Measuring the Economic Complexity at the sub-national level using international and interregional trade

Santiago José Pérez Balsalobre
Departamento de Economía Aplicada. Facultad de Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid; CEPREDE and L. R. Klein Institute, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid.
e-mail: santiagoj.perez@uam.es

Carlos Llano Verduras
e-mail: carlos.llano@uam.es

Jorge Díaz Lanchas
e-mail: Jorge.diaz@uam.es

Abstract
The Economic Complexity Index (ECI) elaborated by Hausmann and Hidalgo has open a new framework in the discussion of the reasons that explains the growth of countries in the long run. ECI was calculated by considering the international trade of countries. This paper pretends to apply this theory to the sub-national level in Spain estimating ECI at NUTS-3 level, to check if ECI was a good indicator of growth at the long-run in the sub-national level too. But for the first time, ECI is going to be calculated by considering the province to province trade level instead of the international trade. For this purpose, a huge dataset was estimated by considering the trade of the regions by different transportation modes at a very disaggregated product classification both in international and intranational trade.

Key Words: economic complexity; international trade; inter-regional trade; growth; Spain

JEL: H70, O41, O47, O56

1 Alternative titles: A new perspective of Economic Complexity: Province to province trade in the Spanish case; Measuring the Economic Complexity at the sub-national level using international and interregional trade
1. INTRODUCTION

The relevance of the accumulation and diffusion of knowledge as a key driver of long term economic growth is attracting increasing attention in the literature (Alvarez, et al., 2012; Andersen et al 2011; Balland and Rigby, 2016; Boschma, et al. 2015; Corrado and Hulten, 2010; OECD, 2013; Hausmann and Hidalgo 2011).

Hausmann and Hidalgo (2009, 2011, 2014) developed the concept of economic complexity (EC), defined as a national indicator able to measure the non-observable capabilities (know-how) required in the production of goods and services. This concept puts the emphasis in how the production process of a specific product/service implies the interaction of different specific knowledges, and how these interactions permit the innovation and production of more complex products, and therefore, in the configuration of a more complex and dynamic entire economy. In this paradigm, complexity is calculated through an iterative process between countries diversification level, the number of products that countries export with a revealed comparative advantage, and products ubiquity, the number of countries that have a revealed comparative advantage in the product (Hausmann and Hidalgo, 2014).

Countries with more capabilities will be able to make more products (higher diversification), while products that require more capabilities will be accessible to fewer countries (lower ubiquity). Thus, it is expected that more complex countries will be both more diversified and would make less ubiquitous products. Hidalgo and Hausmann (2009; 2011; 2014), also shown how these bimodal complexity indexes, based on the diversification and ubiquity concepts, correlates well with aggregate levels of per capita GDP. Moreover, they show how the error terms of the relationship are predictive of future growth.

Furthermore, Hidalgo et al. (2007) shown that the likelihood that a country develops a particular product depends on its proximity in the product space to the products that the country is already able to export. Countries with wider product links within the product space, are more able to diversify and grow in contrast to countries that produce products that are less connected with the production of alternative items (Hausmann and Klinger 2006). Thus, the product space of a country can also be used to explain the lack of convergence in the world economy (Hidalgo et al. 2007), or the resilience of national economies to external shocks (Hausmann et al., 2006).

Departing from the economic complexity paradigm at the national level, it is clear the interest to extending it to sub-national entities within countries. However, doing it is not straightforward, and is usually impeded by several conceptual and statistical limitations. Indeed, very few attempts are found in the literature.

Among the few exceptions, Reynolds et al, 2017, generated a rich datasets containing economic complexity indexes at the sub-national level for Australia, considering 9 regions and 506 sectors. Their paper made a number of contributions: i) First, they use a novel source of trade data: a high-resolution multi-regional input–output (MRIO) tables; ii) second, this powerful dataset allows them covering both trade of goods and services, a great improvement when trying to measure economic complexity; iii) it also allows considering trade flows in its widest perspective, that is, including flows with the rest of the World, and with the rest of the regions in the country. As the authors suggest, such improvements were able due to the recent developments of global and regional MRIO databases (Lenzen, Moran, Kanemoto, & Geschke, 2013; Lenzen et al., 2014).

Specifically for the Australian case, Wood and Lenzen (2009) also provided a first estimate of economic complexity using input-output data, covering the period from 1975 to 1999, and using 344 intermediate-industry-sector resolution. In contrast to the work done by Reynolds et al, 2017, the previous attempt just investigated the developments of economic structure solely at the national level. Instead, Reynolds et al (2017) applied Hausmann and Hidalgo’s EC concept to the Australia’s interstate and international exports. Regarding the results, some are of special interest for our analysis: using a 2009 Australian multi-regional input–output table, they found that small differences in industrial capability and knowledge are crucial to relative complexity. The majority of states (especially Western Australia)
export primarily resource-intensive goods, yet interstate trade has many complex products that are not currently internationally exported.

Another interesting example is found in Gao and Zhou (2017), who also working within the Hausmann and Hidalgo framework, estimated the economic complexity of China’s provinces between 1990 to 2015. Their main novelty relied on the use of firm level data, which allowed them taking into account international as well as intra-national trade flows of goods and services. As these authors remarked, ‘most of previous economic complexity analysis are based on the world trade data, meaning that industries without exporting products are excluded, such as services. However, not only goods but also services are important to measure economic complexity as the growth in service and its sophistication can provide an additional route for economic growth’. They also wondered about how the link between the economic growth and the economic complexity varies among regions, when this wider definition of trade is used instead of the one just based on international flows of goods. Their results shown that, after estimating the Economic Complexity Index (ECI) of China’s provinces, diversified provinces tend to have industries of less ubiquity. They also verified certain stability in the evolution of the provinces’ rankings by ECI, with provinces located along the coast having higher economic complexity. Then, after linking ECI to the economic development and the income inequality, they find that the predictive power of ECI is positive for the former but negative for the latter. Finally, they compare different measures of economic diversity and explore their relationships with monetary macroeconomic indicators.

Also connected with the Hausmann-Hidalgo paradigm, but using an alternative source of information, Balland & Rigby (2016) offer another interesting example about how to measure the economic complexity at the sub-national level, in this case, for the United States (US). In contrast to the previous articles cited, they measure the distribution and the evolution of knowledge complexity in US cities using an extensive dataset with more than two million patent records from the US Patent and Trademark Office. Following Hausmann and Hidalgo bimodal approach, they identified the techno-logical structure of US metropolitan areas in terms of the patent classes in which they are most active between 1975 and 2010. They find that knowledge complexity is unevenly distributed across the United States and that cities with the most complex technological structures are not necessarily those with the highest rates of patenting. Citation data indicate that more complex patents are less likely to be cited than less complex patents when citing and cited patents are located in different metropolitan areas. They also explore if the spatial diffusion of knowledge is linked to complexity, finding that in general, more complex technologies are less likely to be cited in distant locations, pointing out to a strong level of concentration of highly complex activities in very few [smart and productive] locations.

