCOMTRADE data does not include flows below 1,000 US dollars but accounts for more than 95% of total world trade. In order to have the same countries and products in every year, it is necessary to drop some observations[1]. The final sample used here is composed of 178 countries and 4948 products for each of the 13 years of the period.

The use of this database constitutes an important methodological departure from what is used in Hidalgo and Hausmann (2009) to test MR indicators' properties and is one of the main changes proposed here to test their robustness. Hidalgo and Hausmann (2009) use as main source of information Feenstra et al. (2005) database which gathers UNCOMTRADE, Standard International Trade Classification (SITC), revision 4, data at a four-digit aggregation level and also matches export and import reports covering the period 1962-2000[2]. For the years after 2000 they used raw UNCOMTRADE data.

The selection of the BACI was made pursuing the idea that more disaggregated data can feed the MR indicators with more accurate information: when evaluating product sophistication in terms of the amount of capabilities required to produce one good, it is more suitable to use the most disaggregated data available since this allows a sharper distinction between the capabilities required for each product. For example, to have six-digit data allows to differentiate between product 847010 which is the code for electronic calculators and product 847050 which denotes cash registers, or between product 901710, drafting tables and product 901730, micrometers, callipers and gauges. As will be clear in the next section there is valuable information in the fact that some country is for example exporting both products in one of these pairs but some other only exports one of them, and this is exactly the type of information MR indicators nourish from.

Other auxiliary data comes from the World Development Indicators (WDI) as reported by the World Bank[3], except data on population, per capita GDP at PPP and trade openness for which the work uses data from the Penn World Tables 7.0 (PWT) as reported by Heston et al. (2011). This is because PWT provides information in those variables for a greater number of countries in each year of the time span used here. In fact the only case of missing values in per capita income and population belong to Timor-Leste in the period 1995-1999.

Notice the time span allowed by the use of the Feenstra et al. (2005) dataset is much longer than the one in the BACI (although only data starting from 1985 is used by Hidalgo and Hausmann (2009) when testing the three properties this work focuses on). This constitutes an important shortcoming in the use of the later. In particular this work is not going to be able to test the performance of MR indicators as predictors of future growth in a 20-year period as done in Hidalgo and Hausmann (2009). However this should not prevent this work to find that MR indicators significantly predict future growth given that MR indicators are supposed to perform well as predictors of

[1]Countries being-left out are the Vatican City, Serbia and Montenegro, San Marino and the Occidental Palestinian Territories which represent less than 0.09 % of total trade in the sample for each year considered here.

[2]The authors explain they have checked the validity of their results with different databases. In particular they used UNCOMTRADE HS data at four-digit level (covering 1241 products and 103 countries) and North American Industry Classification System (NAICS) with data at six-digit aggregation level (318 products, 150 countries). Unfortunately results stemming from the use of these datasets are not presented. Although the use of the NAICS has the same aggregation level than the BACI the quantity of products contained in that dataset is much lower.

[3]Available at http://data.worldbank.org/data-catalog/world-development-indicators.

future growth in the short term as well (Hidalgo and Hausmann (2009) find significant results for MR complexity indicators in growth regressions using only five-years periods). The use of a shorter time period has the benefit of being able to work with a larger quantity of countries: while this work uses trade and income information for 177 countries for the whole period, the same can only be said for 95 countries in the period analysed in Hidalgo and Hausmann (2009).

The use of the HS classification instead of the SITC classification does not imply an important change, results are not expected to differ due to this.

4 The method of reflections

4.1 Construction of the indicators

This section explains how MR indicators are constructed strictly following the proposal of Hidalgo and Hausmann (2009) and Hidalgo (2009) unless indicated otherwise. The first step is to consider as exported by a country only those products for which the country has revealed comparative advantages. By doing this MR indicators will consider only those products for which the country has proven to be competitive in world markets. The authors propose to use Balassa's revealed comparative advantage index (Balassa, 1986), $RCA_{c,p}$, which is computed as follows:

$$RCA_{c,p} = \frac{\frac{x_{c,p}}{\sum_p x_{c,p}}}{\frac{\sum_c x_{c,p}}{\sum_{c,p} x_{c,p}}} \tag{1}$$

where $x_{c,p}$ is the export value of product p by country c. The $RCA_{c,p}$ gives the importance of a product p in country c's export basket relative to the importance that the same product has in worldwide trade. The importance a product has for a country could be measured differently, it could include for example different weights for products with diverse intrinsic or strategic values. This work follows strictly MR author's proposal in order to keep results comparable.

A threshold that separates those products that are exported with comparative advantages by a country from those which are not must be established. Then it is possible to build a matrix of countries and products in which every component follows the next rule:

$$M_{c,p} = \begin{cases} 1 & if \quad RCA_{c,p} \geq R^* \\ 0 & \quad\quad otherwise \end{cases} \tag{2}$$

Authors propose R^*=1 as threshold which means the MR indicators will consider as exported by a country only those products that have a higher or equal weight in the country's export basket than in global trade. As will be argued in section 4.4 the R^* threshold can be considered an arbitrary choice and will be target of modifications in this work.

Using the $M_{c,p}$ matrix, it is possible to build the MR's simpler indicators following:

$$k_{p,0} = \sum_{c=1}^{N_c} M_{c,p} \tag{3}$$

$$k_{c,0} = \sum_{p=1}^{N_p} M_{c,p} \quad (4)$$

being N_p the total number of products considered (here $N_p = 4948$) and N_c the total number of countries used in the dataset ($N_c = 178$). Equation (3) establishes that $k_{p,0}$ measures the number of countries exporting product p, so it is a measure of that product's ubiquity. Indicator $k_{p,0}$ can also be seen as a simple measure of product p's sophistication: if a product is exported by few countries it might indicate that technological capabilities required to do so are rare. Similarly, equation (4) shows how $k_{c,0}$ gives a measure of the number of products exported by country c, and so it measures country's diversification. This indicator can also be seen as a very simple index of country c's complexity, since a diversified economy must have acquired many technological capabilities to be successful in many productive processes.

But these are only rough approximations to the concepts of product sophistication and economic complexity. Suppose we have two different countries both with similar diversification levels, but one is small and has achieved its diversification level by acquiring different capabilities and the other one is large and the only reason it has a diversified export basket is because of its size. It would be desirable for an indicator of economic complexity to discriminate between these two very different countries. The same thing can happen when evaluating product sophistication. This is why the MR proposes to complement the initial information about diversification and ubiquity by exploiting to the fullest the information contained in trade data to better establish each country complexity and each product sophistication. This is done by following the iterative process described in the following equations:

$$k_{p,n} = \frac{1}{k_{p,0}} \sum_{c=1}^{N_c} M_{c,p}.k_{c,n-1} \quad (5)$$

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_{p=1}^{N_p} M_{c,p}.k_{p,n-1} \quad (6)$$

where n is the number of iterations used to define indicators $k_{p,n}$ and $k_{c,n}$. The result of these iterations yields two vectors of indicators: on the one hand vector $k_p = \{k_{p,0}, k_{p,1}, ..., k_{p,n}\}$ defined for each product p, and on the other vector $k_c = \{k_{c,0}, k_{c,1}, ..., k_{c,n}\}$ defined for every country c.

4.2 Interpretation of the indicators of economic complexity

This work will focus on k_c components since its aim is to test whether these indicators are related to income and can predict future growth. Let's consider first the simplest cases. Equation (4) shows that $k_{c,0}$ is only counting the number of products exported by country c, and by equation (3), $k_{p,0}$ is the counting of countries exporting product p. Following equation (6), it is possible to see that $k_{c,1}$ is the average ubiquity of products exported by country c, while equation (5) shows that $k_{p,1}$ is the average diversification of countries exporting product p. Moving to the second iteration stage, $k_{c,2}$ is the average diversification of countries exporting products that country c exports as well. Similarly, $k_{p,2}$ is the average ubiquity of products exported by countries also exporting product p.

By comparing $k_{c,i}$ results for very different countries it is possible to get a better idea of how the MR uses trade data to get rid of distortionary effects and reach an evaluation of economic complexity. Let's take a look for example at the different trajectories that India and Japan follow as i grows for $k_{c,i}$, using data for the last year in the sample (2007). Results show that $k_{India,0} = 1693$ which ranks India in the 7^{th} place of the ranking of 178 countries, above Japan that has a $k_{Japan,0} = 1259$ and its in the 17^{th} place. This could be considered as a first rough approximation to economic complexity, but of course the level of $k_{c,0}$ only indicates the quantity of products country c is exporting with $RCA \geq 1$, so it is probably very much influenced by country size. To really evaluate economic complexity more information is required.

With $i = 1$ results are $k_{India,1} = 13.85782$ and $k_{Japan,1} = 11.37679$. This is the result of weighting each exported product by the number of countries exporting that product, which gives an idea of how difficult it is to do what the country under evaluation is doing. Notice that even though India exports more products, the average number of countries exporting what Japan is exporting is lower which could imply that Japan is actually a more complex economy. This result is less related to country size but on the other hand it does not say much about the quantity of products being exported by each country. It could be the case that Japan's exports are rare mainly because it is a country that has a very rare natural resource and it concentrates its exports among low sophisticated derivatives of that resource. This situation would not be close to the concept of economic complexity defined by the literature. So again, at this level of iterations the information extracted from trade data is not providing something that could be considered close enough to the idea of complexity.

Now, when $i = 2$ results are $k_{India,2} = 1011.534$ and $k_{Japan,2} = 1126.783$ ranking India in the 22^{th} place while Japan reaches the 2^{nd} position. The average diversification of countries exporting what Japan exports is greater than the same figure for India. This gives the idea that the Japanese economy can be considered more complex than the Indian economy. The $k_{c,2}$ is clearly less correlated with distortionary features like country size or specialization in extraction-type products than the less iterated components of the k_c vector. This example has shown that an indicator with $i = 2$ is closer to the idea of economic complexity than other indicators where $i < 2$.

The interpretation of the MR indicators gets harder as the number of iterations is increased, since every vector component gathers information from the preceding components. But this also means that elements coming from higher iterations will have more information and their correlation with economic complexity or product sophistication will be stronger. Therefore, every component of vector k_c can be considered as a measure of an economy's complexity and the higher the iteration the more information it has.

4.3 Main features of the indicators of economic complexity

If highly iterated k_c indicators approach economic complexity one could expect these indicators to be explained by different kinds of capabilities. Table 1 shows results of pooled OLS estimations with even k_c indicators as dependent variables, and different indexes representing different kinds of capabilities as explanatory variables. Each indicator from the k_c vector is standardized in order to make coefficients comparable with each other. White's estimator is used to obtain robust standard errors.

