Measuring Services Complexity: A Novel Machine Learning Approach Using U.S. Input–Output Data

dECON
Facultad de Ciencias Sociales

dECON Facultad de Ciencias Sociales
UNIVERSIDAD DE LA REPÚBLICA

Ciencias Sociales Universidad de la República
URUGUAY

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Documentos de Trabajo

Titulo

Measuring Services Complexity:
A Novel Machine Learning Approach Using U.S. Input–Output Data

Santiago Picasso

Documento No. 01/26
Febrero 2026

ISSN 0797-7484

Measuring Services Complexity:
A Novel Machine Learning Approach Using U.S. Input–Output Data*

Santiago Picasso †

19th February 2026

Abstract

A stylized fact in modern economies is that the more developed a country is, the greater the weight of the service sector. The economics of complexity has provided a new perspective that explains this growth in modern economies. However, the study of economic complexity through the standard measure of the complexity index presents an increasingly relevant omission in understanding the economic process and its growth. In general, the data used to measure the Economic Complexity Index (ECI) are based on information about goods; however, there is a lack of information on services. This paper proposes a new methodology to retrieve information on the economic complexity in services. For this purpose, the US input-output matrix is used. This work is novel because, thanks to the structure of the data as a network, it is possible to infer the missing information on the complexity of services. Using a machine learning method, it is possible to impute the complexity index for 146 services, a level of disaggregation, that is strikingly higher than in other works. The index recovered by this method is consistent with previous results that found service sectors to be more complex than goods. The second result shows that the more restricted the core is in the center of the network, the greater the centrality of services and their complexity. Finally, the results confirm the relevance of the economic complexity index. However, the ECI for services is better than the ECI for goods for predicting growth; a one-unit increase in the ECI of services increases GDP growth by more than 1 percentage point.

Keywords: Economic Complexity; Services Sector; Input–Output Networks; Machine Learning; k-Nearest Neighbors; Structural Transformation; Economic Growth; Spatial Econometrics.
JEL Classification: C45; C55; O11; O14; O47; L80.

*A Spanish version of this paper is available from the author upon request via e-mail. This is a preprint version of one chapter of my thesis. I thank my supervisors, Prof. Carlos Bianchi and Prof. Guzmán Ourens, for their valuable guidance and support. I also thank Decon-FCS, FCEA, the development group of research in FCEA, and colleagues for their valuable comments. Financial support from CSIC, through the CAP scholarship is gratefully acknowledged.
Una versión en español de este artículo está disponible previa solicitud al autor por correo electrónico.

†E-mail:santiago.picasso@fcea.edu.uy // santiago.picasso@cienciassociales.edu.uy

Resumen

Un hecho estilizado en las economías modernas es que cuanto más desarrollado está un país, mayor es el peso del sector servicios. La economía de la complejidad ha proporcionado una nueva perspectiva que explica este crecimiento en las economías modernas. Sin embargo, el estudio de la complejidad económica a través de la medida estándar del índice de complejidad presenta una omisión cada vez más relevante para comprender el proceso económico y su crecimiento. En general, los datos utilizados para medir el Índice de Complejidad Económica (ICE) se basan en información sobre bienes; sin embargo, hay una falta de información sobre los servicios. Este artículo propone una nueva metodología para recuperar información sobre la complejidad económica en los servicios. Para ello, se utiliza la matriz de insumo-producto de Estados Unidos. Este trabajo es novedoso porque, gracias a la estructura de los datos como red, es posible inferir la información que falta sobre la complejidad de los servicios. Mediante un método de aprendizaje automático, es posible imputar el índice de complejidad para 146 servicios, un nivel de desagregación notablemente superior al de otros trabajos. El índice recuperado por este método es coherente con resultados anteriores que consideraban que los sectores de servicios eran más complejos que los de bienes. El segundo resultado muestra que cuanto más restringido es el núcleo en el centro de la red, mayor es la centralidad de los servicios y su complejidad. Por último, los resultados confirman la relevancia del índice de complejidad económica. Sin embargo, el ECI para los servicios es mejor que el ECI para los bienes a la hora de predecir el crecimiento; un aumento de una unidad en el ECI de los servicios aumenta el crecimiento del PIB en más de un punto porcentual.

Palabras clave: Complejidad Económica; Sector Servicios; Redes Insumo–Producto; Aprendizaje Automático; k-Vecinos Más Cercanos; Transformación Estructural; Crecimiento Económico; Econometría Espacial.

Clasificación JEL: C45; C55; O11; O14; O47; L80.