In the context of these previous articles, the main objective of our paper is to compute and analyse the concept of economic complexity at the sub-national level in Spain, covering the 50 provinces (Nuts 3 following Eurostat terminology) and 17 regions (Nuts 2 following Eurostat terminology) during the period 1995-2015. In line with the previous mentioned articles (Reynolds et al, 2017; Gao and Zhou, 2017), our main contribution is on the combination of inter-national and intra-national trade for each province and product, bridging the gap between the [desired and usually unobservable] complexity at the production level, and the [observed but maybe biased] complexity measure based on [just] international trade flows2. Due to data restrictions, trade of services is not considered in our case. Regarding the estimation of the economic complexity based on international trade flows, we depart from Hausmann and Hidalgo indexes of complexity, provided by The Observatory of Economic Complexity and the Atlas of Economic Complexity (2011). Then we develop a methodology able to derive ECI for 50 Spanish provinces, considering a high level of product disaggregation. Then, an additional dataset is

---

2 As Hausmann and Hidalgo has pointed out in several occasions, although their focus on the link between production and knowledge, their empirical analysis is mainly concentrated on the use of international trade data. The reason for this jump from knowledge to production, and from production to international trade is mainly explained by the availability of data: while the data on the accumulation and use of knowledge in the production process is scarce and incomplete, the one for international trade is highly standardized and disaggregated for most countries in the world. Accepting this practical sacrifice, it is convenient to have a more detail view of its implications. For example, having in mind the current literature on the internationalization of heterogeneous firms and the border effect literature, one may wonder to what extent a measure of economic complexity built on just international trade exports may be biased, since just the most productive firms are usually able to export abroad.
computed containing ECI indexes that rely on the inter-provincial trade flows within Spain (intra-national trade) for the same period. As said before, our approach is similar in this respect to the one applied in the most recent papers on the topic (Reynolds et al, 2017; Gao and Zhou, 2017), although diverges in the use of bilateral deliveries by transportation statistics, rather than on the use of input-output tables, firm level data or inter-city patent citations. To the best of our knowledge, our analysis is the first attempt to estimate economic complexity indexes at the sub-national level in Europe, which combines actual inter-national and intra-national trade flows of goods, covering a large period (1995-2015), and with a high level of product and spatial disaggregation (NUTS-3). Moreover, our approach allow us to add an interesting new layer not considered before in any other analysis available in the literature: to analyse the complexity indexes by the transport mode used in the deliveries, both when computing the ECI for the international and the intra-national flows. To this regard, we remark the hidden link between the transport mode and the value/volume (proxy of quality) of the products being delivered, a factor not expressly considered as a source of complexity in the Hausmann-Hidalgo framework. Since one important aim of this paper is to explain the methodology for producing a dataset that will be maintained and used afterwards by several researchers and institutions for policy making, we put a strong emphasis on the transparency of every step taken in the estimation of the indexes. With this purpose, the paper considers three alternative variants of the methodology, whose results are then compared and tested in the growth model.

The structure of the article is the following: First, we revise the theoretical background associated with the concept of economic complexity. Second, we describe the data and methodology used in the construction of the economic complexity indexes for the Spanish provinces. The next section contains the descriptive analysis, which allows identifying the main features of the indexes obtained and their relationship with other variables, and their evolution overt time. Finally, we develop an econometric analysis to analyse the obtained data and its validity.

2. Theoretical background

As suggested before, the concept of economic complexity was developed by Hausmann, Hidalgo et al. (2011). The dataset produced by the Observatory of Economic Complexity includes, for every year and country, two main indexes that try to measure the economic complexity of products and countries: the Product Complexity Index (PCI), and the Economic Complexity Index (ECI).

For these authors, “The Product Complexity Index is a product characteristic. (...) It is calculated as the mathematical limit of a measure based on how many countries export the product and how diversified those exporters are”.

Due to the division of labour, inherent to modern systems of production, individuals develop a set of skills and knowledge that are combined in different stages of the production process. The concept of economic complexity emphasizes, specifically, the interaction between individuals with different knowledges. Then, the capacity of producing a product reveals the presence of certain capabilities. This intuition is applicable for the construction of an airplane, the elaboration of a medicament or for growing apples. The complexity of products increases with the number of different capabilities to be combined in the production process. Furthermore, the complexity of countries increases when it is able to combine a wide set of capabilities, being able to produce a diversified range of products.

In addition to that, these authors also introduce the concept of ‘singularity’, thanks to which some specific products are considered as complex not because of the knowledge embodied in their production process, but, because of the scarcity of the inputs with which they are made off. Haussman, Hidalgo et al (2011), condense all this in the concept of ‘ubiquity’. This ubiquity reflects the number of countries that make a product. As commented before, usually ubiquity and diversity are negatively related: The larger the number of countries producing and exporting a product (high ubiquity), the lower the level of diversity associated with this product. On the other hand, the products that are traded by few countries (low ubiquity), either because the product requires very special knowledge that not many countries have, or because it is a very exclusive product, are going to have high levels of complexity.
By the combination of these two concepts of ubiquity and diversity, and departing the H&H product complexity index (PCI), they obtain the concept of “Economic Complexity Index” (ECI), defined as a country characteristic that measure ‘how diversified (...) a country’s export basket is. It is calculated as the mathematical limit of a measure based on how many products a country exports and how many exporters each product has.’

Departing from these definitions, the two sets of indexes PCI and ECI are computed by the Observatory of Economic Complexity for each product, country and year. Then, these indexes as been proved to be of great interest explaining countries income level and growth potential. The rational is as follows: the complexity reflects what the country is capable of making, given its specific factor and knowledge endowment. It also defines their position in the product space, indicating the likelihood of producing alternative products to the ones that they are able to produce and export now. Thus, the ECI is directly related with the productive structure of the countries now, but also give relevant information about its capacity to transform into a more complex economy in the next future, evolving towards the core in the product space. Haussman, Hidalgo et al (2011), explain this relation and shows how the complexity index is related with the income per capita of the countries. As can be seen in Figure 3.1, there is a clear and positive relationship between the complexity of a country for a year and the income per capita of the next year.

Figure 1: Relationship between the Income per capita in USD for the year 2009 and the Economic Complexity Index of the year 2008 for the countries in the world.

In addition the authors remark another important conclusion: the countries that have a level of economic complexity higher than the one that corresponds to their level of income per capita, are going grow faster in the following years, than the ones that are actually in the correspondent’s levels of income with respect to their complexity. To exemplify this relation, Figure 3.1 compares the ECI (2008) against the country income per capita in 2009. Some cases are especially illustrative: on the one hand, Qatar (QAT) or Kuwait (KWT), stand out for being countries with very large levels of income per capita but low levels...
of economic complexity. Their productive structure does not reflect high levels of collective knowledge: i) they produce low complex products; ii) they are too specialized in one type of product. On the other hand, China and Ukraine are countries with larger complexity indexes than the ones that correspond to an average country with its income. This can be considered as a sign that these countries account for a collective knowledge that is under-exploited. Thus, the capacity of using it conveniently in the future, moving into the product space towards the production of more sophisticated or diversified basket of products, will end up in a better growth performance.

3. Data

The construction of product and economic complexity indexes at the sub-national level is a hard task, if not impossible due to data restrictions. In fact, the exact replication of Hausmann and Hidalgo methodology at the sub-national level will require the use of an international trade dataset that includes the bilateral flows between equivalent sub-national entities within each country in the world, with enough product disaggregation. Just in this case, the exact methodology applied in the Atlas of economic complexity, can be replicated. Moreover, if one wants to add the intra-national trade of goods and services, in line to what has been applied in specific countries such as China or Australia, the needs for additional data increases exponentially.