Table 1: Pooled OLS estimations of different variables explaining k_c even indicators

	1	2	3	4	5	6	7	8	9	10
	$sk_{c,0}$	$sk_{c,2}$	$sk_{c,4}$	$sk_{c,6}$	$sk_{c,8}$	$sk_{c,10}$	$sk_{c,12}$	$sk_{c,14}$	$sk_{c,16}$	$sk_{c,18}$
lpop	0.360***	0.200***	0.148***	0.096***	0.053**	0.039*	0.035*	0.034	0.033	0.033
	(26.39)	(19.22)	(11.75)	(5.16)	(2.54)	(1.84)	(1.65)	(1.60)	(1.59)	(1.59)
openk	0.003***	0.002***	0.002***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(5.62)	(3.97)	(3.40)	(3.65)	(3.71)	(3.77)	(3.80)	(3.81)	(3.81)	(3.81)
inflation	-0.002*	-0.001**	-0.001***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(1.66)	(2.36)	(3.21)	(4.25)	(4.25)	(4.09)	(4.03)	(4.01)	(4.01)	(4.01)
real_int	-0.010***	-0.008***	-0.009***	-0.012***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***
	(4.38)	(4.61)	(4.79)	(4.67)	(4.34)	(4.21)	(4.18)	(4.17)	(4.17)	(4.16)
gov_C	0.048***	0.036***	0.031***	0.009	-0.010	-0.016**	-0.018***	-0.018***	-0.018***	-0.018***
	(12.89)	(10.05)	(7.26)	(1.55)	(1.49)	(2.46)	(2.74)	(2.82)	(2.85)	(2.85)
ltertiary	0.259***	0.370***	0.385***	0.268***	0.133***	0.085***	0.072**	0.068**	0.067**	0.067**
	(15.21)	(22.56)	(18.92)	(9.64)	(4.45)	(2.85)	(2.40)	(2.27)	(2.24)	(2.23)
indvalue_p	0.003	0.008***	0.004	-0.009**	-0.015***	-0.017***	-0.017***	-0.017***	-0.017***	-0.017***
	(1.01)	(3.87)	(1.35)	(2.29)	(3.55)	(3.86)	(3.93)	(3.95)	(3.96)	(3.96)
natural_res	-0.030***	-0.031***	-0.020***	0.001	0.013***	0.016***	0.017***	0.017***	0.017***	0.017***
	(11.10)	(13.62)	(7.54)	(0.37)	(3.74)	(4.59)	(4.79)	(4.84)	(4.86)	(4.86)
Obs	1052	1052	1052	1052	1052	1052	1052	1052	1052	1052
Adj. R-sq	0.66	0.69	0.55	0.18	0.06	0.05	0.05	0.05	0.05	0.05
F-test	218.17	241.84	139.25	29.23	9.51	8.96	9.22	9.33	9.36	9.37
Prob> F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes. Pooled OLS estimations using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable $sk_{c,i}$ is the i-th component of the k_c vector divided by its standard deviation. *lpop* is the logarithm of total population, *openk* stands for economic openness (%) at 2005 constant prices, *inflation* is the inflation rate, *real_int* is the real interest rate (%), *gov_C* is the share of GDP destined to consumption, *ltertiary* is the logarithm of the gross rate of enrolment in tertiary education, *indvaluep* is the share of GDP comming from the secondary sector, *natural_res* is the share of GDP comming from rents over natural resources. All explanatory variables were extracted from the WDI Database except for *openk* which was taken from PWT. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

The logarithm of the gross rate of enrolment in tertiary education (*ltertiary*) is used to approximate human capital, the inflation rate (*inflation*) is used to approximate macroeconomic stability and the real interest rate is incorporated as a measure of the financial cost of engaging a productive project (*realint*). Other variables were included like a measure of economic openness (*openk*), the share of GDP destined to government consumption (*gov_C*) which measures the importance of the public sector in the economy, the share of GDP coming from industry (*indvalue_p*) and total natural resources rents as a percentage of GDP (*natural_res*). Finally a control for the size of the economy was included which is measured by the logarithm of the economy's population (*lpop*).

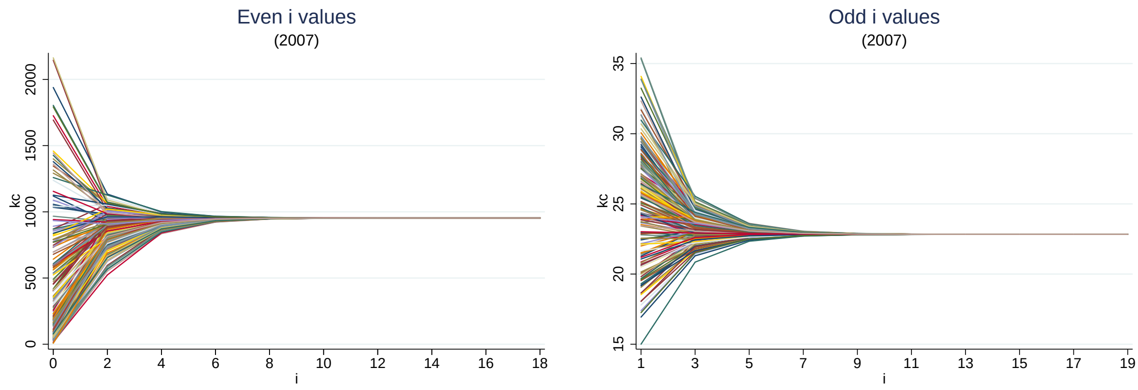
It is remarkable how, although all regressions are significant as a whole at 1%, the percentage of the k_c indicators explained by these variables is high for low iterated indicators, and low for high iterated indicators. These results present evidence supporting the hypothesis that highly iterated indicators are much richer and capture many more dimensions of economic complexity than less iterated indicators, since the greatest part of their variation is explained by unobservables.

Table 1 also shows that, as expected, population looses significance as i increases. Moreover, the worse the macro-environment (the higher the inflation rate or the real interest rate) the lower all k_c indicators will be. Finally, human capital and openness are both always positive and significant at 1% for any iteration level.

Figure 1 shows the value of every $k_{c,i}$ indicator for each country as i grows (for even and odd indicators separately). Only data from 2007 has been used to simplify the exposition but the same picture emerges every year. As shown in both panels, when i grows the MR indicators converge to their mean, which is not surprising given that they are built as averages of other averages. The figure also shows that odd components inside a vector will converge to a certain mean while even components tend to another. This is due to the way indicators are constructed: in building $k_{c,i}$ information from $k_{p,i-1}$ is used but information from $k_{c,i-1}$ is not. Thus, odd components do not contribute with any information in the construction of even components within the same vector, and vice versa.

Figure 1 also shows that the final mean for even components (i.e. that of $k_{c,18}$) is 953.63 while that for odd components (corresponding to $k_{c,19}$) is 22.84. This shows an evident difference in the units of measure of each of these families, given by the fact that even components are computing averages of products while odd components average countries (remember the interpretation done for equation (6) at different i levels).

Figure 1: $k_{c,i}$ results for all countries as i grows

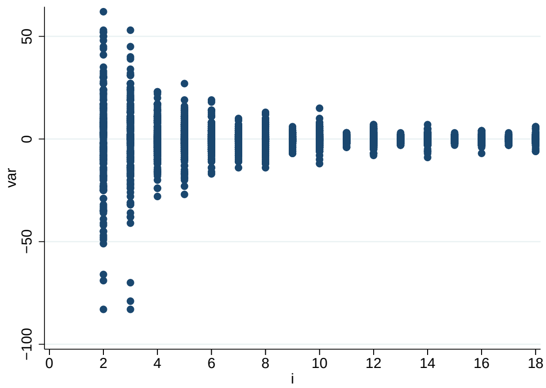


Notes. The x-axis indicates the degree of iteration i of the $k_{c,i}$ indicator, while the y-axis measures the value of the corresponding indicator. Each line comprises values for one country. Only data from the last year of the sample (2007) is included. Even indicators (left) and odd indicators (right) are presented separately.

The convergence-to-the-mean effect implies that highly iterated indicators have a very narrow range (the greatest standard deviation among all years for $k_{c,18}$ is of 0.007 in 1995 when the minimum value is 930.98 and the maximum is 931.02). Despite this, the small differences between countries that exist in highly iterated indicators yield a more stable ranking of countries than those stemming from less iterated indicators. Figure 2 shows the variation in each country ranking positioning as the iteration increases. The vertical axis measures the change in the positioning each country experiences when changing the indicator from $k_{c,i-2}$ to $k_{c,i}$. This figure also uses only data for 2007 but the same conclusion arises for every year. Notice how the sorting tends to stabilize as i increases. This fact implies that when i is low, information extracted from trade data has an important marginal contribution, while when i is high the marginal informational contribution of an extra iteration is very low.

The different nature of even and odd indicators implies that even indicators are positively correlated with per capita income (the more products being exported the higher economic complexity should be) and odd indicators are negatively correlated with per capita income (more countries exporting what country c exports means that capabilities required to do so are not rare and therefore the lower country c's complexity). Naturally, this also means that the two families of indicators are negatively related to each other. Figure 3 shows the correlations of indicators from each family with per

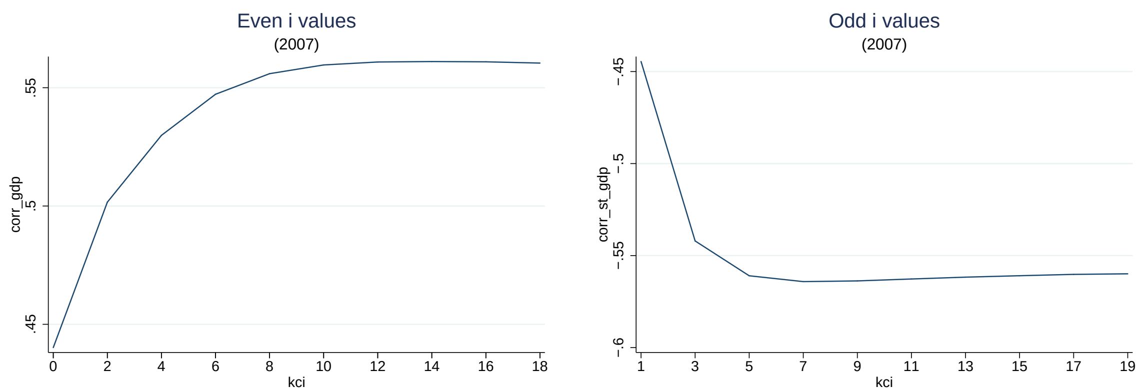
Figure 2: Ranking change between $k_{c,i-2}$ and $k_{c,i}$ for all countries as i grows



Notes. The x-axis indicates the degree of iteration i of the $k_{c,i}$ indicator, while the y-axis measures the difference $k_{c,i-2} - k_{c,i}$ for each country. Each dot shows that difference for a country and an iteration value i. Only data from the last year of the sample (2007) is included.

capita income as the number of iterations increases. It is possible to see that the correlation with country's per capita GDP is much higher for higher iterated indicators implying that the sorting stemming from higher iterations could be considered as the one that better reflects countries' complexity.

Figure 3: Correlation between $k_{c,i}$ indicators and per capita GDP



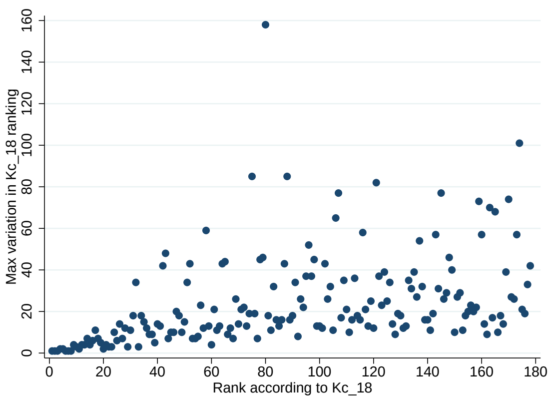
Notes. The x-axis indicates the degree of iteration of the $k_{c,i}$ indicator, while the y-axis measures the correlation between that indicator $k_{c,i}$ and per capita GDP for the entire sample of countries. Only data from the last year of the sample (2007) is included. Even indicators (left) and odd indicators (right) are presented separately.