1 Motivation

Economic complexity is a concept that refers to the amount of knowledge embedded in the productive structure of an economy (Hidalgo & Hausmann, 2009). The concept was operationalised with the complexity index of (Hidalgo & Hausmann, 2009). This index approximates complexity through the concept of the revealed comparative advantage of goods and their interaction with other goods in the basket of exportable goods. Thus, a good that is rarely traded on a global scale is unique; however, if it is also demanded by many countries, it indicates a relatively innovative good that is difficult to replicate globally, signifying a high level of complexity.

The reference to goods effectively refers to those tangible products, typically documented by the COMTRADE database, that pass through customs and are registered. No services are registered in this database. As is well known, records for services have not been measured well historically. The information is scarce and disjointed (Li et al., 2025) for different reasons, such as intangibility and non-storability, which impede the compilation of precise international trade in services data (Mishra et al., 2020). Recently, efforts have been made to incorporate them into the analysis as they are a crucial part of the economic growth of most of the world's economies. Leaving services out of the complexity index is increasingly problematic in measurements as a proxy for economic structure.

An additional problem with the complexity methodology that could lead to criticism is that the information revealed by this index reflects only international relations. As shown in the table for the United States, trade is 9% for goods and 3% for services.

The methodology of complexity when using internationally traded products cannot reflect the complexity of services because most of them are not traded globally (see table 1). The services reflect only 20-30% of the value traded globally (Ortiz-Ospina et al., 2018). The question, then, is how to measure their complexity if they cannot be captured in the transaction matrix using the complexity methodology.

Table 1: Exports of goods and services in 2017 in US.

type	export ratio	products that are export	ratio products that are export
goods	9%	227	56%
services	3%	44	11%

Source: Own elaboration based on matrix of use of United State´s economy extracted from BEA

The aim of this paper is to propose a new methodology to retrieve information on economic complexity in services across an extended number of sectors. This index will be analyzed using a network approach and employed to assess the economic complexity of different economies (U.S. Federal Units) and its effect on growth. Based on the relevance of services due to the boom of the last few decades related to their importance in the economy and the value that they generate (Herrendorf et al., 2013; Stojkoski et al., 2016), the hypotheses are proposed. Services are, on average, more complex than goods, are central to an economy's trade network, and therefore, their complexity is an important predictor of an economy's growth.

To this end, this research incorporates a nonparametric machine learning methodology that allows us to infer the complexity of services using exogenous information regarding the process of generating the data used to calculate complexity, typically exportable flows between countries. The external source arises from the US input-output matrix. This method allows us to measure the complexity of goods and services and observe the differences in complexity between them. Additionally, we construct the network of the USA economy for three periods (2007, 2012, and 2017). The results show consistency; services are more central in the network of transactions for all years and are also more complex. Finally, we use a growth econometric approach to show that the variation in the complexity of services allows us to predict the economic growth of each state.

The article continues in section 2 with a review of the literature on the concept of complexity (section 2.1, previous evidence of the link with growth using goods (section 2.2) and services (section 2.3), and with this background, postulates the hypotheses of the work. Section 3 explains the ECI and PCI (section 3.1), the construction of the PCI for services (section 3.2 and section 3.3), the machine learning techniques used to impute it (section 3.3) and the data use for the analysis (section 3.4). After explaining the estimation strategies, the article moves on to section (4), where the hypotheses are confirmed, and finally concludes in section (5).

2 Review of literature

2.1 Economic complexity concept

The concept of economic complexity refers to the amount of specialized knowledge and its combinations required to produce a good (Balland et al., 2022). To produce that product, a large number of inputs are needed, not only tradable but also non-tradable. These non-tradable inputs, which include specific infrastructure, regulations, norms, and productive knowledge and skills, are the capabilities (Hausmann & Hidalgo, 2010). The capabilities required to produce a given product emerge when a state has revealed comparative advantages in producing it. An economy is more complex the greater the amount of such capabilities that are effectively used in production.

The core idea behind the concept of economic complexity is that revealed comparative advantages conceal the skills (a) that a territory (c), in this case states, possesses to produce. However, this is not sufficient, since producing a good (p) requires specific skills. Therefore, production in a territory with revealed comparative advantages arises from having the available skills that are also necessary to produce the good. This is summarized by Balland et al. (2022) in the following equation.

$$M_{cp} = C_{ca} \odot P_{pa} \tag{1}$$

Where M_{cp} is the country/state/city–product matrix that indicates the revealed comparative advantages of a state (c) in producing a product (p). These comparative advantages are explained by the state's capabilities C_{ca}, which must be aligned with the capabilities required by a product P_{pa}.