Instead, following Hausmann and Hidalgo pragmatically approach, the aim of this article is to develop a methodology able to produce the best possible complexity indexes at hand, given the information available in Spain for inter-national and intra-national trade at the lowest spatial level (provinces, Nuts 3). The main data required for this purpose is summarized in Table 1.

Table 1: Summary of trade data used

<table>
<thead>
<tr>
<th>Trade flow</th>
<th>Territorial Dissagregation</th>
<th>Transportation mode</th>
<th>Initial product dissagregation</th>
<th>Time period</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>International</td>
<td>NUTS-3</td>
<td>Ship</td>
<td>More than 10000</td>
<td>1995-2015</td>
<td>AEAT(^4)</td>
</tr>
<tr>
<td>International</td>
<td>NUTS-3</td>
<td>Train</td>
<td>More than 10000</td>
<td>1995-2015</td>
<td>AEAT</td>
</tr>
<tr>
<td>International</td>
<td>NUTS-3</td>
<td>Road</td>
<td>More than 10000</td>
<td>1995-2015</td>
<td>AEAT</td>
</tr>
<tr>
<td>International</td>
<td>NUTS-3</td>
<td>Aircraft</td>
<td>More than 10000</td>
<td>1995-2015</td>
<td>AEAT</td>
</tr>
<tr>
<td>Intranational</td>
<td>NUTS-3</td>
<td>Ship</td>
<td>62</td>
<td>1995-2015</td>
<td>C-intereg</td>
</tr>
<tr>
<td>Intranational</td>
<td>NUTS-3</td>
<td>Train</td>
<td>51</td>
<td>1995-2015</td>
<td>C-intereg</td>
</tr>
<tr>
<td>Intranational</td>
<td>NUTS-3</td>
<td>Road</td>
<td>171</td>
<td>1995-2015</td>
<td>C-intereg</td>
</tr>
<tr>
<td>Intranational</td>
<td>NUTS-3</td>
<td>Aircraft</td>
<td>171</td>
<td>1995-2015</td>
<td>C-intereg</td>
</tr>
</tbody>
</table>

Regarding the intra-national trade flows relys on the C-intereg dataset, which includes inter-provincial trade flows between the 50 Spanish provinces at the NUTS-3 level, covering a long period of time (1995-2015). This dataset is estimated every year using the best information available in the country regarding bilateral flows by the main transport modes (road, railway, ship, aircraft). The level of disaggregation by products in each on of this transport mode is different. A maximum common level of disaggregation for all transport modes corresponds to 30 types of products. However, some transport modes provide a larger range of product. For example, in the case of the road (the main mode for intra-national deliveries) and the aircraft, we work with 171 different products, based on the NST classification. Also for the intra-national deliveries, the product disaggregation available for deliveries by railway and ship are: AI-RENFE-51 for the former and PdE-62 for the later. In sum, our point of departure for the intra-national trade flows in this long period of time corresponds to around 25 million observations, and considerable

\(^3\) Note: 50% of Qatar’s exports in 2008 correspond to crude petroleum, while another 45% to liquefied hydrocarbons. This is, 95% of the total exports are concentrated in two types of products. The high concentration of Qatar’s exports may reflect lack of enough collective knowledge to produce/export alternative products. This will impede diversification, what can be reflected in lower growth rates in the future. The same happens in the case of Kuwait, which concentrates a 92% of the exports in crude petroleum.

\(^4\) AEAT: State Tax Administration Agency
strong base for any analysis regarding any movement of products occurred within the country. We want to remark here that such deliveries are much larger than the ones corresponding to the international deliveries. Moreover, as it has been reported in previous papers mixing this dataset with the official international trade flows for Spain, the range of products and different destinations are also wider than for the international deliveries (Diaz-Lanchas et al, 2017).

The international trade flows corresponds to the official files published on yearly basis by the Spanish Tax Agency (AEAT) at the province level (Nuts-3), with the largest product disaggregation available (CN8-digits product classification). In sum an initial international dataset of around 38 million observations for the considered period. All these dimensions of our initial dataset are summarized in Table 1.

The other key ingredient for building database is the Product Complexity Index elaborated by Hausmann and Hidalgo (2014). This index is available from year 1995 to 2015, at HS4 product disaggregation level (around 1,240 products per year). This data goes from a maximum value of 7.715 to a minimum value of -5.406. To avoid possible problems with the different methodological approaches, this data is going to be normalised, being the maximum (1) and minimum (0) values, the same obtained by Hausmann and Hidalgo.

For the econometric analysis, some control variables are added. That variables (Net capital of the region per capita and Economic value of the Human Capita) are extracted from the IVIE datasets for the period 1995-2013.

4. Methodology

4.1 Economic complexity index calculation.

Before going into the possible estimations for sub-national complexity indexes it is important to summarize the Hausmann methodology to have in mind the main phases in the calculation of their indexes and the limitations of the ones calculated in this paper.

As the data explained in section 3, the data used for the calculation of the PCI and ECI are based in export data because of their availability rather than production data. But for making comparable the different countries in the world they used the Balassa’s definition of Revealed Comparative Advantage (RCA). That is, they correct the export data by excluding the non-competitive products in the international trade market. That RCA can be written as:

\[ RCA_{cp} = \frac{X_{cp}}{\sum_c X_{cp}} \left/ \frac{\sum_p X_{cp}}{\sum_c \sum_p X_{cp}} \right. \]  (1)

Where, \( X_{cp} \) represents the total exports of a specific product (\( p \)) by each country (\( c \)). This result in a value’s series specific for each year, where each product in each country obtains different values. This is transformed in a matrix of 0 and 1 by replacing with a 1 when \( RCA_{cp} \geq 1 \) and 0 otherwise. With all this \( RCA_{cp} \) data, they construct the matrix \( M_{cp} \) and is going to be the key tool for the construction of the index.

By using this matrix, they calculated the Diversity and Ubiquity defined as:

\[ Diversity = k_{c,0} = \sum_p M_{cp} \]  (2)

---

5 This PCI value corresponds with the product 7805 (Lead tubes, pipes and fittings) at the HS4 product classification in 2012.

6 This PCI value corresponds with the product 5303 (Jute and other textile fibers) at the HS4 product classification in 2010.
Ubiquity = \( k_{p,0} = \sum_c M_{cp} \)  

This measure is generated in a more accurate way by using each one of the measures to correct the other. That is, correct the diversity of countries with the average ubiquity of that products. And for products, correct the ubiquity of the different products, by the average diversity of countries that produce these products. This can be expressed by the recursion:

\[
k_{c,N} = \frac{1}{k_{c,0}} \sum_p M_{cp} \cdot k_{p,N-1} \\
k_{p,N} = \frac{1}{k_{p,0}} \sum_c M_{cp} \cdot k_{c,N-1}
\]

And by introducing (5) in (4), the term \( k_{c,N} \) can be finally rewritten as:

\[
k_{c,N} = \sum_{cr} \overline{M_{ccr}} \cdot k_{cr,N-2}
\]

Where:

\[
\overline{M_{ccr}} = \sum_p M_{cp} M_{cpr} \frac{k_{c,0} k_{p,0}}{k_{c,0}}
\]

By having that, ECI and PCI can be calculated, and they are finally defined as:

\[
\overline{ECI} = \frac{\overline{K} - <\overline{K}>}{stdev(\overline{K})} \\
\overline{PCI} = \frac{\overline{Q} - <\overline{Q}>}{stdev(\overline{Q})}
\]

Where \( \overline{K} \) is the eigenvector of \( \overline{M_{ccr}} \) associated with second largest eigenvalue and \( \overline{Q} \), the eigenvector of \( \overline{M_{ppr}} \) associated with second largest eigenvalue. Finally, \(< >\) represents an average and stdev stands for the standard deviation.