Although even and odd indicators nourish from different sources of information, highly iterated indicators from both families yield very similar rankings for countries and products. In fact the absolute value of the correlations between $k_{c,18}$ and either $k_{c,17}$ or $k_{c,19}$ is greater than 0.995 for every year in the sample used. Strong correlations can also be found between lower iterated indicators, e.g. the absolute value of the correlation between $k_{c,4}$ and $k_{c,5}$ is greater than 0.91 for every year.

When looking how countries are ranked according to their $k_{c,18}$ throughout the years, it appears to be some cases of clear upward or downward trends that go along intuition. Countries like Malaysia, Thailand or Viet Nam, which are well known cases of increasingly complex economies, exhibit a markedly upward trend in their ranking positioning over the years. Most countries however do not present such a clear trend positioning variation from one year to the next is very frequent, being the variations very strong especially among low complexity countries. Remember that economic complexity is, according to the literature, a long term phenomenon: countries build their economic complexity through a long and costly process of acquiring (and destroying) capabilities. This implies that strong and sudden changes should not be usual in an indicator that approximates economic complexity.

Figure 4 plots the maximum variation each country experiences in the $k_{c,18}$ ranking from one year to the next (called v) against its $k_{c,18}$ level in 2007. The figure shows how this variation is increasing from the first positions until the 80^{th} place and remains stable after that. This means that more complex countries, i.e. those occupying the higher places in the $k_{c,18}$ ranking, present a more stable position in it. Given that the dataset used here considers matched values from exporters and importers to reach the final value of a given purchase, it is hard to disregard the observed volatility in the lower positions of the $k_{c,18}$ ranking on the basis of lack of trust in these countries' reports. Rather it should be pointed out that relying solely on export data, MR indicators are vulnerable to sudden changes in export figures (which can happen in contexts of important changes in trade or industrial policies, political or institutional environment, etc.). This prevents the achievement of an accurate measure of economic complexity for countries where these kind of changes occur. Explaining the situation in each of these volatile countries escapes the aim of this work, but it will be important to keep in mind that there is an important source of noise here that should be addressed.

Figure 4: Maximum changes in $k_{c,18}$ ranking for each country



Notes. The x-axis shows the ranking positioning each country has according to $k_{c,18}$ in the year 2007. The y-axis measures the maximum variation each country shows in that ranking between any two consecutive years. Each dot represents a country.

The same figure also shows that it is not usual to see changes in more than 40 positions and most countries never face a change of 20 positions. Table A.1 presents the list of all 178 countries sorted by v (computed under the benchmark case $R^* = 1$). The table also includes population (in thousands), per capita GDP at PPP ($rgdpl$) and the Hirschmann-Herfindahl index of export diversification (HH). This later index will be properly defined in equation (7) of section 6 but it should be pointed that it ranges between 0 and 1 and that a value close to 1 indicates a very concentrated export basket. The table allows to conclude that those countries having very large changes (say v >40) are either very small, very concentrated or have a well known record of economic instability during the period considered here.

4.4 Limitations of the MR

MR indicators have some important limitations that should be kept in mind. First, to measure economic complexity by looking at what a country is exporting implies assuming that every country exports with revealed comparative advantages all products for which it has the required capabilities. But of course this might not be the case. In a context of uncertainty about the results of a given enterprise, as described in Hausmann and Rodrik (2003), some productions might have not being discovered yet, although required capabilities are available in the economy. It could be argued that MR indicators can be undervaluing complexity in countries that have accumulated a great number of capabilities more recently, since these economies probably had less time to learn what they can do with them.

The fact that indicators are considering as exported by a country only those products with $RCA_{c,p} \geq R^*$ is another limitation since it implies that they are also ignoring many processes that might be adding complexity to the country's economy although their RCA level is not that large.

It should also be noticed that the original proposal for the construction of MR indicators, which uses only product exports, is completely ignoring production of goods for the domestic market and services. This is a strong impediment when trying to get closer to an economy's technological capabilities, since both kinds of production are able to add technological learning into the productive structure of a country. It is not easy to replace the database used since records on domestic market production and services are not available with the same level of comparability and with the same periodicity as product exports are. Hausmann and Hidalgo (2011) addressed this issue by constructing the MR indicators upon a database of total production from Chile and concluded that results obtained from the MR indicators are not strongly influenced by the fact that it uses only product export data.

None of the limitations stated above imply that MR indicators are useless. Rather they are pointing out that information being considered might not be complete. The fact that MR indicators might be underestimating economic complexity does not prevent the indicators to increase the available knowledge regarding countries' complexity, which can be a very useful thing.

5 The filter used

The volatility of countries' $k_{c,18}$ across years points at some existing noise in the link between the information MR indicators are considering and what they are measuring. When a country displays a volatile index, which of its different values should be considered to evaluate its true economic complexity?

This noise can harm regression analysis greatly and so this work proposes to use a filter of countries based on the maximum $k_{c,18}$ ranking change from one year to the next (previously denoted v). That is, countries that have a ranking change between any two years greater than a certain number are being dropped from the analysis. This choice for the filter provides this work with a flexible criterion in which the filter threshold can be changed and it is possible to analyse the impact this has on the results[4]. The filter can also help to put forward the conditions the sample must fulfil in order for the MR indicators to be useful predictors of future growth.

As the limit imposed to v (called v^*) is decreased, increasingly significant results are expected since this implies getting rid of the noise brought by excess variability in the complexity indicator. Of course if the limit is too low then too many observations will be dropped and the relationship will not appear to be so clear. The reader can use Table A.1 to check which countries are being included in the set of observations of every regression performed here under the benchmark case. As mentioned in Section 4.3 the large majority of countries remain when v^* is set at 40. The filters actually used in most cases are, as will be shown, less restrictive than that.

6 Main control variables

The only two control variables used in regressions performed by Hidalgo and Hausmann (2009) to show the main three properties that are going to be tested here are the Hirschmann-Herfindahl (HH_c) concentration index and Theil's Entropy index (E_c) of diversification (Theil, 1972). It is not surprising that the authors are not using more control variables in their regressions since, as explained, MR complexity indicators are supposed to be measuring capabilities in a broad sense. The inclusion of most control variables usually included in growth regressions to capture different kinds of capabilities (different resources abundance, geography, institutional quality, etc.) can therefore be redundant.

The main purpose of introducing export diversification (or concentration) controls in income and growth regressions is to show that MR complexity indicators are able to explain a larger percentage of the dependent variable's variability compared to what diversification indicators can explain.

The HH_c index is a standard measure of market concentration but can be applied to a country's export basket to evaluate its concentration level, as is done here. The index is defined as follows:

$$HH_c = \sum_{p=1}^{N_p} \left(\frac{x_{c,p}}{\sum\limits_{p=1}^{N_p} x_{c,p}} \right)^2 \tag{7}$$

HH_c ranges between 0 and 1. As can be seen, the term in brackets is the share of product p in country's c export basket, so the higher the index the more concentrated the exports of country c are in fewer products.

[4]There are many ways to perform the same task. A different alternative could have been to base the criterion upon the quantity of times a country changes from one quantile of the $k_{c,18}$ distribution to another, but that would imply a greater probability of deletion of countries close to the quantile limit. The filtering choice made here avoids the latter problem and allows flexibility, without loosing simplicity and transparency.

The E_c index is a widely used measure of inequality and can be applied to measure the diversification of a country export basket when defined as follows:

$$E_c = -\sum_{p=1}^{N_p} \left(\frac{x_{c,p}}{\sum_{p=1}^{N_p} x_{c,p}} \right) .log \left(\frac{x_{c,p}}{\sum_{p=1}^{N_p} x_{c,p}} \right) \tag{8}$$

E_c is always positive and high values of the index implies that country c has a highly diversified export basket.

7 Results

This section will submit to different robustness checks three of the main properties of the MR complexity indicators: 1) that they are related with countries' current income levels, 2) that they predict future long term growth, and 3) that they predict future short term growth. These properties are originally tested in Tables S6-S10 of the Appendix in Hidalgo and Hausmann (2009). Following the authors, focus will be given to even and highly iterated indicators from the k_c vector. As explained in section 4.3 highly iterated indicators present a stronger correlation with per capita income than less iterated indicators. Also, as i grows correlation between even and odd indicators go to 1, which makes the exposition of odd indicators redundant. Special focus will be given to $k_{c,18}$ which is the highest iterated even indicator computed by the authors and constitutes the main reference used by them when economic complexity needs to be evaluated (see for example Hidalgo (2009)).

For each of the properties the procedure will be as follows. First, results obtained following the same methodological steps as in Hidalgo and Hausmann (2009) will be presented. The only methodological departure in this first step is the use of the more disaggregated database. If no significant results are found, the next step is to explore conditions under which significant results could be found. This is done by filtering for countries that present too much volatility according to the $k_{c,18}$ indicator, by including more control variables in the regressions and by dropping outliers in countries' income distribution. When significant results appear the last step will be to test how much they depend on what is considered here to be the most important methodological decision made in the construction of MR indicators, i.e. the choice of the R^* parameter.

7.1 Relationship between MR indicators and income

The first exercise to do, is follow Hidalgo and Hausmann (2009) by computing simple cross-section OLS regressions for a single year where the log per capita income at purchasing parity power is the dependent variable and a measure of complexity is the main regressor. Different components of vector k_c are used as measures of complexity and their performance as explanatory variables are compared against the two diversification indexes presented. Table 2 shows the results for the same specifications presented in Table S6 in the Appendix of Hidalgo and Hausmann (2009) using data for the same year (2000).

Table 2: Income regressions (year 2000).

	1	2	3	4	5	6	7	8	9	10	11
	Dependent variable: log per capita GDP										
E_c	0.332***								0.062		0.346***
	(6.01)								(1.00)		(3.30)
HH_c		-1.315***								0.483	2.542***
		(2.87)								(1.18)	(3.44)
$k_{c,0}$			0.001***								
			(7.73)								
$k_{c,1}$				-0.138***							
				(6.14)							
$k_{c,4}$					0.022***						
					(9.04)						
$k_{c,8}$						0.322***					
						(10.14)					
$k_{c,12}$							4.254***				
							(10.65)				
$k_{c,18}$								192.714***	177.166***	202.715***	159.224***
								(10.75)	(7.76)	(10.86)	(6.72)
Obs	178	178	178	178	178	178	178	178	178	178	178
Adj. R-sq	0.16	0.03	0.21	0.21	0.33	0.35	0.35	0.36	0.36	0.36	0.37
F-test	36.12	8.22	59.72	37.67	81.74	102.83	113.34	115.55	57.23	62.06	47.19
Prob> F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes. Cross-section regressions using White's consistent estimator. Only data from the year 2000 was inlcuded. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is the logarithm of per capita GDP. E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,i}$ is the i-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

All coefficient signs are as expected. The even components of the k_c vector have positive effects on income which is as expected since they measure economic complexity. Additionally, E_c has a positive sign which is as expected too since, according to the literature, diversification is positively related to countries' income levels (see for example Imbs and Wacziarg (2003)). The only odd indicator included in Table 2, $k_{c,1}$, presents a negative coefficient which is in line with intuition since this indicator measures the average ubiquity of products exported by each country. Furthermore, HH_c has a negative coefficient which is not surprising either since it is a measure of export concentration. It can also be seen that, among even components of the k_c vector, higher iteration levels yield greater coefficients. This is due to the fact that, as shown in Figure 1, the variability of the indicator decreases with its iteration level.