The degree of complexity of a state follows from observing its production diversity relative to others—that is, not only how many products are produced, but also how exclusive those products are. A state may produce many goods that are produced by everyone else, or it may produce exclusive goods that few or no other territories produce. It is the combination of both conditions—high diversity and high exclusivity—that makes a territory more complex.

One aspect to consider at this point is that developed capabilities, unlike the skills approach, have not only an individual aspect but also a relational aspect in their link with others (sectors, capabilities). These capabilities have a combinatorial property (Hausmann & Hidalgo, 2010; Hidalgo & Stojkoski, 2025), which makes them unique. For example, if a certain skill is required to produce a product, the appearance of that product will not be achieved by simply incorporating that skill; rather, it will depend on the exact combination of capabilities that makes it possible to produce that product. This is a crucial difference from this approach.

Under the previous idea, two indicators were developed to measure complexity: the Economic Complexity Index (ECI) and the Product Complexity Index (PCI) (Hidalgo, 2021). These indices arise from a dimensionality reduction of the matrix M_{cp} and, as shown by Mealy et al. (2019), can be interpreted as the outcome of a spectral clustering algorithm that partitions the network into two. For this reason, these indicators (measured by the eigenvector) are referred to as a method for grouping countries and products according to their capabilities and skills (Balland et al., 2022). The ECI and PCI indicators rank countries according to their degree of territorial complexity: the higher the ECI, the greater the complexity. At the same time, PCI values are obtained for products, which also rank their complexity, with higher values indicating greater complexity.

These measures gained relevance because they reduce the dimensionality of the information contained in the matrix M_{cp}, offer a straightforward interpretation, and display strong correlations with economic and social indicators. Accordingly, economic complexity appears to be associated with a broad set of economic and social outcomes such as inequality and economic growth (Balland et al., 2022).

2.2 Economic complexity of goods

Focusing on the core of this research, there is extensive literature showing a positive relationship between economic complexity and economic growth. In general, a variety of works have shown how economic complexity is a robust predictor of economic growth (Hidalgo, 2021; Ourens et al., 2013). This relationship has been shown to be robust to different controls. (Hausmann et al., 2014). Gao and Zhou (2018) finds a relationship between complexity and the level of production. Antonelli et al. (2022) finds a positive relationship with innovation and productivity, which is indirectly associated with higher value added per unit of employment or output.

This finding is consistent across other historical periods. For the period 1850–1900, estimates of complexity show a positive and statistically significant relationship with long-run GDP (Domini, 2022). Koch (2021) proposes a new measure of complexity using the value added of exports, again finding a positive relationship. A one-unit increase in the ECI implies a positive growth variation of between 9% and 12%.

Stojkoski and Kocarev (2017) find a positive relationship for Southeastern and Central Europe, although they show that this relationship materializes in the long run rather than in the short run. In this case, the long run corresponds to the period 1995–2013, during which a one-unit increase in the ECI (one standard deviation) raises GDP by 45%. This result is consistent with Ourens et al. (2013), who find that the method of reflections predicts growth, but not in the short run.

This result is also observed at the regional level within countries. Poncet and De Waldemar (2013) documents this relationship for Chinese cities. Similarly, in Brazil, Teixeira et al. (2022) shows a positive relationship between complexity and growth, although it is not always statistically significant. However, another study on Brazil shows that what truly matters for explaining growth is local industry and services rather than export complexity (Cardoso et al., 2023). Other subnational studies confirm this relationship: Pérez-Balsalobre et al. (2019) show that for Spain, over the period 1995–2016, regional GDP increases by between 30% and 40% following a one-unit increase in the ECI. Mao and An (2021) show that GDP per capita would increase by around 30% when the ECI increases by one unit in middle- and high-income economies from 1995 to 2010. Beyond this review, the work of Hidalgo (2021) found evidence that economic complexity accounts for a growth effect of approximately 0.7% and 3-4%.

All these studies corroborate the idea that increasing the complexity of an economy leads to economic growth. This empirical evidence corroborates the relationship, but one might ask what causes greater complexity to generate greater added value. Although there are few developments linking the indicator to its theoretical basis, Hidalgo and Stojkoski (2025) develops a theoretical model that attempts to address this gap. They show that the prices of goods in this model follow a concave function with respect to their capacity requirements, which aligns with the argument that complex goods in a territory generate more value. It would appear that proven capabilities in producing goods are allocated to those that generate the most value.