4.2 Methodology 1: International and Intra-national ECI. A first approximation.

To calculate a totally comparable sub-national ECI to the one calculated in previous section for countries, a complete dataset with all NUTS-3 regions in world will be required. If not, applying the same methodology considering only Spanish regions, can drive important bias. The most important is the one related with ubiquity. Can happen that there are some products that are very commonly exported in one country, and consequently, that product present a high ubiquity. But it can happen that this product can be not so common in world because it is only produce on that country. So, it will be dangerous to apply the same methodology if not considering all regions around world.

On the other hand, having the world PCI calculated yet, the different regional complexities can be calculated by different ways. But there are some important data treatment aspects that are needed to be mentioned before.

Having our international trade dataset of the Spanish regions at the same HS4 product disaggregation that Hausmann, a first alternative to calculate sub-national economic complexity is assuming the same product complexity that the one obtained by them. The problem with that alternative, is the comparability with the intranational trade, where disaggregation is lower.
For the following methodologies, the calculations are going to be done considering the euro trade flows. Additionally, the same methodologies are going to be applied considering the ton trade flows. In this paper the attention is going to be in the euro trade flows, that are the comparable PCI and ECI with the one calculated in the world level.

Another important aspect of the data used for this methodology, is that the flows smaller to 500 euros are going to be remove in the international dataset to guarantee a great similarity with the intranational ones and avoiding the effect of possible strange trade flows in some specific years or transportation modes. So, below this reference line the flows are not going to be considered relevant enough.

Finally, like our interest is the study of the intranational trade flows, but considering the inter-regional exports between the regions of the country, the intraregional trade flows of each region are also removed.

Starting now with the methodology calculations. Firstly, it is needed the elaboration of a new PCI based on the ones obtained by Hausmann. Now, PCI are going to be calculated at a more aggregate product classification, that are going to be different in each transportation mode, as shown in Table 1. This allow to impute this new aggregate PCI to the intra-national trade flows and make possible to calculate the new intra-national ECI.

With this methodology, the assumption of having the same product complexity that the ones obtained considering all world trade is also respect. When aggregating product complexity for obtaining the new PCI at a more aggregate level is going to differ in the regions, but only because of the different product’s basket of the region at HS4 level. So,

\[
P_{\text{CI}}_{jrtp} = \frac{\sum_{j} PCI_{jrtp} X_{jrtp}}{\sum_{j} X_{jrtp}}
\]

(10)

Being \( PCI^0 \), the different product complexities at HS4 level obtained by Hausmann and \( X_{rtp} \) the international exports of each product \((p)\) included in each NST/PdE/Ai product \((j)\) in the aggregated classifications, in each region \((r)\) by each transportation mode \((t)\).

This new PCI are going to be the same when considering international and intra-national trade and are going to be the initial point for the calculation of intra-national complexity indexes.

Continuing with the complexities considering international trade. Having this PCI is now possible to calculate the complexity of regions. The process is going to be the as follows, considering the product classification \((j)\) for each transportation mode \((t)\) instead of the \((p)\) classification, and considering the more aggregated PCI, after aggregating all regions:

\[
P_{\text{CI}}_{jt} = \frac{\sum_{j} PCI_{jrtp} X_{jrtp}}{\sum_{j} X_{jrtp}}
\]

(11)

So, imputing the new product complexity, equal in all Spanish regions, to the exported products and aggregating this new data by each region, ECI can be calculated:

\[
ECI_{rt} = \frac{\sum_{r} PCI_{jt} X_{jt}}{\sum_{r} X_{jt}}
\]

(12)

\[
ECI_r = \frac{\sum_{r} ECI_{rt} X_{rt}}{\sum_{r} X_{rt}}
\]

(13)

Also, a new PCI can be calculated, and is the PCI aggregating all transportation modes. This is done for having a common measure that is going to be comparable with the intra-national trade. So, it is going to be possible to compare the complexity of products in international and intra-national trade. Having in mind that this is going to represent only a comparison of the complexity of the basket of products included in each aggregate product because of the aggregation of the different previous classifications.
For this aggregation, can be used the classification of the C-interREG project (29 products, or R-29) that is a common classification for all transportation modes. For that, first is needed to calculate a new PCI at a more aggregate level (14) and finally, the aggregation of the different transportation modes and regions (15).

\[
P_{CI_{k,t}} = \frac{\sum_{j,t} P_{CI_{j,t}} X_{k,t,j}}{\sum_{k,t} X_{k,t,j}} \tag{14}
\]

\[
P_{CI_{k}} = \frac{\sum_{k} P_{CI_{k,t}} X_{k,r,t}}{\sum_{k} X_{k,r,t}} \tag{15}
\]

For the intra-national trade flows, the calculations are going to be the same that the ones described when considering the international trade, but imputing to the intra-national trade dataset the PCI calculated in (10). For differentiating both complexity indexes, the PCI and ECI calculated with international flows are going to be called from now on Inter-PCI and Inter-ECI, and the ones calculated with intra-national flows, Intra-PCI and Intra-ECI.

### 4.3 Methodology 2: International and Intra-national ECI. Modifications of the index.

There is a second alternative estimating international and intra-national complexity indexes at the subnational level. The previous methodology, although is consistent with the methodology of Hausmann, not have into account the diversity and ubiquity of regions and products respectively. So, with the previous methodology, ECI is really the complexity of the basket of products exported by the region in the international level and in the intranational market.

Another alternative is to consider that the internal diversity and ubiquity of products is also important. Particularly when talking about intra-national complexity. For doing this, the first step is the same like in first methodology. That is, calculate an aggregate PCI to impute in the intra-national flows. This is shown in equation (10).

This first step is equal to the one of the second methodology, because although now the intra-national diversity and ubiquity issues are going to be into consideration, to consider the international complexity calculated by Hausmann is key for not obtaining bias results. That is, maybe one product can be very ubiquitous inside the frontiers of a country, but can be the only country producing this product. So, in the world market, this product obtains total opposite results. So, departing from the Hausmann’s PCI and incorporating this new ubiquity and diversity measures, allow us control both aspects.

Having calculated \( P_{CI_{j,r,t}} \) the following step consist in the elaboration of a RCA matrix like the process showed in equation (1).

\[
RCA_{j,r,t} = \frac{X_{j,r,t}}{\sum_{j,r,t} X_{j,r,t}} / \frac{\sum_{j,r,t} X_{j,r,t}}{\sum_{j,r,t} X_{j,r,t}} \tag{16}
\]

Representing now \( X_{j,r,t} \) the international exports of each product (p) included in each NST/PdE/Ai product (j) in the aggregated classifications, in each region (r) by each transportation mode (t), and \( \sum_{j,r,t} X_{j,r,t} \) the total international exports of Spain, instead of the total world trade. For the intra-national flows, the process is going to be the same but considering \( X_{j,r,t} \) the intranational flows instead of the international ones.