Every component of the k_c vector is significant at 1% in each regression. Moreover columns 3-8 show how regressions using k_c indicators have greater adjusted-R^2 than regressions using only HH_c or E_c (columns 1 and 2). Given that the number of regressors and observations are the same in each of these specifications this means that the percentage of income explained by MR indicators is greater than that explained by diversification indicators. Notice also that the adjusted-R^2 grows when the specification uses a higher iterated k_c components as regressor, reaching a level of 0.36 for $k_{c,18}$ in column 8. This is much lower than what Hidalgo and Hausmann (2009) find for the same year using a less aggregated dataset (0.535).

Finally, columns 9-11 show how little HH_c and E_c add to the explanation of per capita income above what already is explained by $k_{c,18}$. The adjusted-R^2 remains almost unchanged when including both diversification indicators and their coefficients are not significant.

Despite the difference in coefficient magnitudes, all previous conclusions are similar to those extracted by Hidalgo and Hausmann (2009) and arise when performing this exercise for every year of the sample. Those results are omitted here due to space constraints.

7.1.1 Changing the R^* threshold in income regressions

This section checks how sensitive the former results are to changes in the R^* threshold. Increasing R^* would mean to be more restrictive regarding what the matrix $M_{c,p}$ assigns a 1 to. This would imply to consider as exported by a country a smaller number of highly competitive exports ignoring a greater number of products. Setting a smaller value for R^* implies the opposite.
Table 3 shows results for pooled regressions using all years and including the same variables involved in column 11 of Table 2. MR indicators used here were constructed with alternative values of the R^* threshold, namely 0.8, 0.9, 1.0, 1.1 and 1.2. Results show how MR complexity indicators still present highly significant coefficients in every case and the percentage explained in each specification is very similar between cases. When performing the same cross-section analysis as done in Table 2 for each year and each R^* threshold, the same conclusions always arise. Due to space constraints those results are omitted here. Overall results indicate that MR complexity indicators, and especially those stemming from a high number of iterations, explain countries' per capita income quite robustly.

Table 3: Pooled income regressions for different R^* threshold levels

	1	2	3	4	5
	Dependent variable: log per capita GDP				
$R^*=$	0.8	0,9	1	1,1	1,2
E_c	0.326***	0.323***	0.325***	0.327***	0.330***
	(3.379)	(3.370)	(3.396)	(3.452)	(3.479)
HH_c	2.274***	2.256***	2.222***	2.224***	2.217***
	(3.682)	(3.679)	(3.611)	(3.659)	(3.621)
$k_{c,18}$	245.054***	189.695***	150.403***	124.430***	106.764***
	(7.132)	(7.325)	(7.338)	(7.439)	(7.404)
Obs	2,309	2,309	2,309	2,309	2,309
Adj. R-sq	0.368	0.374	0.374	0.375	0.374
F-test	40.93	42.07	43.25	27.43	36.56
Prob> F	0.000	0.000	0.000	0.000	0.000

Notes. Pooled cross-section regressions using White's consistent estimator. Data from all years was used. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is the logarithm of per capita GDP. E_c is Theil's entropy index of diversification.HH_c is the Hischmann-Herfindahl concentration index. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

7.2 Potentiality of MR indicators to predict long-term growth

In order to test the performance of the MR indicators as predictors of future growth the first step is to perform a similar analysis as done in Tables S7 and S8 in the Appendix of Hidalgo and Hausmann (2009). Both tables present OLS regressions to explain the average growth rate of long time periods (20-years in Table S7 and 10-years in Table S8) using different complexity indicators as main regressors. Each specification uses as main regressors a different pair of even and odd k_c indicators. Although the magnitude of the coefficients of k_c indicators and the adjusted R-squared reported for 20-year periods are greater than those for 10-year periods, the authors conclude that MR indicators significantly predict future growth in both time spans. Therefore, even though the longest time span this work can cover is 13 years, it should still be possible to find significant results.

When computing the same regressions upon the database used here results do not support the conclusion that any of the k_c indicators can be used as predictors of future growth. It should be noticed that original specifications include even and odd indicators which are highly correlated, specially when the number of iterations is high. Table A.2 shows results for specifications that do not include odd indicators in order to better capture the effect that a given indicator has on growth avoiding multicollinearity. Again, robust standard errors are used. It is possible to see there that coefficients are not significant, so Sections 7.2.1, 7.2.2 and 7.2.3 will explore conditions under which significant results could arise.

7.2.1 Using the country filter

Significant results are found when filtering for observations with v <55, that is, dropping from the analysis countries for which the maximum variation between any two years is greater than 55 $k_{c,18}$ ranking positions. The reader can check in Table A.1 which are the countries being left aside, which in this case rise up to 17. Table 4 shows results, when this filter is applied, for the same OLS regressions specified in Table A.2 (also with robust standard errors). Except for $k_{c,0}$, all other k_c indicators help predict

Table 4: Long term growth regressions (filter used $v^* = 55$)

	1	2	3	4	5	6	7	8	9	10
	Dependent variable: average growth rate of per capita GDP (1995-2007)									
rgdpl	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.71)	(0.45)	(0.54)	(1.28)	(1.37)	(1.34)	(1.31)	(1.32)	(1.37)	(1.39)
E_c	0.001							-0.000		0.000
	(1.21)							(0.14)		(0.10)
HH_c		-0.008							0.003	0.004
		(0.80)							(0.27)	(0.23)
$k_{c,0}$			0.000							
			(0.59)							
$k_{c,4}$				0.000**						
				(2.05)						
$k_{c,8}$					0.002**					
					(2.17)					
$k_{c,12}$						0.021**				
						(2.10)				
$k_{c,18}$							0.800**	0.831*	0.835**	0.818*
							(2.04)	(1.91)	(2.14)	(1.80)
Obs	161	161	161	161	161	161	161	161	161	161
Adj. R-sq	-0.00	-0.01	-0.01	0.03	0.03	0.03	0.03	0.02	0.02	0.02
F-test	0.73	0.32	0.19	2.25	2.62	2.48	2.34	1.57	1.66	1.27
Prob> F	0.48	0.72	0.82	0.11	0.08	0.09	0.09	0.20	0.18	0.28

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,i}$ is the i-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

future growth with a confidence level of 5%. It should be noticed that the inclusion of the diversification indexes together with $k_{c,18}$ turns the regressions non significant (see the p-value of the F statistic of the test of joint significance for columns 8-10). Judging by these results, and taking into account the magnitude of the coefficients, the best specification to predict future growth would be one including only $k_{c,18}$ (column 7).

These results are greatly sensible to the filter applied. Table A.3 report results for estimations using $k_{c,18}$ as main regressor, both with and without diversification controls, for different v^* thresholds. Both the significance and magnitude of the effect the indicator has on growth increases as v^* is reduced. This is also the case for the Adjusted-R^2 and the significance of the regression as a whole, and these conclusions hold whether diversification controls are included (columns 1-3) or not (columns 4-6). Presented evidence therefore suggests then that $k_{c,18}$ can function as predictor of future growth when the country filter is selective enough, that is, when the sample is not including countries for which complexity variability is too high.

7.2.2 Including more control variables

Although Hidalgo and Hausmann (2009) do not include any control variables other than diversification indexes in their income and growth regressions, variables accounting for human capital are included in Table 8 of Hausmann et al. (2007) when testing the predictive power of $k_{c,18}$ predecessor, $EXPY$. Table A.4 present results using two different measures of human capital: *ltertiary* (already defined) and *leducexp* which is the log of the percentage of GNI devoted to education as reported by the WDI. Dummy variables to signal low and middle income countries were also included to better capture the effect that income can have in growth regressions. To construct these dummy variables the World Bank Classification of countries[5] is used. Results are reported for three different levels of v^*: 45, 70 and 95.

Table A.4 shows again how the significance of $k_{c,18}$ coefficients grows as v^* is decreased. Columns 1-3 show that, when the filter is set at $v^* = 45$, $k_{c,18}$ significantly predicts future growth even when both human capital controls are included. It is also possible to find significant values for larger country samples (i.e. for less restrictive v^* levels) if *leducexp* is included[6]. Columns 4, 8 and 12 of Table A.4 include the *poor* and *middle* dummy variables to the specification presenting higher and more significant coefficients for $k_{c,18}$ (i.e. the one including only *leducexp* to approximate human capital). Both dummy variables are jointly significant and present negative signs which implies that richer countries had a greater average growth over this period after controlling for export diversification, complexity and human capital. Most important to the purposes of this work, it can be seen that the inclusion of these controls do not hinder the significance of $k_{c,18}$, but rather they enhance it. This suggests that there is some difference in observations (specifically a negative effect upon growth for poor and middle income countries) that is not captured by $k_{c,18}$ alone, but once this effect is controlled for, then the predictive power of $k_{c,18}$ increases.

7.2.3 Removing fast growing and decreasing countries

Besides variability in the complexity index, outliers from the growth rate distribution could bring noise to the true relationship between the two variables. As explained in Section 4.4, MR indicators probably underestimate economic complexity in countries that are experiencing rapid structural changes. This justifies the introduction of an alternative filter: one that ignores countries going through extraordinarily rapid processes, either positive of negative, of economic growth.

Table A.5 presents results for the same specifications in Table A.4 having dropped observations belonging to either the top or bottom 5% of the distribution of the average growth rate (*avggr*). Results show how this modification makes $k_{c,18}$ more significant in some specifications but less significant in others so there is no clear conclusion for the effect that these outliers have on the regressions used here.

7.2.4 Changing the R^* threshold

The link between MR indicators of complexity and income proved to be fairly robust to changes in R^*. Non significant results arise however when performing the same exercise on growth regressions. Table A.6 show results for the same specifications computed in columns 7-10 of Table 4 (which follow the specifications used in Hidalgo and Hausmann (2009)). Results indicate that, growth regression outcomes are very sensitive to the R^* threshold chosen. Significant coefficients for $k_{c,18}$ are only found in some specifications when R^*=1.1.

[5]Available at http://data.worldbank.org/about/country-classifications/country-and-lending-groups.

[6]Notice that both human capital measures are available for a limited number of countries so the sample gets smaller when these variables are included. Country loss is greater using *ltertiary* than using *leducexp*. Given that most countries with high v are also countries for which human capital variables are not available, the inclusion of these variables makes the filter less determinant of the country sample. This explains why only two observation are gained by relaxing the filter from $v^* = 70$ to $v^* = 95$.

It is possible to find a v^* for which significant results are obtained with all alternative R^* values, but this v^* implies a very restrictive filter. In Table A.7 results are presented for each of the alternative R^* thresholds, with $v^* = 30$. The table shows significant results for $k_{c,18}$ as predictor of future growth in every case. Evidence is therefore suggesting that under specifications used by Hidalgo and Hausmann (2009) the significance of $k_{c,18}$ as predictor of long term future growth suffers greatly when changing the R^* threshold and depends strongly upon the country sample used.

A different situation arises when more control variables are included into specifications. As shown in section 7.2.2 the growth predicting power of $k_{c,18}$ can be greatly enhanced when introducing more controls, especially dummies for poor and middle income countries. Table A.8 shows results of growth regressions including both diversification indexes, *leducexp* to proxy human capital and both dummies for poor and middle income countries. Variable *ltertiary* is not included because there are many countries that do not report that information and thus, when the variable is included, filters do not make a difference in the country sample. The inclusion of these controls makes $k_{c,18}$ highly significant for all the selected R^* thresholds. This implies that R^* is not strongly determining the results when these controls are included. Regarding differences brought about by changing the country filter level the usual conclusion holds: magnitude and significance of the $k_{c,18}$ coefficient increases as the allowed v is reduced.