2.3 Economic complexity of services

Complexity measures have traditionally concentrated on goods; however, recent papers emphasizes the need to incorporate services to provide a more comprehensive assessment of economic complexity (Mishra et al., 2020; Zaccaria et al., 2018). Following the framework of Hidalgo and Hausmann (2009), societies have expanded their knowledge through specialization and ask themselves how to combine the different parts of knowledge. Under this idea, specialization implies not only goods but also services that are combined to generate added value. If these services are omitted, a part of the possible combinations of current and new knowledge is omitted. The underlying intuition is a traditional concept developed by the Lewis mechanism of substitution from traditional to modern sectors and readapted in recent contributions (e.g. Hidalgo and Hausmann, 2009) provide a modern network-based or complexity-oriented reinterpretation.

In the work of Hidalgo and Hausmann (2009), one of the limitations is the absence of services data in the analysis. However, they performed an exercise including data on services from the International Monetary Fund and found that the results of complexity as predictor of economic growth are positive and almost the same regardless of whether the complexity index included services or not. Moreover, the correlation between the complexity of only services and economic growth is close to zero and is not significant.

This result is used to support the argument that consider only goods is a good enough approach to the whole economic dynamics. The article uses a clearly weak measure and the authors explicitly comment on it as a debility because the inclusion of services has a much lower level of disaggregation. In this vein, the authors argument that ECI with services does not add much information. However, if this is true, in light of new data and methods, the view that the complexity of services does not change the result begins to change. For example, the extension of complexity to include service exports has shown that sophisticated services often align with advanced manufacturing sectors, significantly influencing a nation's competitiveness (Stojkoski et al., 2016; Zaccaria et al., 2018).

Regarding the effort to measure complexity in services, Lo Turco and Maggioni (2022) do so, but through workers' skills. They find for the United States a stronger effect on economic growth in services than in manufacturing. Mishra et al. (2020) find that the inclusion of the services increases economic fitness (algorithm similar to Hidalgo and Hausmann (2009) but in a non-linear version) of an economy, and its relationship with GDP.

Although arguments were presented to justify that complex goods add more value, services have specific channels compared to goods through which they generate value and stimulate growth. Services linked to the development of new technologies, as well as logistical and technostructural support for the services of other sectors, determine that these are decisive for the generation of added value. In light of the above empirical evidence and theoretical arguments, it is expected that the complexity of services is a determinant of economic growth.

In this sense, Stojkoski et al. (2016) shows that not only goods are relevant for predicting the economic growth of countries. The complexity index for services is greater than that for goods, but these results have a composition problem. The index has only 10 categories for goods and 12 categories for services. Patelli et al. (2022) improves this analysis by generating an integrated database of services and goods for 160 countries and 124 sectors, of which 27 are services. This work uses an algorithm to impute the missing services named as K-Nearest Neighbors (k-NN) method (see section 3.3 for more details).

There are many machine learning techniques that allow for learning or inferring parameters, patterns, and predictions from a large and heterogeneous set of data (Hastie et al., 2009). The techniques that reduce the dimensionality of the information are agnostic about which individual factors shape growth and distribution, but they use the information without further assumptions (Hidalgo, 2021).

This emerging literature has done well to recognize the weaknesses of the original complexity methods. However, they have not been able to overcome a fundamental weakness of the method, which is that international trade is mostly in goods but also in services that, being inherently non-tradable, present limited measurement opportunities. In this respect, we propose to contribute to the measurement of the complexity of the services sector, taking into account their non-tradable characteristics.

In general, the literature review shows us an extensive body of work that measures the complexity of goods, but not that of services. Therefore, one of the objectives of this work is to measure this complexity and test it within the framework of growth regressions. Under this framework, the general hypothesis postulates that the complexity of services should be central to the economic structure and should explain economic growth.

Hypothesis 1: Economic growth is directly explained by the complexity of services. $\beta_s > 0$.

Although the studies reviewed have not focused on the network structure of the economy, the importance and weight of services lead to a natural hypothesis associated with the previous hypothesis. If the service sector is relevant as a motor for the dynamic of the economy, it is expected that the center of the economy will be characterized by services. As the services, on mean, are more complex (Stojkoski et al., 2016), the core of the network should be characterized by the complexity of services. Taking advantage of the availability of network data through the I-O matrix, work will be done to explore this hypothesis.

Hypothesis 2: The core of the network transactions in the I-O matrix is more complex than that in the whole economy.