By doing this, we can construct the matrix with the different RCA that is going to be called, \( M_{j,r,t} \), where all the RCA larger or equal than 1 are going to be substituted to 1 and the other with a 0. Having these matrices for each transportation mode, all the flows of the regions are going to be multiplied by the matrix. So, the flows with an RCA of 0 are going to be eliminated. The diversity and ubiquity are going to be calculated with the matrices data as follows:
\[
Diversity = k_{rt} = \frac{1}{J_t} \sum_j M_{jrt}
\]
\[
Ubiquity = k_{pt} = 1 - \frac{1}{R} \sum_r M_{jrt}
\]

Being \(J_t\) the total number of products NST/Pde/Ai depending on the transportation mode considered \((t)\) and \(R\) the total number of regions in Spain \((50)\). So, the diversity measure is going to point largely to regions with higher number of products in their export’s basket. On the other hand, the ubiquity measure, is going to put lower, if the products are present largely in the Spanish regions.

Having this measures calculated, is time to continue with the calculations of the product and economic complexities departing from the \(PCI_{jrt}\) calculated in equation (10)
\[
PCI_{jt} = \frac{\sum_{jt} PCI_{jrt} X_{jrt} M_{jrt}}{\sum_j X_{jrt} M_{jrt}} k_{pt}
\]

Then, having this PCI calculated, ECI can be computed by assigning these new product complexities to the international trade flows, as in the process followed in equations (12) and (13) but now introducing also the diversity measure. So, ECI is going to be:
\[
ECI_{rt} = \frac{\sum_{rt} PCI_{jrt} X_{jrt} M_{jrt}}{\sum_r X_{jrt} M_{jrt}} k_{rt}
\]
\[
ECI_r = \frac{\sum_r ECI_{rt} X_{rt}}{\sum_r X_{rt}}
\]

As in previous methodology, the intranational indexes are calculated by using the same procedure but with the intranational trade flows dataset.

5. Descriptive analysis.

We start our analysis scanning the raw data with the aim of providing a first snapshot of the different types of trade flows combined in the methodology, and their possible influence in the final indexes.

First, we show how much trade (exports in euros) correspond to international or intra-national deliveries. To do that, Figure 2 shows the percentage of international exports over the total exports by provinces, both in 1995 and 2015. The results clearly points out to a stable trade structure during the period, something that has also been observed in other papers (Gao and Zhou, 2017). There are some exceptions, particularly the two provinces of the Canary Islands and Ourense. But in general, the same provinces with the highest international export shares in 1995 are also the ones more internationalised in 2015. Some results are worth mentioning regarding the provincial heterogeneity in terms of the international trade share: first, we observe that the mean remain stable around a 20% in both years; however, we find important differences across-provinces, finding cases such as Barcelona or Pontevedra with around a 40% in the two considered years, and others such as Segovia or Teruel with less than a 10% of international exports.

This preliminary result points out to the relevance of computing the economic complexity indexes considering all type of trade, and not just the international deliveries. As said, Hausmann and Hidalgo methodology relies just on international trade of goods due to data limitations for international comparisons. There, for cross-country analysis, assuming international exports of goods as a proxy may be acceptable, or at least, the best one can get. But in view of what is shown here, when the analysis zooms into the sub-national trade structure, the inclusion of intra-national trade flows could be fundamental, mainly for the regions with not so much international trade.
It is also convenient to have an initial look into the trade structure by products. With this aim, Figure 3 shows the international export share for each one of the 30 products available for the international and the intra-national flows. As in the previous case, the means of the share of international exports does not change a lot, being almost the same in the two years considered. The behaviour is like the one observed in the regional analysis. In general, the products that are more international exported in 1995 are the ones that are more exported internationally nowadays. There are important exceptions like the R12 (Minerals) or R25 (Textiles) that increment the intra-national trade share during the period; more surprisingly, products like R21 (Plastic products) more than double its international export’s share. Overall, it is interesting to see how some products show high international export shares (R26-Leather and footwear; R18-Construction materials), with ratios close to 70%; while others are clearly more oriented towards the domestic market (R9-Coals); R4-Wood).

Figure 2: The evolution of the international exports shares by provinces: 1995 vs 2015.

Figure 3: The evolution of the international exports shares by product (R30). 1995 vs 2015.
Next, we analyse the diversity and ubiquity dimensions of our two sets of trade flows. We want to emphasize that both measures are the key aspects when calculating the PCI and ECI, both at the national and sub-national level.

First, our analysis starts with the international trade raw data at the same level than the one considered by Hausmann and Hidalgo for the entire world (HS4). The results are reported in Figure 4, which shows, on the one hand, the number of products exported abroad by each province (diversity), and on the other, the ubiquity, measured as the average of the ubiquity\(^7\) of the different products that each province export abroad.

The result clearly shows how during the period of study, every Spanish province has improved in term of the two dimensions, since, in general, all provinces export more products abroad (more diversity), while each product is traded by a larger set of provinces (higher ubiquity). We also observe certain degree of consistency between the patterns of these two dimensions in both years. The provinces with largest diversity levels, are the ones also the lowest ubiquity, indicating that, in both years, they export abroad not so common products. On the other hand, the ones with not a large variety of products in their international export baskets, have products with a high ubiquity. In fact, the mean ubiquity of some regions in 2015 is around 40. That means that the products that these Spanish provinces exported abroad are like the ones that all the other province exported too.

---

\(^7\) This ubiquity is calculated as the number of provinces that exported that product. That is, the presence of the product in the international export basket of each region.
Figure 4: Diversity and Ubiquity by province in 1995 and 2015 considering raw international data exports (HS4 level).

For making an equivalent analysis with the intranational data it is necessary to compute in our ‘methodology 1’ diversity and average ubiquity measures. A similar analysis can be done with the diversity and ubiquity measures calculated for the intra-national trade flows in our ‘methodology 2’. We use that estimates for plotting them split by IV panels in Figure 5.

Panel I shows the equivalent graph than Figure 4, but using international trade flows once it has been treated according to our ‘methodology 1’, that is, not using raw data, but after treating the diversity and ubiquity measures using the RCA’s. The first important conclusion that arise is that the graph doesn’t change too much, which can be seen as a sign of consistency between both methodologies, because guaranty that the results are not disturbed by the aggregation of the different transportation modes or the more aggregated products. It is just worth mentioning that, with that new measures, corrected with the RCA, the ubiquity of any province gets closer to the maximum level 1 (or 50 in the previous Figure).

We now focus on the results regarding the intra-national trade. Panel II shows a complete different pattern for the binomial diversity-ubiquity in this type of flows. Instead of trading a larger variety of products, when considering intra-national trade, the range of product traded domestically by the Spanish provinces decreased, on average, from 1995 to 2015. This paradoxical result can be observed by the average vertical orange line (2005) in Panel II, which is on the left to the vertical blue line (1995). Note that the opposite happened in Panel I for inter-national trade, where the range of different products exported abroad has increase with time.
Panel III and IV dig deeper on the differences between the two types of trade and their evolution in time. Panel III compares the binomial diversity-ubiquity of both types of trade in 1995. Intranational trade has, in general, more ubiquity and diversity than international trade. Panel IV shows the same results for 2015, obtaining a similar pattern than in 1995, but with less marked differences. Another important result is the reduction of the diversity range in year 2015 for the intranational trade. The interpretation is as follows: the range of diversity across-provinces observed in 1995 for the intranational trade was larger, due to the presence of certain provinces able to produce and trade domestically a large variety of products, while others, were just able to trade few products; By contrast, this situation changes in 2015, showing that, although there are regional differences, the diversity of different products

Having in mind the recent literature on trade and firm heterogeneity (Melitz, 2003), it is reasonable to assume that exporting products abroad is less likely than to trade them domestically. Moreover, if common institutions and national factor endowments are behind part of sectoral specialization, it is more likely that provinces within a given country share these factors, increasing the ubiquity of certain products when considering intra-national trade (i.e. almost each province in Spain produce and trade domestically an specific variety of wine; however, it is not so common that this provinces are able to export them internationally; moreover, several countries in the world are not able to produce and export abroad even a gallon)
traded domestically by the 50 regions is much more similar; although this results requires more research, it can be associated with the impact of the economic crisis, the disappearance of important sectors linked to the construction sector, and the corresponding structural change occurred during the recovery.