7.3 Potentiality of MR indicators to predict short-term growth

In Table S9 of Hidalgo and Hausmann (2009), the authors present regressions with MR indicators in one year as explanatory variables and the following 5-year average growth rates as independent variables. They use observations from years 1985, 1990, 1995 and 2000 to do this. Since the total time span in this work is 1995-2007 Table A.9 is constructed using years 1995 and 2000 with the purpose of finding similar results, but of course there are only two observations per country instead of four. Regressions in Table A.9 are also performed with robust standard errors and odd complexity indicators are not included, again, to avoid multicollinearity. Results show how every MR indicator used has a very small and non significant coefficient and no single regression can be considered a good fit.

These results are dependent on which five-year period is considered. It is possible to find significant coefficients for $k_{c,18}$ if different five-year periods are used. Still, the magnitude of these coefficients is very small. As has been discussed, the variability of $k_{c,18}$ is very small which means that the coefficients in the regressions should be large if the indicator is to have some impact upon the dependent variable.

It is possible to get more observations for each country by taking five-years moving averages which yields nine observations per country. Table A.10 shows results for the same specifications from Tables A.9 using moving averages. It is noticeable that $k_{c,18}$ has a significant but very small and negative coefficient. The negative sign is at odds with intuition and it is also appearing in Table S9 for the specification including only $k_{c,18}$ and income level as explanatory variables. In order to check whether the sign becomes negative when $k_{c,19}$ is added to the regression, as it is the case in Table S9 of Hidalgo and Hausmann (2009), that indicator is introduced. However signs do not change as shown in columns 11-14 of the table.

Another interesting exercise is to perform regressions using the panel structure of the database which allows to incorporate country fixed-effects as done in Table S10 of Hidalgo and Hausmann (2009). Results show again coefficients for complexity indicators being negative and very small.

7.3.1 Exploring for more meaningful results

Results do not change when control variables are added and different country filter levels are used as shown in columns 1-5 of Table A.11. Columns 6-10 compute the same specifications upon a database in which outliers from the growth rate distribution have been removed and, as can be seen, this does not make the coefficient of $k_{c,18}$ greater or positive either.

The fact that coefficient the of $k_{c,18}$ is so small leads to conclude that MR indicators are not contributing much in predicting future growth in the short term. This result does not seem surprising. After all, the literature explains that economic complexity is something that materializes over time and therefore the link between complexity and growth is typically a long term phenomenon.

8 Conclusions

Many works have discussed the relationship between a country's current production and its long term growth possibilities. Recently, Hidalgo and Hausmann (2009) presented a very strong tool with their Method of Reflections (MR). The work finds that MR indicators of economic complexity significantly explain current per capita GDP and also contribute in the prediction of its growth. Previous works from the same group of researchers suggested measures of distance between current production and potentially produced goods. All these tools combined can provide policy-makers with detailed insights to better design industrial policies that help economies diversify "the right way" and accumulate valuable capabilities, which would enhance their growth prospectives.

This work has tested the explanatory and predictive power of MR indicators of economic complexity by comparing results in Hidalgo and Hausmann (2009) with new results obtained under different conditions (i.e. changing the used dataset, altering parameters in the construction of the indicators and inlcuding more variables in the regressions). Results show that the main characteristics of MR indicators also arise when computed upon the BACI. MR indicators get more complex and closely related to per capita GDP as i increases: low iterated indicators are rough approximations to economic complexity and are vulnerable to different distortionary effects, but highly iterated indicators are less dependent on these effects and present a much richer informational basis.

Results also support the conclusion that MR indicators of economic complexity perform well as explanatory variables of a country's current per capita GDP. Every component of the k_c vector tested here has a significant coefficient in cross section regressions performed for every year in the sample. This is also found in pooled OLS regressions using data for the entire period and in regressions in which the R^* threshold is modified around its original value of 1.

MR complexity indicators are vulnerable to sudden changes in export figures and so there are plenty of sources of noise that affect their predicting power of long-term growth. The regression analysis done here shows that the noise can be too much when the country sample does not filter out countries for which $k_{c,18}$ is very volatile and this can prevent the indicators to significantly predict future growth. Adding control variables previously used in the literature, like human capital indicators and dummy variables for poor and middle income countries may increase the significance of $k_{c,18}$ in these regressions. The inclusion of these variables can also make $k_{c,18}$ overcome the test of changing the R^* threshold. The fact that the time span used here is only 13 years long could partly explain why MR indicators do not appear to be as strong predictors of long term growth as in Hidalgo and Hausmann (2009). That work shows that there are great differences in the predicting power MR indicators exhibit over the 20-year average growth rate and over the 10-year average growth rate. Still, the strong dependence on the country sample and the inclusion of covariates showed here is not likely to vanish with longer time spans.

Finally, results do not support the conclusion that MR economic complexity indicators are good predictors of growth in the short term. Outcomes differ greatly when changing the initial year of the five-year periods used in estimations. In every regression performed, coefficients for MR indicators are too small and some of them present signs that are at odds with intuition. Therefore complexity indicators are not contributing in the prediction of short-term growth. This conclusion is supported by the literature reviewed here which states that the relationship between economic complexity and growth materializes over the long term.

References

Abdon, A. and Felipe, J. (2011), "The Product Space: What Does It Say About the Opportunities for Growth and Structural Transformation of Sub-Saharan Africa?" *Levy Economic Institute of Bard College Working Paper*, 670, 1-35.

Aghion, P. and Howitt, P. (1992), "A model of growth through creative destruction." *Econometrica*, 60(2), 323-51.

Araujo, R. A. and Lima, G. T. (2007), "A structural economic dynamics approach to balance-of-payments-constrained growth." *Cambridge Journal of Economics*, 31(5), 755-74.

Balassa, B. (1986), "Comparative advantage in manufactured goods: a reappraisal." *The Review of Economics and Statistics*, 68(2), 315-19.

Baumol, W. J. (1967), "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review*, 57(3), 415-26.

Cadot, O; Carrère, C. and Strauss-Kahn, V. (2011), "Export Diversification: What's behind the Hump?." *The Review of Economics and Statistics*, 93(2), 590–605.

Dosi, G. (1982), "Technological paradigms and technological trajectories." *Journal of Economic Literature*, *Research Policy*(11), 147-62.

Fagerberg, J. (1994), "Technology and International Differences in Growth Rates." *Journal of Economic Literature*, 32, 1147-75.

Feenstra, R. C; Lipsey, R. E; Deng, H; Ma, A. C. and Mo, H. (2005), "World Trade Flows: 1962-2000." *NBER Working Paper*, 11040.

Gaulier, G. and Zignago, S. (2010), "BACI: International Trade Database at the Product-level The 1994-2007 Version", *CEPII Working Paper* 2010-23.
Grossman, G. and Helpman, E. (1991), *"Innovation and Growth in the Global Economy."* The MIT Press, Cambridge MA.
Hausmann, R. and Hidalgo, C. (2011), "The Network Structure of Economic Output." *Journal of Economic Growth*, 16(4), 309-42.
Hausmann, R. and Klinger, B. (2007), "The Structure of the Product Space and the Evolution of Comparative Advantage." *CID Working Paper*, Harvard University, 146, 1-40.
Hausmann, R. and Klinger, B. (2006), "South Africa's Export Predicament" *CID Working Paper*, Harvard University, 129, 1-51.
Hausmann, R; Hwang, J. and Rodrik, D. (2007), "What you export matters." *Journal of Economic Growth*, 12(1), 1-25.
Hausmann, R. and Rodrik, D. (2003), "Economic development as self-discovery." *Journal of Development Economics*, 72(2), 603-33.
Heckscher, E.F. and Ohlin, B. (1991), *"Heckscher-Ohlin Trade Theory."* Ed. Flam H. and Flanders, M. J., The MIT Press, Cambridge MA.
Heston, A; Summers, R. and Aten, B. (2011), Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
Hidalgo, C. (2009), "The Dynamics of Economic Complexity and the Product Space over a 42 year period." *CID Working Paper*, Harvard University, 189, 1-20.
Hidalgo, C. and Hausmann, R. (2010), "Inferring Macroeconomic Complexity from Country-Product Network Data." *Association for the Advancement of Artificial Intelligence*, 1-4.
Hidalgo, C. and Hausmann, R. (2009), "The Building Blocks of Economic Complexity." *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10570-75.
Hidalgo, C; Klinger, B; Barabasi, A. L. and Hausmann, R. (2007), "The Product Space Conditions the Development of Nations." *Science*, 317(5837), 482-87.
Imbs, J. and Wacziarg, R. (2003), "Stages of Diversification." *American Economic Review*, 93(1), 63-86.
Jankowska, A; Nagengast, A. and Perea, J. R. (2012), "The Product Space and the Middle-Income Trap: Comparing Asian and Latin American Experiences" *OECD Development Center Working Paper*, 311, 1-70.
Klinger, B. and Lederman, D. (2006), "Diversification, Innovation, and Imitation inside the Global Technology Frontier." *World Bank Policy Research working paper*, 3872, 1-48.
Klinger, B. and Lederman, D. (2004), "Discovery and development : an empirical exploration of "new" products." *World Bank Policy Research working paper*, 3450, 1-24.
Lall, S. (2000), "The Technological Structure and Performance of Developing Country Manufactured Exports, 1985-98." *Oxford Development Studies*, 28(3), 337-69.
Lewis, W. A. (1954), "Economic Development with Unlimited Supplies of Labor." *Manchester School of Economic and Social Studies*, 22, 139-91.
Pasinetti, L. L. (1981), *"Structural Change and Economic Growth: a Theoretical Essay on the Dynamics of the Wealth of Nations."* Cambridge University Press, Cambridge.
Prebisch, R. (1949), *"The Economic Development of Latin America and its Principal Problems."* United Nations, New York.

Ramsey, F. P. (1928), “A mathematical theory of saving.” *Economic Journal*, 38(152), 543–59.
Ranjan Basu, S. and Das, M. (2011), “Export Structure and Economic Performance in Developing Countries: Evidence From Nonparametric Methodology.” *Policy Issues in International Trade and Commodities Study Series*, United Nations, 48, 1-59.
Record, R. and Nghardsaysone, K. (2010), “Export Dynamics and Diversification in Lao PDR: An analysis of the product space”, *Lao PDR Development Report, Background Paper*, 1-43.
Romer, P. (1986), “Increasing Returns and Long-Run Growth.” *Journal of Political Economy*, 94(5), 1002-37.
Sachs, J. D. and Warner, A. M. (1995), “Natural Resource Abundance and Economic Growth.” *NBER Working Papers* 5398, 1-47.
Solow, R. (1956), “A Contribution to the Theory of Economic Growth.” *Quarterly Journal of Economics*, 70(1), 65-94.
Theil, H. (1972), *“Statistical Decomposition Analysis.”* North-Holland Pub. Co., Amsterdam.
Thirlwall, A. P. (1979), “The Balance of Payments Constraint as an Explanation of International Growth Rate Differences.” *BNL Quarterly Review*, 32(128), 45-53.
Thirlwall, A. P. and Hussain, M. N. (1982), “The Balance of Payments Constraint, Capital Flows and Growth Rate Differences between Developing Countries,” *Oxford Economic Papers*, 34(3), 498-510.