3 Methodology

The methodology for retrieving the complexity service values consists of three steps. Firstly, a database with a high level of disaggregation of services is chosen, with good quality information, an aspect that is not evident in the case of services, especially at the level of international trade. After obtaining this information, the fundamental task is to carry out a correspondence algorithm between sectors of the US census Input-Output matrix and the sectors of foreign trade, which are used to generate the economic complexity index for goods (see 3.2). Then, the third step is to estimate the complexity index for services using a non-parametric supervised learning classifier algorithm. The idea is to use the Input-Output matrix as exogenous information to learn or infer the complexity index for services (PCI_S) using the complexity index for goods (PCI_G)

3.1 Complexity index

Let us consider the following example, which summarizes the idea of using the matrix M_{cp} to derive complexity:

Example 1 (Comparative advantages in products (p) by country (c))**.** *We consider a simple setting with 2 countries, 3 products, and 4 capabilities.*

1. Capability endowment matrix of countries

$$C_{ca} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Country 1 has capabilities $\{1,2,4\}$*; Country 2 has capabilities* $\{1,3\}$*.*

2. Capability requirements of products

$$P_{pa} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Product A requires $\{1,2\}$*; Product B requires* $\{1,3,4\}$*; Product C requires* $\{2,3\}$*.*

3. Application of the operator ⊙ (Leontief rule)

We define:

$$M_{cp} = C_{ca} \odot P_{pa}$$

where

$$M_{cp} = \begin{cases} 1, & \text{if } P_{p,*} \subseteq C_{c,*}, \\ 0, & \text{otherwise.} \end{cases}$$

When applying the logical operator, only good A will be produced in country 1:

$$\begin{aligned} \text{Country 1 produces A} \quad & (\{1,2\} \subseteq \{1,2,4\}) \\ \text{Country 2 does not produce A} \quad & (\{2\} \notin \{1,3\}) \\ \text{Country 1 does not produce B} \quad & (3 \notin \{1,2,4\}) \\ \text{Country 2 does not produce B} \quad & (4 \notin \{1,3\}) \\ \text{Country 1 does not produce C} \quad & (3 \notin \{1,2,4\}) \\ \text{Country 2 does not produce C} \quad & (2 \notin \{1,3\}) \end{aligned}$$

4. Resulting country–product matrix

$$M_{cp} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Country 1 produces only Product A; Country 2 is unable to produce any of the three products because it does not have the indicated capabilities. This example clearly shows how the capabilities in a territory are a necessary condition to be more complex but are not sufficient because the knowledge has to be applicable in the production of a good.

Therefore, how can we represent this concept in a way that the M_{cp} reveals it properly? There was a method from the complexity approach developed or adapted by Hidalgo and Hausmann (2009). The authors create new indicators that reveal the concept in a measurable way. The complexity of a country depends on the complexity of its goods and vise versa Hidalgo (2021).

$$K_c = f\left(M_{cp}, g\left(M_{cp}, K_c\right)\right)$$
$$K_p = g\left(M_{cp}, f\left(M_{cp}, K_p\right)\right) \tag{2}$$

In the usual case, the system of equations is defined by simple averages:

$$K_c = \frac{1}{M_c}\sum_p M_{cp}K_p$$
$$K_p = \frac{1}{M_p}\sum_c M_{cp}K_c \tag{3}$$

Given the linearity of the system, it can be rewritten as:

$$K_c = \widetilde{M}_{cc}, K_c$$
$$K_p = \widetilde{M}_{pp}, K_p, \tag{4}$$

Being $\widetilde{M}_{cc}$ y $\widetilde{M}_{pp'}$:

$$\widetilde{M}_{cc} = \sum_p \frac{M_{cp}M_{cp}}{M_cM_p}$$

and:

$$\widetilde{M}_{pp'} = \sum_c \frac{M_{cp}M_{cp}}{M_cM_p} \tag{5}$$

The second eigenvectors of each matrix are the complexity indices k_c (ECI) and k_p (PCI). In economic terms, the ECI is the best way to divide economic units into groups according to the activities that are present in them.

$$ECI = \frac{K_c - mean(K_c)}{\sigma\left(K_c\right)} \tag{6}$$

The Product Complexity Index (PCI) ranks the diversity and sophistication of the production skills required to manufacture a product.
Products with a high PCI value (the more complex products that only a few countries can manufacture) include electronics and chemicals (Hausmann et al., 2014).

$$\text{PCI}_G = \frac{K_p - \text{mean}\left(K_p\right)}{\text{stdev}\left(K_p\right)} \tag{7}$$

In this paper the PCI is composed of the complexity of goods and services, i.e. the PCI is a function of the complexities of both types of products, $PCI = f(PCI_S, PCI_G)$.

It is useful to note here what the system of equation (3) tells us about the complexity of products and territories. The ECI is a weighted average of the PCI and vice versa. This means that once we achieve the complexity of all products we can go to see the complexity by country (or state).