5.1 PCI and ECI: Results by the different methodologies.

Once this initial analysis is done, it is time to go into the results of the different methodologies proposed. We start by analysing the product complexity index (PCI), reported in Table 2.

Table 2: PCI for the R30 product classification considering different methodologies, trade flows and years.

<table>
<thead>
<tr>
<th>PCI</th>
<th>Methodology 1</th>
<th>Methodology 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.496</td>
<td>0.486</td>
</tr>
<tr>
<td>R2</td>
<td>0.359</td>
<td>0.439</td>
</tr>
<tr>
<td>R3</td>
<td>0.344</td>
<td>0.363</td>
</tr>
<tr>
<td>R4</td>
<td>0.295</td>
<td>0.296</td>
</tr>
<tr>
<td>R5</td>
<td>0.398</td>
<td>0.439</td>
</tr>
<tr>
<td>R6</td>
<td>0.343</td>
<td>0.338</td>
</tr>
<tr>
<td>R7</td>
<td>0.327</td>
<td>0.427</td>
</tr>
<tr>
<td>R8</td>
<td>0.451</td>
<td>0.467</td>
</tr>
<tr>
<td>R9</td>
<td>0.489</td>
<td>0.430</td>
</tr>
<tr>
<td>R10</td>
<td>0.450</td>
<td>0.450</td>
</tr>
<tr>
<td>R11</td>
<td>0.344</td>
<td>0.342</td>
</tr>
<tr>
<td>R12</td>
<td>0.361</td>
<td>0.304</td>
</tr>
<tr>
<td>R13</td>
<td>0.578</td>
<td>0.549</td>
</tr>
<tr>
<td>R14</td>
<td>0.550</td>
<td>0.527</td>
</tr>
<tr>
<td>R15</td>
<td>0.414</td>
<td>0.410</td>
</tr>
<tr>
<td>R16</td>
<td>0.373</td>
<td>0.377</td>
</tr>
<tr>
<td>R17</td>
<td>0.532</td>
<td>0.542</td>
</tr>
<tr>
<td>R18</td>
<td>0.445</td>
<td>0.468</td>
</tr>
<tr>
<td>R19</td>
<td>0.415</td>
<td>0.492</td>
</tr>
<tr>
<td>R20</td>
<td>0.567</td>
<td>0.573</td>
</tr>
<tr>
<td>R21</td>
<td>0.582</td>
<td>0.571</td>
</tr>
<tr>
<td>R22</td>
<td>0.683</td>
<td>0.674</td>
</tr>
<tr>
<td>R23</td>
<td>0.595</td>
<td>0.590</td>
</tr>
<tr>
<td>R24</td>
<td>0.717</td>
<td>0.705</td>
</tr>
<tr>
<td>R25</td>
<td>0.443</td>
<td>0.456</td>
</tr>
<tr>
<td>R26</td>
<td>0.339</td>
<td>0.328</td>
</tr>
<tr>
<td>R27</td>
<td>0.613</td>
<td>0.568</td>
</tr>
<tr>
<td>R28</td>
<td>0.411</td>
<td>0.401</td>
</tr>
<tr>
<td>R29</td>
<td>0.529</td>
<td>0.501</td>
</tr>
<tr>
<td>Mean</td>
<td>0.563</td>
<td>0.505</td>
</tr>
</tbody>
</table>
Firstly, we want to attract your attention to the differences between the PCI obtained for the international (inter-PCI) and intranational trade flows (intra-PCI). For the two years considered, the Inter-PCI is larger than the Intra-PCI.

This differentiation suggests that the products exported abroad are more complex than the ones traded domestically within Spain. This means that the products that require more capabilities to be produced are more likely to be exported to foreign markets. Or in other words, they are the ones that embodied larger amounts of knowledge (value added), and are able to cover the greater trade cost that an international delivery involves. This result is in line with the intuition about how competitive advantages are running international trade: singular products, less ubiquitous, are more likely to be exported abroad, both for being special local varieties (based on singular local factor endowments), or because they required very singular combination of capabilities and knowledge (i.e. tacit and express knowledge, in line with what was shown by Ballard & Rigby (2016) for the US). Moreover, it can be the case that firms producing very singular products, tend to concentrate their sales in foreign markets, while those producing more common items, will be more centred in the domestic market. The later intuition, points out to the expected result that, in general, provinces with larger shares of international exports, are the ones with higher economic complexity. Indeed, going one step further, and bringing here the implications of economic complexity over growth, one may expect that provinces with larger shares of international exports will be the ones expecting better future expectative in terms of growth.

Another conclusion is that complexity seems to be persistent in time. As commented before for the case of China, the ranking of complexity for the Spanish Provinces seems to be quite stable: independently of the methodology used, the PCI both for the international and for the intra-national trade, are very similar in the period considered. Given this stability, there are also some remarkable changes. For example, looking at the maximum values of the PCI, in year 1995, the most complex product was R24 (Transport materials), while in 2015 it was R22 (Non-electric machinery). These variations have their implications within the Hausmann and Hidalgo paradigm, and may draw important conclusions for policy making: if the Spanish provinces want to exploit their capabilities in order to gain the maximum growth rates in the future, they might gradually change their production structure towards the most complex products, whatever these might be.

For the maximum, there are no differences when considering international and intranational trade flows. But when considering the minimum values, important changes arise. In 1995 the minimum PCI for both types of trade were the same: R4-Wood. However, for 2015, the product with the lowest inter-PCI is R26-Cereals, while the one with the minimum PCI-intra is R26-Leather and footwear (methodology 1) or R4-Wood (methodology 2).

Next, we analyse the economic complexity index (ECI) for each province. Figure 5 show the results for the different ECI calculated following the two methodologies used for the intra-ECI. Panels I, II, V and VI show the results with ‘methodology 1’ and Panels III, IV, VII, VIII the ones calculated with the ‘methodology 2’. Panels I-IV shows the results for the ECI compared with the GDP of the next year and the Panels V-VIII the results against the provincial GDP lagged 10 years. This is done to check if the predictability of ECI’s help to predict better in the long-run than in the short-run.

The different panels show a positive relation of ECI and GDP per capita in all the scenarios. So, independently of what methodology is used, the results continue being positive. This is also true if using Intra-ECI instead of Inter-ECI. There are neither a big different when using the GDP per capita of 10 years later. The fitness continues being good, although the number of outliers increases when using the lagged GDP per capita. Moreover, removing this outlier, the relation is even better in the long-run than in the short-run. The only panel where the outliers have a smaller presence is when using the Intra-ECI of the first methodology (Panel VI).
Regarding the different methodologies, although the results are similar, the second methodology seems to obtain better results. For example, comparing Panel I with III and Panel V with VII, the relation is a slightly more positive when using ‘methodology I’. These seem to happen also when comparing the panels with Intra-ECI. All these facts are going to be checked later in the econometric analysis.
6. **Econometric analysis.**

Once that the main economic complexity indexes have been analysed in the previous section, it is time to focus on the role of these indexes as new variables explaining the economic growth of the Spanish provinces. For this purpose, the regression is going to be controlled with another two variables that are largely discussed in the literature: the *net capital* and the *human capital*.