Appendix

Table A.1: All countries sorted by v.

n	country	v	pop	$rgdpl$	HH_c	n	country	v	pop	$rgdpl$	HH_c
1	Timor-Leste	158	1086,17	1086,85	0,301565	55	Burkina Faso	27	14797,2	904,828	0,504744
2	Chad	101	9885,66	1216,7	0,893369	56	Côte d'Ivoire	27	19746,9	1299,97	0,088069
3	Somalia	85	9238	455,29	0,075042	57	Dominican Rep.	26	9426	9341,09	0,038829
4	Micronesia	85	107,848	3459,25	0,240642	58	Georgia	26	4646	4708,76	0,043104
5	Eritrea	82	5357,68	668,078	0,484823	59	Cameroon	26	18060,4	1826,42	0,274496
6	Kiribati	77	107,915	3623,06	0,111618	60	Togo	26	6042	743,425	0,106294
7	Palau	77	20,591	15043	0,413497	61	Macao SAR	26	525,531	50192,2	0,033739
8	Iraq	74	27500	3948,59	0,979519	62	Fiji	25	852	4470,62	0,044983
9	Marshall Isds	73	61,815	7350,88	0,300723	63	Viet Nam	25	86519	2584,62	0,031208
10	Brunei D.	70	374,577	49063	0,520847	64	Gambia	23	1630	1325,67	0,100369
11	Guinea-Bissau	68	1472,78	790,91	0,786911	65	Benin	23	8278,16	1254,22	0,132982
12	Rwanda	65	10141	924,106	0,12687	66	Grenada	23	106	12910,2	0,113957
13	Bahamas	59	302	29114,8	0,179394	67	Lebanon	22	3896	11102,8	0,021924
14	Ctrl African R.	58	4544	614,929	0,187505	68	Kyrgyzstan	22	5284,15	2034,36	0,077921
15	Djibouti	57	694	2163,58	0,163228	69	Nigeria	22	143312	1941,63	0,739259
16	Liberia	57	3270	403,546	0,27621	70	Bahrain	21	1054	24388,7	0,46103
17	Eq. Guinea	57	599,763	18065,5	0,746179	71	Zambia	21	12341	1794,78	0,317118
18	Maldives	54	364,968	4689,11	0,161234	72	Sudan	21	40526	2065,73	0,787094
19	Burundi	52	8783	370,88	0,158174	73	Albania	21	2992	6017,55	0,024363
20	Qatar	48	814,897	123307	0,272779	74	Kuwait	21	2376	51608,9	0,450997
21	Mali	46	12769	937,231	0,543219	75	Bosnia Herz.	20	4552,2	6245,46	0,013148
22	Comoros	46	711,417	927,715	0,125649	76	Lao	20	6035	2338,74	0,211075
23	Cape Verde	45	486	3498,44	0,188435	77	Ghana	20	22981	1143,45	0,164632
24	Jamaica	45	2782,22	9232,01	0,256183	78	Jordan	19	5997	4412,85	0,014041
25	Sierra Leone	44	4918	830,022	0,138264	79	Uganda	19	30263	1063,86	0,043404
26	Saint Lucia	43	158,875	12684	0,116261	80	Trin. & Tobago	19	1233	26286,3	0,146997
27	Seychelles	43	85,702	26659,9	0,24076	81	Papua New Guin.	19	5806,04	2417,56	0,148204
28	Cuba	43	11160	10955,6	0,202379	82	Guyana	19	764	4028,27	0,09473
29	Ant. and Barb.	43	83,425	17449,4	0,276869	83	Madagascar	18	19448,8	776,605	0,033567
30	Bhutan	42	673,353	4077,33	0,150523	84	Sao Tome & P.	18	165	1522,38	0,179595
31	Gabon	42	1456,45	10300,2	0,523353	85	UAE	18	4444	53847,8	0,280468
32	Mozambique	40	20905,6	715,726	0,247699	86	Syria	18	20488	3955,62	0,15499
33	Solomon Isds	39	566,948	1598,01	0,513748	87	Cyprus	18	1049	19463,2	0,032326
34	Libya	39	6036,91	19822	0,72609	88	Paraguay	18	6113	3785,18	0,130105
35	Angola	39	12263,6	4353,26	0,907839	89	Vanuatu	18	212	6042,99	0,25204
36	Tonga	37	104	7585,63	0,122921	90	Macedonia	18	2055,92	7330,34	0,031947
37	Niger	37	14214,7	532,473	0,191329	91	China	18	1310584	5889,78	0,006719
38	Senegal	37	11394	1494,68	0,038779	92	Mauritania	17	2981,45	1619,2	0,169208
39	Algeria	36	33362,7	6062,33	0,331637	93	Zimbabwe	17	11443,2	158,696	0,065195
40	Samoa	35	188	7000,89	0,355918	94	Nepal	16	27827,9	1092,18	0,029289
41	Suriname	35	470,784	10169,3	0,286858	95	Oman	16	2800	20761,8	0,40396
42	Afghanistan	34	31889,9	736,48	0,031112	96	Bolivia	16	9425,94	3619,5	0,1912
43	Armenia	34	2971,65	5603,83	0,063014	97	Nicaragua	16	5408	2113,12	0,035454
44	St. V.and Gren.	34	105,307	6774,5	0,219774	98	Peru	16	28050	6750,19	0,070943
45	Bermuda	34	66,921	50487,9	0,263617	99	Iran	16	74093	10059,1	0,647413
46	Congo	33	3802,33	1912,89	0,652168	100	Moldova	16	4328,82	2376,18	0,017708
47	Belize	32	294,61	9251,59	0,08747	101	Malta	15	401,88	21891,9	0,095977
48	Yemen	32	21591	2421,06	0,555162	102	Dominica	15	72,377	5585,79	0,07416
49	Azerbaijan	32	8120,25	7969,98	0,609793	103	Belarus	14	9724,72	11177,4	0,106575
50	Tajikistan	31	7076,6	1764,08	0,429195	104	Congo	14	64355	227,135	0,135064
51	Tanzania	31	39384,2	1039,1	0,047559	105	Malaysia	14	26896	11643,6	0,021171
52	Turkmenistan	29	4774	6195,14	0,563299	106	Kazakhstan	14	15284,9	10845	0,257445
53	Mongolia	29	2951,79	3040,43	0,201678	107	Mauritius	14	1263,9	9001,03	0,059851
54	Guinea	27	9569,22	834,13	0,236281	108	Malawi	14	14233	581,423	0,17664

Table A.1(cont.): All countries sorted by maximum change in $k_{c,18}$ ranking (v)

n	country	v	pop	$rgdpl$	HH_c	n	country	v	pop	$rgdpl$	HH_c
109	Uzbekistan	13	27079	2081,98	0,083025	144	El Salvador	8	5982	6501,72	0,017453
110	Thailand	13	65110	7752,5	0,011399	145	Portugal	8	10642,8	20700,2	0,006611
111	Egypt	13	75677	4684,57	0,053065	146	Argentina	7	40048,8	11344,8	0,030331
112	Costa Rica	13	4331	11215,6	0,069125	147	Ukraine	7	46300	7123,5	0,009935
113	Panama	13	3258,33	9290,75	0,024457	148	Hungary	7	9956,11	17486,9	0,017818
114	Ecuador	13	14135	5972,97	0,259532	149	Bulgaria	7	7322,86	10529,3	0,018364
115	Romania	13	22106	9523,52	0,009853	150	Turkey	7	74768	10549,5	0,007656
116	Uruguay	13	3279	9959,75	0,025546	151	Rep. of Korea	7	48250,1	25061,7	0,019491
117	Tunisia	13	10213	5938,28	0,02254	152	Colombia	7	42597	7522,18	0,060776
118	Indonesia	12	234694	3626,22	0,016828	153	Australia	7	20749,6	40290,1	0,033097
119	Croatia	12	4493,31	15620,3	0,015098	154	Slovakia	6	5447,5	19495,1	0,02168
120	Honduras	12	7516	3625,6	0,041514	155	Poland	6	38518,2	15249,1	0,004783
121	Morocco	12	30594	3327,02	0,013326	156	Norway	5	4627,93	50959,3	0,169396
122	Greece	12	10706,3	27963,8	0,020284	157	New Zealand	5	4132,34	28297	0,015447
123	Hong Kong SAR	12	6980,41	37366,7	0,010937	158	Slovenia	4	2009	26593,1	0,00884
124	Mexico	12	108701	12696,9	0,03316	159	Ireland	4	4420	39168	0,04142
125	Ethiopia	11	79935,8	593,063	0,103722	160	Czech Rep.	4	10228,7	23518,8	0,005864
126	Philippines	11	94157,5	2963,96	0,046826	161	Denmark	4	5468,12	36335,8	0,008107
127	Sri Lanka	11	20508	3738,78	0,017016	162	Netherlands	4	16570,6	40691,2	0,011061
128	Singapore	11	4553,01	48215,1	0,050863	163	India	4	1124135	2999,89	0,023729
129	Venezuela	11	26415	9545,42	0,46222	164	Belgium	3	10392,2	35574	0,009099
130	Estonia	11	1315,91	19047,4	0,016827	165	Italy	3	59627	30199,6	0,003516
131	Guatemala	11	12728,1	6043,28	0,020175	166	Spain	3	45212	29133,5	0,008828
132	Bangladesh	10	148894	1290,07	0,048414	167	Brazil	3	193919	9040,19	0,013486
133	Saudi Arabia	10	24499	20449,2	0,621102	168	Israel	3	6990	25473,6	0,062931
134	Haiti	10	9500	1383,01	0,176634	169	Switzerland	2	7555	39912,1	0,019024
135	Russia	10	141378	14495,7	0,150595	170	France	2	63681,7	32015,4	0,006256
136	Kenya	10	36913,7	1233,46	0,035966	171	Canada	2	32936	37703	0,023094
137	Iceland	10	301,931	43125,3	0,092013	172	Finland	2	5238,46	34888,8	0,017606
138	Barbados	10	282,359	24556,1	0,045758	173	Sweden	1	9031	37358,7	0,007255
139	Cambodia	9	13719	1800,39	0,055581	174	Japan	1	127433	34223,8	0,011747
140	Lithuania	9	3575,44	15648,4	0,022448	175	Austria	1	8199,78	38232,8	0,003751
141	Latvia	9	2259,81	15486,3	0,039695	176	United Kingdom	1	61249	35649	0,011113
142	Pakistan	9	175495	2291,78	0,012672	177	USA	1	301580	43697,5	0,005304
143	Chile	9	16303,9	12135,4	0,134625	178	Germany	1	82237	33638,2	0,005786

Notes. v represents the maximum variation in the $k_{c,18}$ ranking that each country exhibited in the period 1995-2007. *pop* is the total population of the country in thousands as reported by the WDI. HH_c is the Hischmann-Herfindahl concentration index. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011)

Table A.2: Long term growth regressions (no filter used).

	1	2	3	4	5	6	7	8	9	10
	Dep. var.: avg. growth rate of per capita GDP (1995-2007)									
rgdpl	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(1.17)	(0.96)	(1.13)	(1.14)	(1.18)	(1.20)	(1.22)	(1.21)	(1.25)	(1.35)
E_c	0.001							0.000		0.002
	(0.73)							(0.29)		(0.85)
HH_c		-0.002							0.003	0.016
		(0.19)							(0.27)	(0.88)
$k_{c,0}$			0.000							
			(0.65)							
$k_{c,4}$				0.000						
				(0.48)						
$k_{c,8}$					0.001					
					(0.56)					
$k_{c,12}$						0.009				
						(0.60)				
$k_{c,18}$							0.338	0.280	0.369	0.202
							(0.64)	(0.47)	(0.66)	(0.34)
Obs	177	177	177	177	177	177	177	177	177	177
Adj. R-sq	-0.00	-0.01	-0.01	-0.00	-0.00	-0.00	-0.00	-0.01	-0.01	-0.01
F-test	0.76	0.46	0.65	0.71	0.75	0.77	0.78	0.53	0.61	0.80
Prob> F	0.47	0.63	0.52	0.49	0.47	0.47	0.46	0.66	0.61	0.53

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,i}$ is the i-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.3: Long term growth regressions. Sensitivity to different filter levels.