3.2 Biunivocal correspondence between NAICS and HS 1992

The United States (US) I-O matrix developed by the Bureau of Economic Analysis (BEA) is classified according to the Industry Country Code (ICC). On the other hand, the complexity of goods is coded according to the Harmonised System (HS). Meanwhile, US goods are classified according to North American Industry Classification System (NAICS) codes. Therefore, a correspondence must be made between goods according to HS and NAICS, and once this equivalence has been established, the complexity of goods according to the NAICS classification must be imputed to the goods in the I-O matrix that are classified according to ICC.

1. Using the database of Atlas "Product Complexity Rankings 1995 - 2021" we get all the product complexity index for the period 1995-2021 with HS 1992 codes labels.

2. To merge the complexity information to the Input-Output database we need an equivalence between HS 1992 and NAICS. For that reason we use the information of Pierce and Schott (2009) available at https://sompks4.github.io/sub_data.html

3. Once we get the correspondence between HS 1992 and NAICS we have to generate a correspondence between NAICS and BEA Industry Code.

3.3 How to get the complexity information for services?

A set of information could have missing values of different types: completely at random, missing at random and not missing at random (Batista & Monard, 2002). The first has not risk of bias when the data is imputed. However, in the other cases, biases are likely to exist. The usual options of ignoring missing cases or imputing an average value to them are not appropriate because they ignore the correlated structure of the data (Troyanskaya et al., 2001). Here the bias is clearly not random, but corresponds to services that are generally non-tradable, so ignoring it presents biases.

The identifying assumption for measuring the complexity of services is that of homophily. This assumption establishes that the more similar the agents are, the greater the probability that they will be related. It is in this sense that a machine learning methodology is established to retrieve averages in an environment of the sector we want to measure, its neighbors will be a good proxy of its complexity. This assumption is not unreasonable in the network literature and is, in fact, commonly observed in many systems. Beyond this generality, evidence from the complexity literature also confirms this homophily relationship between sectors. The core of the product space, its most densely connected part, is composed of highly complex activities (Hidalgo et al., 2007).

3.3.1 The network input-output

The structure of the I-O matrix will be analysed descriptively using classic network techniques. Each sector will be a node in a network, and each dollar transacted will determine the weight of the relationship between two sectors. In this way, the network will be constructed for each period. The complete network is one in which there are many links per node, where graphing it does not provide much information, but the central core of the network will be shown. This shows the main links in the economy, where the most significant transactions occur, the largest dollar transactions. The criterion for dissecting the network is to look at the 95th percentile, i.e. the 5% of the heaviest links, and then expand. When we have 100% of the links, it means that we are at the 0% percentile, or we are considering the entire network.

To describe the network, we will consider the following variables:

Table 2: Basic structural indicators of a network

Indicator	Symbol	Definition / Formula
Number of nodes	N	Total number of industries in the network.
Number of edges	L	Total number of transactions between industries.
Network density	D	$D = \frac{2L}{N(N-1)}$
Average Degree	$\bar{deg}$	$\bar{deg} = \frac{1}{N}\sum_{i=1}^{N} deg_i = \frac{2L}{N}$
Distance	$d_{A,B}$	The shortest path algorithm is applied (West, 1996)

3.3.2 Weighted K-nearest neighbors (KNNimpute)

The K nearest neighbours method is used to impute missing data. This method uses network topology to infer missing data for some attributes. This method is a well-known technique that has been used for many decades. It has been developed and successfully implemented in other disciplines such as in the study of genetics (Troyanskaya et al., 2001).

While there are several methods, such as Singular Value Decomposition (SVD) or row average. The K-NN method has been one of the most popular because of its simplicity and robustness (Troyanskaya et al., 2001). Basically, what the method does is obtain the information for the node in the network that has missing information— in our case, the service sector i— and infer its value from the average of the values of the attribute with missing data from its nearest neighbours.

Homophily is the underlying assumption of this method, i.e., that nearby nodes will be similar to each other. In our case, the idea is to assume that a sector that has many transactions with another sector within the input-output matrix has a certain complementarity or relationship that leads us to assume that their strong interactions presuppose a similar relatedness and therefore a similar level of complexity.