With this aim, using the equation (22), a random effect panel data regression is estimated:

\[ \log(GDP_{pc})_{rt} = \beta_0 + \beta_1 ECI_{rt} + \beta_2 \log(Capital_{pc})_{rt} + \beta_3 \text{Human Capital}_{pcr} + u_{rt} \]  

(24)

Where ‘pc’ refers to the per capita variables, ‘r’ to the provinces and ‘t’ to the year.

The results are reported in Table 3. First Models 1-4 present the results with the GDP per capita of the following year (t+1) to analyse if those variables are good explanatory variables in the short-run. M1-2 show the results with the economic complexity indexes calculated with the *methodology 1*; while M3-4 the results for the indexes obtained using *methodology 2*. M1 and M3 show the results using the Inter-ECI, while M2 and M4 present the Intra-ECI cases.

Then, with the aim of analysing if the ECI work as a predictor in the long-run, we use an alternative dependent variable, which is now defined as the GDP per capita lagged ten years. The results for M5-M6 are the ones obtained for the ECIs obtained using the *methodology 1*; while the ones in M7-M8 corresponds to *methodology 2*. The inter-ECI is used for the M5 and M7 estimations, while M6 and M8 used the intra-ECI.

Regarding the results, and in contrast to the ones obtained for the other two variables, the coefficients are always positive, independently of the methodology or type of flow considered. This result points out to a clear positive effect of ECI in the determination of the lagged GDP per capita of the Spanish provinces. The significance of these coefficients is not as consistent as their signs. Looking first to the regressions when using the dependent variables “GDP per capita of the next year”, Inter-ECI is only significant in the case of the second methodology. In this scenario, the coefficient is also the largest compared with the coefficients of the other variables, so, with that results Inter-ECI seems to affect more positively the provincial GDP per capita of the following year than Net capital or Human capital. On the other hand, Intra-ECI is only significant with the estimation of the first methodology. In this scenario, it is important to note that although the Intra-ECI affects positively the GDP per capita, it seems to affect less positively than other variables like Net capital of the region.

In the case of the long-run estimation the results are even better. Inter-ECI is positive and significant independently of the methodology used. And in both cases the coefficient is related with higher GDP per capita in the long-run compared with the other variables. The results of the other variables are also remarkable, because these variables, in contrast to what happens with the ECI, obtained negative signs in all the scenarios estimated.

On the other hand, Intra-ECI is only significant in the estimation with the first methodology. But the results continue being positive, because of the positive sign of the coefficient, against the other variables. It is also remarkable that with the mentioned methodology, in the long-run, Intra-ECI obtained better results that Inter-ECI, with a larger positive coefficient.
Table 3: Panel data analysis of the economic complexity index.

<table>
<thead>
<tr>
<th>Methodology</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log GDP per capita (t+1)</td>
<td>Log GDP per capita (t+11)</td>
<td>Log GDP per capita (t+1)</td>
<td>Log GDP per capita (t+11)</td>
<td>Log GDP per capita (t+1)</td>
<td>Log GDP per capita (t+11)</td>
<td>Log GDP per capita (t+1)</td>
<td>Log GDP per capita (t+11)</td>
</tr>
<tr>
<td>Inter-ECI</td>
<td>0.0692 (0.122)</td>
<td>1.319*** (0.389)</td>
<td>0.180* (0.0979)</td>
<td>0.881*** (0.300)</td>
<td>0.213** (0.0882)</td>
<td>0.256 (0.206)</td>
<td>0.213** (0.0882)</td>
<td>0.256 (0.206)</td>
</tr>
<tr>
<td>Intra-ECI</td>
<td>0.344** (0.142)</td>
<td>0.173 (0.328)</td>
<td>-0.0957** (0.0427)</td>
<td>-0.0951** (0.0426)</td>
<td>-0.0752* (0.0426)</td>
<td>-0.0739* (0.0434)</td>
<td>-0.113*** (0.0390)</td>
<td>-0.101*** (0.0390)</td>
</tr>
<tr>
<td>Log Net capital per capita</td>
<td>1.268*** (0.0229)</td>
<td>1.278*** (0.0232)</td>
<td>1.262*** (0.0226)</td>
<td>1.262*** (0.0231)</td>
<td>-0.0929** (0.0387)</td>
<td>-0.0953** (0.0387)</td>
<td>-0.113*** (0.0389)</td>
<td>-0.101*** (0.0389)</td>
</tr>
<tr>
<td>Economic value of Human capital per capita</td>
<td>0.00775 (0.0368)</td>
<td>0.00831 (0.0367)</td>
<td>0.00982 (0.0366)</td>
<td>0.0167 (0.0373)</td>
<td>-0.0929** (0.0387)</td>
<td>-0.0953** (0.0387)</td>
<td>-0.113*** (0.0389)</td>
<td>-0.101*** (0.0389)</td>
</tr>
<tr>
<td>Constant</td>
<td>-13.04*** (0.398)</td>
<td>-13.36*** (0.411)</td>
<td>-12.98*** (0.378)</td>
<td>-12.92*** (0.395)</td>
<td>11.81*** (0.678)</td>
<td>11.79*** (0.676)</td>
<td>11.54*** (0.678)</td>
<td>11.52*** (0.693)</td>
</tr>
</tbody>
</table>

R-squared: within | 0.8169 (0.411) | 0.8185 (0.378) | 0.8179 (0.395) | 0.8166 (0.395) | 0.2351 (0.678) | 0.2496 (0.676) | 0.2463 (0.678) | 0.2446 (0.693) |

R-squared: between | 0.4681 (0.5017) | 0.4802 (0.5017) | 0.4689 (0.5017) | 0.4689 (0.5017) | 0.3468 (0.5017) | 0.4762 (0.5017) | 0.2340 (0.5017) | 0.5436 (0.5017) |

R-squared: overall | 0.5918 (0.6112) | 0.5955 (0.6112) | 0.5936 (0.6112) | 0.5936 (0.6112) | 0.1880 (0.6112) | 0.2698 (0.6112) | 0.1346 (0.6112) | 0.3250 (0.6112) |

Observations | 950 | 950 | 950 | 950 | 450 | 450 | 450 | 450 |

Number of prov | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
In conclusion, these results, although preliminary, reinforce the idea of the importance of economic complexity as an indicator of the income growth across provinces within a country, which has proven to perform in the same way than in cross-country analysis for the entire World. Moreover, in support to our main contribution of combining intra-national and international trade flows, our results confirm that the intra-ECI obtains similar results than the traditional inter-ECI when predicting GDP per capita in the short-run \((t+1)\), but has no rival when predicting the future GDP per capita lagged 10 years.

7. Conclusions.

The production and diffusion of ideas is becoming a new critical factor explaining growth differentials in a knowledge based economy. However, the information connecting the production and use of knowledge in the generation of goods and services is scarce and sometimes limited to singular cases studies.