	1	2	3	4	5	6
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)					
$v^{*}=$	45	55	65	45	55	65
rgdpl	-0.000***	-0.000	-0.000	-0.000***	-0.000	-0.000
	(3.28)	(1.39)	(0.90)	(3.23)	(1.31)	(0.83)
E_c	0.000	0.000	0.001			
	(0.12)	(0.10)	(0.23)			
HH_c	0.004	0.004	0.008			
	(0.21)	(0.23)	(0.42)			
$k_{c,18}$	1.128***	0.818*	0.225	1.122***	0.800**	0.210
	(2.63)	(1.80)	(0.33)	(3.35)	(2.04)	(0.34)
Obs	154	161	166	154	161	166
Adj. R-sq	0.06	0.02	-0.02	0.07	0.03	-0.01
F-test	3.23	1.27	0.36	6.10	2.34	0.44
Prob> F	0.01	0.28	0.84	0.00	0.10	0.64

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.4: Long term growth regressions. Including controls (different filters used).

	1	2	3	4	5	6	7	8	9	10	11	12
	Dependent variable: average growth rate of per capita GDP (1995-2007)											
$v^*=$	45				70				95			
rgdpl	-0.000***	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000***	-0.000***	-0.000	-0.000***	-0.000***	-0.000***
	(4.070)	(3.688)	(3.990)	(4.657)	(1.597)	(3.346)	(4.047)	(4.727)	(1.622)	(3.332)	(3.856)	(4.693)
E_c	-0.000	0.000	0.000	-0.002	0.000	0.001	-0.000	-0.001	0.002	0.002	0.001	-0.001
	(0.099)	(0.089)	(0.052)	(0.671)	(0.148)	(0.266)	(0.043)	(0.525)	(0.476)	(0.546)	(0.435)	(0.326)
HH_c	-0.005	0.008	0.005	0.001	0.013	0.004	0.002	0.004	0.023	0.010	0.017	0.007
	(0.172)	(0.356)	(0.177)	(0.059)	(0.475)	(0.213)	(0.093)	(0.211)	(0.868)	(0.463)	(0.674)	(0.421)
ltertiary	0.003		0.004		0.004		0.004		0.006*		0.006*	
	(1.072)		(1.207)		(1.382)		(1.202)		(1.924)		(1.718)	
leducexp		-0.002	0.004	-0.005		0.000	0.004	-0.004		0.002	0.006	-0.003
		(0.316)	(0.457)	(0.924)		(0.009)	(0.432)	(0.755)		(0.335)	(0.773)	(0.546)
poor				-0.034***				-0.037***				-0.037***
				(3.528)				(3.991)				(3.948)
middle				-0.013*				-0.013**				-0.012*
				(1.950)				(2.054)				(1.873)
$k_{c,18}$	1.221**	1.517***	1.232*	1.568***	0.473	1.257**	1.251*	1.401***	0.121	0.944*	0.656	1.121**
	(2.027)	(2.851)	(1.870)	(3.190)	(0.612)	(2.359)	(1.923)	(2.936)	(0.172)	(1.731)	(0.951)	(2.351)
Obs	96	139	92	139	99	148	94	148	100	150	95	150
Adj. R-sq	0.124	0.076	0.133	0.182	0.030	0.059	0.133	0.198	0.042	0.052	0.131	0.198
F-test	3.820	3.649	3.065	4.789	0.987	3.081	3.224	5.297	1.259	3.266	3.262	5.737
Prob> F	0.003	0.004	0.009	0.000	0.430	0.011	0.007	0.000	0.288	0.008	0.006	0.000

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. *ltertiary* is the logarithm of the gross rate of enrolment in tertiary education and *educexp* is the logarithm of the percentage of GNI devoted to educationas reported by the WDI Database. *poor* and *middle* are dummy variables indicating countries that are ranked as low or middle income countries according to the World Bank Classification of countries. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Long term growth regressions. Cleaning for outliers in the distribution of the dep. variable.

	1	2	3	4	5	6	7	8	9	10	11	12
	Dependent variable: average growth rate of per capita GDP (1995-2007)											
$v^* =$	45				55				65			
rgdpl	-0.000***	-0.000***	-0.000***	-0.000**	-0.000	-0.000***	-0.000***	-0.000**	-0.000	-0.000***	-0.000***	-0.000***
	(4.408)	(3.350)	(4.275)	(2.145)	(1.120)	(3.202)	(4.365)	(2.227)	(1.120)	(4.043)	(4.365)	(2.838)
E_c	-0.002	0.001	-0.001	0.001	-0.001	0.001	-0.002	0.001	-0.001	0.001	-0.002	0.001
	(0.868)	(0.585)	(0.584)	(0.599)	(0.444)	(0.563)	(0.726)	(0.514)	(0.444)	(0.547)	(0.726)	(0.509)
HH_c	-0.023	0.014	-0.014	0.014	-0.004	0.009	-0.016	0.009	-0.004	0.009	-0.016	0.009
	(1.087)	(0.742)	(0.658)	(0.747)	(0.180)	(0.504)	(0.822)	(0.511)	(0.180)	(0.494)	(0.822)	(0.505)
ltertiary	0.000		0.001		0.002		0.001		0.002		0.001	
	(0.171)		(0.420)		(0.659)		(0.410)		(0.659)		(0.410)	
leducexp		0.002	0.005	0.002		0.003	0.005	0.003		0.003	0.005	0.003
		(0.411)	(1.140)	(0.426)		(0.848)	(1.113)	(0.793)		(0.843)	(1.113)	(0.792)
poor				0.002				-0.000				0.000
				(0.283)				(0.012)				(0.043)
middle				0.002				0.001				0.001
				(0.314)				(0.176)				(0.249)
$k_{c,18}$	1.256***	0.790*	1.138**	0.784*	0.490	0.656	1.149**	0.664	0.490	0.637	1.149**	0.648
	(2.735)	(1.728)	(2.218)	(1.688)	(0.687)	(1.440)	(2.258)	(1.426)	(0.687)	(1.481)	(2.258)	(1.479)
Obs	89	128	85	128	92	136	87	136	92	137	87	137
Adj. R-sq	0.153	0.051	0.164	0.035	-0.003	0.044	0.166	0.029	-0.003	0.062	0.166	0.048
F-test	4.961	2.498	3.641	1.787	0.389	2.380	3.784	1.751	0.389	3.978	3.784	3.178
Prob> F	0.001	0.034	0.003	0.096	0.855	0.042	0.002	0.103	0.855	0.002	0.002	0.004

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. *ltertiary* is the logarithm of the gross rate of enrolment in tertiary education and *educexp* is the logarithm of the percentage of GNI devoted to educationas reported by the WDI Database. *poor* and *middle* are dummy variables indicating countries that are ranked as low or middle income countries according to the World Bank Classification of countries. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.6: Long term growth regressions ($v^* = 55$ and different R^* values).

	1	2	3	4	5	6	7	8
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)							
R^*=	0.8				0.9			
rgdpl	-0.000**	-0.000**	-0.000**	-0.000**	-0.000	-0.000	-0.000	-0.000
	(2.21)	(2.19)	(2.09)	(2.07)	(1.06)	(1.07)	(1.08)	(1.06)
E_c		0.001		-0.001		-0.000		-0.000
		(0.41)		(0.37)		(0.24)		(0.16)
HH_c			-0.008	-0.014			0.002	-0.001
			(0.93)	(0.81)			(0.18)	(0.04)
$k_{c,18}$	1.001	0.870	0.837	0.946	0.446	0.511	0.477	0.515
	(1.02)	(0.80)	(0.80)	(0.88)	(0.57)	(0.60)	(0.58)	(0.61)
Obs	162	162	162	162	162	162	162	162
Adj. R-sq	0.01	0.01	0.01	0.00	-0.00	-0.01	-0.01	-0.02
F-test	3.23	2.47	2.65	1.98	0.61	0.44	0.45	0.34
Prob> F	0.04	0.06	0.05	0.10	0.54	0.73	0.71	0.85

	9	10	11	12	13	14	15	16
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)							
R^*=	1.1				1.2			
rgdpl	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(1.26)	(1.27)	(1.31)	(1.34)	(0.85)	(0.84)	(0.83)	(0.83)
E_c		-0.000		0.000		0.000		0.000
		(0.04)		(0.17)		(0.10)		(0.00)
HH_c			0.002	0.005			-0.001	-0.001
			(0.19)	(0.26)			(0.13)	(0.07)
$k_{c,18}$	0.632*	0.639*	0.652**	0.627	0.150	0.134	0.137	0.137
	(1.93)	(1.75)	(2.01)	(1.64)	(0.33)	(0.27)	(0.29)	(0.28)
Obs	162	162	162	162	161	161	161	161
Adj. R-sq	0.02	0.02	0.02	0.01	-0.01	-0.01	-0.01	-0.02
F-test	2.08	1.38	1.46	1.15	0.47	0.31	0.31	0.24
Prob> F	0.13	0.25	0.23	0.34	0.62	0.82	0.82	0.91

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.7: Long term growth regressions ($v^* = 30$ and different R^* values).

	1	2	3	4	5
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)				
$R^*=$	0.8	0.9	1.0	1.1	1.2
rgdpl	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(2.89)	(2.87)	(2.65)	(2.97)	(3.01)
$k_{c,18}$	1.538***	1.352***	0.933***	0.780***	0.664***
	(2.73)	(2.99)	(2.64)	(2.63)	(2.63)
Obs	127	129	127	130	132
Adj. R-sq	0.06	0.06	0.05	0.06	0.05
F-test	4.51	4.85	3.96	4.53	4.65
Prob$>F$	0.01	0.01	0.02	0.01	0.01

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.8: Long term growth regressions. Adding controls. Sensitivity to R^* and v^*