To see how the algorithm works, consider the following example. Given a matrix $I - O$ of size 4, reflecting all transactions between the 4 hypothetical sectors of the example:

$$M = \begin{bmatrix} & A & B & C & D \\ A & 3902232 & 6894 & 45637 & 23002 \\ B & 9708 & 2585227 & 35701 & 107629 \\ C & 18493 & 28516 & 7546586 & 483020 \\ D & 208459 & 3022 & 167719 & 18600637 \end{bmatrix}$$

In turn, there is a vector (PCI) of the Product complexity index attribute associated with each node.

$$PCI = \begin{bmatrix} A & 3.5 \\ B & IPC_B \\ C & 2.7 \\ D & IPC_D \end{bmatrix}$$

Note that there are nodes with missing information, labelled as IPC_B and IPC_D. These nodes correspond in our case to the services that have no information in the complexity vector of Hidalgo and Hausmann (2009).

The idea of the algorithm is to take advantage of the information in the network M to learn or impute the values of the missing data subvector.

The procedure is as follows:

- Given a matrix $I-O$, we obtain the distance matrix between nodes; let's call it $I-O_d$. The distance is calculated using the function "distances" from R software, which "calculates the length of all shortest paths from or to the vertices of the network. The function"shortest paths()" in R calculates one shortest path (the path itself, and not just its length) from or to the given vertex." The distance matrix for the example matrix M is as follows:

$$I-O_d = \begin{bmatrix} 0.0000 & 0.1409 & 0.0687 & 0.0480 \\ 0.1409 & 0.0000 & 0.1136 & 0.0929 \\ 0.0687 & 0.1136 & 0.0000 & 0.0207 \\ 0.0480 & 0.0929 & 0.0207 & 0.0000 \end{bmatrix}$$

- Once $I-O_d$ is obtained, the information about the k neighbouring nodes of node i is used. Where i is the subset of nodes for which there is no information. In this case, the nodes are 2 and 4.

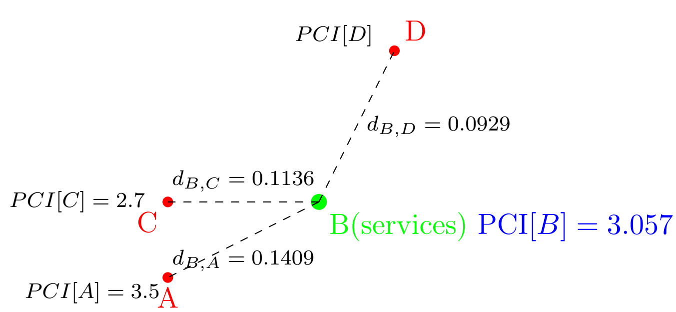
- To node i, the imputation is applied, weighted by the weights matrix (W) under the following formula:

$$\mathrm{W}[i,k] = \frac{1}{I-O_d[i,k]}$$

$$\mathrm{PCI}[i] = \frac{\sum_k PCI[i,k]\cdot W[i,k]}{\sum_k W[i,k]} \qquad \text{(where k is the number of near neighbors)}$$

  In this example, the value of node 2 would be:

$$\mathrm{PCI}[B] = \frac{\left(\frac{1}{0.1409}\cdot 3.5\right)+\left(\frac{1}{0.1136}\cdot 2.7\right)}{\frac{1}{0.1409}+\frac{1}{0.1136}} = \frac{48.6111}{15.9014} \approx \boxed{3.057} \text{ in this case k=2}$$



Where d_{ij} is the distance defined as the length of the shortest path between vertices i and j.

Although some works are conclusive on the advantages of the method (Troyanskaya et al., 2001), the robustness of the method is explored for this particular case. For this purpose, the predictive capability of the method is simulated with a training sample. We consider a subgraph composed only of the nodes for which there is known information about the PCI_G attribute. For complexity attributes, it is considered that a partition of nodes presents missing information in 42% of the cases. This ratio is chosen because it reflects the amount of missing data for the complexity attributes in the US I-O matrix. There are 235 nodes, 99 of which have their complexity values replaced by missing data. In this way, we proceed to impute this value using the k-NN method. As we know the true values of the missing entries, we can calculate the R^2 for this subsample. We hope that the R^2 will be high; this would mean that the method is a good approximation for the missing values. The higher the R^2, the better the method works.