Hausmann and Hidalgo created a useful framework to bridge the gap between the intuition of how countries account for different capabilities, and how they combine them in the production of a diverse range of products. Their approach relied mainly on a bimodal analysis of highly disaggregated international trade data, which allows computing levels of diversification of countries and ubiquity of the products exported. Thus, countries increase their level of complexity when they are able to produce a diversified range of products, which are produced by very few set of countries. The empirical application of this framework to the cross-country analysis has relied mainly on the availability of rich datasets on inter-national trade flows of goods, which raise the criticism about potential bias in the estimation of such complexity indexes, which have been computed without considering services or the share of the production channelled towards the national market (not exported abroad).

The concept of economic complexity explained for the case of countries, can also be extended to sub-national units within countries. This can permit elaborate a new set of indexes with the purpose of testing to what extent the link between complexity and growth across countries also holds for each one of the territorial units within their boundaries. The availability of new sub-national indexes can be critical when explain cross-regional differences in income per-capita, or when trying to define clever strategies for developing optimal sectoral structures of lagging regions, and reinforcing their resilience and capacity to succeed in a more global economy subject to strong structural transformations.

In contrast to this interest, very few attempts to extend the economic complexity framework to the sub-national entities within countries are found. Just three exceptions are revised here for China, Australia and the US.

In this paper, we develop the first attempt to compute and analyse the concept of economic complexity at the sub-national in Spain, considering the 50 Spanish provinces during the period 1995-2015. Our main contribution is on the combination of inter-national and intra-national trade of goods, and the consideration of the transport-mode used in the deliveries. To the best of our knowledge, our analysis is the first attempt to estimate economic complexity indexes at the sub-national level in Europe.

Our results reinforce the idea of the importance of economic complexity as an indicator of the income growth across provinces within a country, which has proven to perform in the same way than in cross-country analysis for the entire World. Moreover, in support to our main contribution of combining intra-national and international trade flows, our results confirm that the intra-ECI obtains similar results than the traditional inter-ECI when predicting GDP per capita in the short-run, but has no rival when predicting the future GDP per capita lagged 10 years.
8. References.


Annex 1: Product and regional codes.

Annex 1.1 Product references.

<table>
<thead>
<tr>
<th>Code</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Live animals</td>
</tr>
<tr>
<td>R2</td>
<td>Cereals</td>
</tr>
<tr>
<td>R3</td>
<td>Unprocessed food products</td>
</tr>
<tr>
<td>R4</td>
<td>Woods</td>
</tr>
<tr>
<td>R5</td>
<td>Processed food products</td>
</tr>
<tr>
<td>R6</td>
<td>Oils</td>
</tr>
<tr>
<td>R7</td>
<td>Tobacco</td>
</tr>
<tr>
<td>R8</td>
<td>Drinks</td>
</tr>
<tr>
<td>R9</td>
<td>Coals</td>
</tr>
<tr>
<td>R10</td>
<td>Minerals (not ECSC)</td>
</tr>
<tr>
<td>R11</td>
<td>Liquid fuels</td>
</tr>
<tr>
<td>R12</td>
<td>Minerals (ECSC)</td>
</tr>
<tr>
<td>R13</td>
<td>Steel products (ECSC)</td>
</tr>
<tr>
<td>R14</td>
<td>Steel products (not ECSC)</td>
</tr>
<tr>
<td>R15</td>
<td>Stones and ground, salt</td>
</tr>
<tr>
<td>R16</td>
<td>Cements and limestones</td>
</tr>
<tr>
<td>R17</td>
<td>Glass</td>
</tr>
<tr>
<td>R18</td>
<td>Construction materials</td>
</tr>
<tr>
<td>R19</td>
<td>Fertilizers</td>
</tr>
<tr>
<td>R20</td>
<td>Chemical products</td>
</tr>
<tr>
<td>R21</td>
<td>Plastics and rubber</td>
</tr>
<tr>
<td>R22</td>
<td>Machinery, non-electric engines and motors, tractors, agricultural machinery and equipment</td>
</tr>
<tr>
<td>R23</td>
<td>Machinery, apparatus and electric motors</td>
</tr>
<tr>
<td>R24</td>
<td>Transport material</td>
</tr>
<tr>
<td>R25</td>
<td>Textile and clothing</td>
</tr>
<tr>
<td>R26</td>
<td>Leather and footwear</td>
</tr>
<tr>
<td>R27</td>
<td>Paper</td>
</tr>
<tr>
<td>R28</td>
<td>Wood and cork</td>
</tr>
<tr>
<td>R29</td>
<td>Furniture and furnishings, new. Other manufactured articles.</td>
</tr>
</tbody>
</table>
Annex 1.2 Classification of the Spanish provinces by the INE

<table>
<thead>
<tr>
<th>Code</th>
<th>Province</th>
<th>Code</th>
<th>Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Álava</td>
<td>35</td>
<td>Las Palmas</td>
</tr>
<tr>
<td>02</td>
<td>Albacete</td>
<td>24</td>
<td>León</td>
</tr>
<tr>
<td>03</td>
<td>Alicante</td>
<td>25</td>
<td>Lérida</td>
</tr>
<tr>
<td>04</td>
<td>Almería</td>
<td>26</td>
<td>La Rioja</td>
</tr>
<tr>
<td>05</td>
<td>Ávila</td>
<td>27</td>
<td>Lugo</td>
</tr>
<tr>
<td>06</td>
<td>Badajoz</td>
<td>28</td>
<td>Madrid</td>
</tr>
<tr>
<td>07</td>
<td>Baleares</td>
<td>29</td>
<td>Málaga</td>
</tr>
<tr>
<td>08</td>
<td>Barcelona</td>
<td>30</td>
<td>Murcia</td>
</tr>
<tr>
<td>09</td>
<td>Burgos</td>
<td>31</td>
<td>Navarra</td>
</tr>
<tr>
<td>10</td>
<td>Cáceres</td>
<td>32</td>
<td>Orense</td>
</tr>
<tr>
<td>11</td>
<td>Cádiz</td>
<td>33</td>
<td>Palencia</td>
</tr>
<tr>
<td>12</td>
<td>Castellón</td>
<td>34</td>
<td>S. C. Tenerife</td>
</tr>
<tr>
<td>13</td>
<td>Ciudad Real</td>
<td>35</td>
<td>Pontevedra</td>
</tr>
<tr>
<td>14</td>
<td>Córdoba</td>
<td>36</td>
<td>Salamanca</td>
</tr>
<tr>
<td>15</td>
<td>Coruña (La)</td>
<td>37</td>
<td>Soria</td>
</tr>
<tr>
<td>16</td>
<td>Cuenca</td>
<td>38</td>
<td>Tarragona</td>
</tr>
<tr>
<td>17</td>
<td>Gerona</td>
<td>39</td>
<td>Teruel</td>
</tr>
<tr>
<td>18</td>
<td>Granada</td>
<td>40</td>
<td>Toledo</td>
</tr>
<tr>
<td>19</td>
<td>Guadalajara</td>
<td>41</td>
<td>Valencia</td>
</tr>
<tr>
<td>20</td>
<td>Guipúzcoa</td>
<td>42</td>
<td>Valladolid</td>
</tr>
<tr>
<td>21</td>
<td>Huelva</td>
<td>43</td>
<td>Vizcaya</td>
</tr>
<tr>
<td>22</td>
<td>Huesca</td>
<td>44</td>
<td>Zamora</td>
</tr>
<tr>
<td>23</td>
<td>Jaén</td>
<td>45</td>
<td></td>
</tr>
</tbody>
</table>