	1	2	3	4	5	6	7	8
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)							
$R^*=$	0.8				0.9			
$v^*=$	35	55	75	no filter	35	55	75	no filter
rgdpl	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(4.721)	(4.812)	(4.860)	(4.810)	(4.783)	(4.803)	(4.800)	(4.760)
E_c	-0.002	-0.002	-0.002	-0.001	-0.002	-0.003	-0.002	-0.001
	(0.715)	(1.010)	(0.695)	(0.450)	(0.754)	(1.034)	(0.599)	(0.356)
HH_c	-0.002	-0.008	0.004	0.010	-0.002	-0.009	0.004	0.010
	(0.082)	(0.454)	(0.219)	(0.553)	(0.108)	(0.502)	(0.221)	(0.556)
leducexp	-0.007	-0.005	-0.004	-0.003	-0.007	-0.007	-0.004	-0.003
	(1.143)	(0.895)	(0.772)	(0.635)	(1.247)	(1.219)	(0.766)	(0.627)
poor	-0.038***	-0.037***	-0.036***	-0.037***	-0.037***	-0.035***	-0.036***	-0.037***
	(3.964)	(3.959)	(3.874)	(3.956)	(3.809)	(3.792)	(3.836)	(3.921)
middle	-0.015**	-0.014**	-0.012*	-0.013*	-0.015**	-0.014**	-0.012*	-0.012*
	(2.180)	(2.118)	(1.887)	(1.899)	(2.116)	(2.059)	(1.854)	(1.871)
$k_{c,18}$	2.493***	2.427***	2.174***	1.953**	2.005***	2.057***	1.624***	1.449**
	(2.936)	(3.213)	(2.898)	(2.564)	(3.173)	(3.601)	(2.807)	(2.474)
Constant	-2,741.376***	-2,669.424***	-2,390.454***	-2,147.602**	-2,025.755***	-2,078.694***	-1,641.321***	-1,464.290**
	(2.936)	(3.213)	(2.898)	(2.564)	(3.172)	(3.601)	(2.807)	(2.474)
Obs	127	146	149	151	127	144	149	151
Adj. R-sq	0.214	0.208	0.207	0.202	0.205	0.210	0.202	0.198
F-test	5.354	5.514	5.704	5.742	5.367	5.633	5.672	5.728
Prob> F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	9	10	11	12	13	14	15	16
	Dep. var.: avg. gr. rate of p/c GDP (1995-2007)							
$R^*=$	1.1				1.2			
$v^*=$	35	55	75	no filter	35	55	75	no filter
rgdpl	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(4.796)	(4.638)	(4.618)	(4.712)	(4.867)	(4.586)	(4.566)	(4.670)
E_c	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
	(0.321)	(0.561)	(0.427)	(0.227)	(0.448)	(0.487)	(0.367)	(0.169)
HH_c	0.017	0.001	0.005	0.010	0.010	0.001	0.004	0.010
	(0.845)	(0.080)	(0.250)	(0.560)	(0.471)	(0.080)	(0.248)	(0.559)
leducexp	-0.004	-0.004	-0.004	-0.003	-0.004	-0.004	-0.003	-0.003
	(0.741)	(0.799)	(0.677)	(0.565)	(0.695)	(0.753)	(0.629)	(0.518)
poor	-0.038***	-0.037***	-0.037***	-0.036***	-0.039***	-0.037***	-0.037***	-0.037***
	(3.956)	(3.953)	(3.988)	(3.936)	(4.095)	(3.965)	(3.993)	(3.941)
middle	-0.014*	-0.014**	-0.013**	-0.012*	-0.015**	-0.014**	-0.013**	-0.013*
	(1.965)	(2.090)	(2.048)	(1.898)	(2.157)	(2.100)	(2.055)	(1.907)
$k_{c,18}$	1.126***	1.137***	1.104***	0.905**	1.009***	0.941***	0.915***	0.742**
	(2.621)	(2.790)	(2.751)	(2.277)	(2.713)	(2.654)	(2.629)	(2.164)
	(2.621)	(2.790)	(2.751)	(2.277)	(2.712)	(2.654)	(2.629)	(2.164)
Obs	125	146	148	151	128	145	148	151
Adj. R-sq	0.213	0.187	0.191	0.191	0.210	0.182	0.187	0.187
F-test	5.598	5.049	5.137	5.724	5.489	4.998	5.091	5.692
Prob> F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 13-year average growth rate and values from the initial year (1995) are used for the explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. *leducexp* is the log of the percentage of GNI devoted to educationas reported by the WB in the WDI. *poor* and *middle* are dummy variables indicating countries that are ranked as low or middle income countries according to the World Bank Classification of countries. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.9: Short term growth regressions.

	1	2	3	4	5	6	7	8	9	10
	Dep. var.: avg. gr. rate of p/c GDP (5-year periods starting in 1995 and 2000)									
rgdpl	-0.0000	-0.0000	-0.0000*	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000*
	(1.475)	(1.134)	(1.743)	(1.473)	(1.432)	(1.289)	(1.114)	(1.522)	(1.181)	(1.783)
E_c	0.0011		0.0032					0.0012		0.0033
	(0.853)		(1.310)					(0.855)		(1.307)
HH_c		-0.0024	0.0206						-0.0024	0.0207
		(0.186)	(0.858)						(0.187)	(0.855)
$k_{c,0}$				0.0000						
				(1.002)						
$k_{c,4}$					0.0001					
					(0.598)					
$k_{c,8}$						0.0001				
						(0.625)				
$k_{c,18}$							0.0001	0.0001	0.0001	0.0001
							(0.515)	(0.522)	(0.515)	(0.527)
Obs	355	355	355	355	355	355	355	355	355	355
Adj. R-sq	-0.001	-0.003	-0.001	-0.002	-0.000	-0.002	-0.003	-0.003	-0.005	-0.003
F-test	1.298	0.665	1.334	1.297	1.127	1.093	0.899	0.975	0.653	1.021
Prob> F	0.274	0.515	0.263	0.275	0.325	0.336	0.408	0.404	0.581	0.396

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is 5-year average growth rate for periods starting in 1995 and 2000. Values for years 1995 and 2000 are used for explanatory variables. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,i}$ is the i-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.10: Short term growth regressions with moving averages.

	1	2	3	4	5	6	7
	Dep. var.: 5-yr. periods moving avg. gr. rate of p/c GDP (1995-2003)						
$rgdpl$	-0.0000	-0.0000	-0.0000**	-0.0000*	-0.0000	-0.0000	-0.0000
	(1.456)	(0.704)	(2.119)	(1.695)	(1.500)	(0.368)	(0.624)
E_c	0.0014**		0.0038***				
	(2.270)		(3.440)				
HH_c		-0.0031	0.0238**				
		(0.542)	(2.245)				
$k_{c,0}$				0.0000***			
				(2.888)			
$k_{c,4}$					0.0001		
					(1.433)		
$k_{c,8}$						-0.0001**	
						(1.984)	
$k_{c,18}$							-0.0002***
							(2.700)
Obs	1,597	1,597	1,597	1,597	1,597	1,597	1,597
Adj. R-sq	0.003	-0.001	0.007	0.003	0.002	0.001	0.003
F-test	2.958	0.371	4.668	4.327	1.313	2.214	4.012
Prob$>F$	0.052	0.690	0.003	0.013	0.269	0.110	0.018

	8	9	10	11	12	13	14
	Dep. var.: 5-yr. periods moving avg. gr. rate of p/c GDP (1995-2003)						
$rgdpl$	-0.0000	-0.0000	-0.0000**	-0.0000	-0.0000	-0.0000	-0.0000**
	(1.475)	(0.727)	(2.146)	(0.786)	(1.641)	(0.885)	(2.331)
E_c	0.0013**		0.0038***		0.0013**		0.0038***
	(2.251)		(3.448)		(2.258)		(3.489)
HH_c		-0.0030	0.0239**			-0.0028	0.0245**
		(0.520)	(2.264)			(0.491)	(2.304)
$k_{c,18}$	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(2.674)	(2.692)	(2.692)	(3.334)	(3.311)	(3.326)	(3.340)
$k_{c,19}$				0.0063***	0.0063***	0.0062***	0.0063***
				(3.188)	(3.189)	(3.184)	(3.225)
Obs	1,597	1,597	1,597	1,597	1,597	1,597	1,597
Adj. R-sq	0.006	0.002	0.010	0.011	0.014	0.010	0.018
F-test	4.416	2.782	5.313	6.037	5.289	4.563	5.288
Prob$>F$	0.004	0.040	0.000	0.000	0.000	0.001	0.000

Notes. Cross-section regressions using White's consistent estimator. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is moving averages growth rates for all 5-year periods starting between 1995 and 2003 as dependent variable. Explanatory variables take the initial value of each 5-year period. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. $k_{c,i}$ is the i-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.11: Short term growth regressions with moving averages. Country fixed effects and control variables.

	1	2	3	4	5	6	7	8	9	10
	Dep. var.: avg. growth rate of per capita GDP (5-year periods starting every year in 1995-2003)									
v^*=	35	45	55	65	no filter	35	45	55	65	no filter
rgdpl	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***
	(4.160)	(4.161)	(4.292)	(4.292)	(4.578)	(5.808)	(5.812)	(5.866)	(5.866)	(6.257)
E_c	-0.0022	-0.0005	-0.0028	-0.0028	-0.0023	0.0010	0.0010	0.0012	0.0012	-0.0003
	(0.320)	(0.079)	(0.477)	(0.476)	(0.372)	(0.167)	(0.190)	(0.245)	(0.245)	(0.057)
HH_c	0.0605	0.0653	0.0353	0.0354	0.0406	0.0173	0.0138	0.0094	0.0094	-0.0039
	(1.065)	(1.273)	(0.998)	(1.000)	(1.070)	(0.450)	(0.401)	(0.416)	(0.416)	(0.165)
ltertiary	0.0245***	0.0252***	0.0235***	0.0235***	0.0247***	0.0238***	0.0225***	0.0212***	0.0212***	0.0214***
	(4.453)	(4.775)	(4.658)	(4.660)	(4.889)	(4.630)	(4.726)	(4.918)	(4.918)	(5.048)
leducexp	0.0028	0.0012	0.0013	0.0013	0.0017	-0.0041	-0.0048	-0.0050	-0.0050	-0.0023
	(0.330)	(0.146)	(0.163)	(0.164)	(0.189)	(0.513)	(0.637)	(0.673)	(0.673)	(0.316)
poor	-0.0111	-0.0099	-0.0068	-0.0068	-0.0028	-0.0003	-0.0001	0.0023	0.0023	0.0031
	(0.942)	(0.915)	(0.653)	(0.653)	(0.265)	(0.041)	(0.015)	(0.326)	(0.326)	(0.462)
middle	-0.0092**	-0.0085**	-0.0094**	-0.0094**	-0.0091**	-0.0086**	-0.0086**	-0.0090***	-0.0090***	-0.0090***
	(2.127)	(2.040)	(2.202)	(2.202)	(2.119)	(2.443)	(2.596)	(2.712)	(2.712)	(2.713)
$k_{c,18}$	-0.0002***	-0.0002***	-0.0001***	-0.0001***	-0.0001**	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(3.683)	(3.527)	(3.251)	(3.247)	(2.592)	(5.347)	(5.178)	(4.708)	(4.708)	(4.815)
Obs	800	845	874	878	908	751	794	823	824	845
Countries	121	133	139	142	148	120	132	138	139	145
Adj. R-sq	0.154	0.162	0.150	0.150	0.138	0.192	0.183	0.176	0.176	0.182
F-test	6.832	7.362	7.035	7.033	6.830	10.18	10.21	9.908	9.909	10.82
Prob> F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes. Fixed effect estimation with robust standard errors. Heteroskedasticity robust t-statistics in parentheses. Dependent variable is moving averages growth rates for all 5-year periods starting between 1995 and 2003 as dependent variable. Explanatory variables take the initial value of each 5-year period. *rgdpl* is PPP Converted GDP Per Capita (Laspeyres) at 2005 constant prices as reported in Heston et al. (2011). E_c is Theil's entropy index of diversification. HH_c is the Hischmann-Herfindahl concentration index. *ltertiary* is the logarithm of the gross rate of enrolment in tertiary education and *educexp* is the logarithm of the percentage of GNI devoted to educationas reported by the WDI Database. *poor* and *middle* are dummy variables indicating countries that are ranked as low or middle income countries according to the World Bank Classification of countries. $k_{c,18}$ is the 18-th component of the k_c vector. An omitted constant was also included in all specifications. * significant at 10%; ** significant at 5%; *** significant at 1%.

Institut de Recherches Économiques et Sociales
Université catholique de Louvain

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

IRES

ISSN 1379-244X D/2013/3082/08