3.3.3 Growth equation: relation between ECI and GDP

Traditional growth equations have well-known problems (Durlauf, 2009; Temple, 2021) that are assumed to be limitations of the work. Previous works, such as the seminal work in this literature (Hausmann et al., 2014), discuss the validity of the predictive power of complexity in growth. Many studies have already been published that find a relationship, but many of them, including the original ones (Hidalgo, 2021), use growth equations without addressing the specification problems to identify the effect between complexity and growth. The identification assumption must satisfy exogeneity:

$$\mathbb{E}\left[\varepsilon_{i,t} \mid ECI^{S}_{i,t-1}, ECI^{G}_{i,t-1}, \ln GDP_{i,t-2}, \mathbf{X}_{i,t-1}, \gamma_i, \tau_t\right] = 0. \tag{8}$$

In the growth equations that we will estimate, simultaneity will be partially corrected, but biases related to state size and omitted variables remain. Some studies explore ways to deal with these endogeneities. Bartelme et al. (2024) use leave-one-out shift–share instruments and apply the Post–Double Selection (PDS) method of Belloni et al. (2014), together with LASSO, to select only the most relevant control variables—those that predict both the dependent variable and the shocks. With this reduced subset, they estimate OLS consistently and with valid inference. While this study does not aim to depart substantially from the economic complexity literature by pursuing causal effects through parameter instrumentation or natural experiments, it does employ several variations in specification design to strengthen the robustness of the results.

First, growth equations with fixed effects will be used, exploiting the panel structure in order to control for unobserved state-specific characteristics. Given the limited number of observations, a differenced-equations approach is used to identify the parameter associated with the ECI, which reduces the number of controls and increases test power. Finally, the analysis addresses an issue that is usually overlooked in growth equations: endogeneity arising from spillover effects across individuals/states. That is, since the units under study influence one another, it is necessary to consider spatial econometrics to deal with this source of bias. This method is called the Spatial Durbin Model, which uses a W matrix to control the distance between states.

3.4 Data

In this work, we use the input-output matrix of the United States economy from the Bureau of Economic Analysis of the US Department of Commerce. This matrix reflects all economic interactions between industries. This information can be presented in different ways. On the one hand, we have the Make table, which tells us the value of goods and services available in the US economy produced by domestic industries. Then, the Supply matrix also includes the imported value of goods and services. On the other hand, it has the Use Tables that show how the supply of goods and services is utilised by industries, households, government, and the rest of the world. Finally, the Requirements matrix shows the direct and indirect connections between industries.

This paper explores these different matrices as sources of information to impute the complexity of services. At the end of the day, the flow matrix chosen will be the one that best predicts the imputed values; in this case, the Supply matrix will be chosen.

The rows show the industries (see table 1), and the columns show the commodities that the same industries produce. Looking in a row, all the commodities produced by that industry are identified, and the sum of the entries is that Commodity (Horowitz & Planting, 2009). But this table has a difference; it presents the total domestic supply from both local and international producers of goods and services available for use in the national economy (Young et al., 2014).

		Industries																Valuation adjustment										
		Agriculture, forestry, fishing, and hunting	Mining	Utilities	Construction	Manufacturing	Wholesale trade	Retail trade	Transportation and warehousing	Information	Finance, insurance, real estate, rental, and leasing	Professional and business services	Educational services, health care, and social assistance	Arts, entertainment, recreation, accommodation, and	Other services, except government	Government	Total Commodity Output. (Basic prices)	Imports	CIF/FOB Adjustments on imports	Total product supply (basic prices)	Trade margins	Transportation costs	Total trade and transportation margins	Import duties	Tax on products	Subsidies	Total tax less subsidies on products	Total product supply (purchaser prices)
Commodities	Agriculture, forestry, fishing, and hunting																											
	Mining																											
	Utilities																											
	Construction																											
	Manufacturing																											
	Wholesale trade																											
	Retail trade																											
	Transportation and warehousing																											
	Information																											
	Finance, insurance, real estate, rental, and																											
	Professional and business services																											
	Educational services, health care, and social assistance																											
	Arts, entertainment, recreation, accommodation, and food services																											
	Other services, except government																											
	Government																											
	Other																											
	Scrap, used and secondhand goods																											
	Total Intermediate																											
	Compensation of employees																											
	Taxes on production and imports, less																											
	Gross operating surplus																											
	Total value added																											
	TOTAL SUPPLY (basic prices)																											

Figure 1: Supply table example for aggregate sectors. Own elaboration based on Horowitz and Planting (2009) and Young et al. (2014).

On the other hand, the industry-by-industry total requirements matrix details an overview of the entire supply chain, with direct and indirect connections between industries. For example, a bakery depends directly on a mill for flour and indirectly on a farmer who supplies wheat to the mill. This takes into account the final demand, which, when multiplied by the total needs matrix, gives the total requirements matrix.

Because of the level of aggregation, we consider that industries and commodities are the same. So, from now on, columns and rows refer to products (goods and services). It measures all economic interactions between 406 products, of which 149 are services (and we exclude three of them because they have no connections). Of the papers analysed, this is the one with the highest disaggregation of services using a unique homogeneous database. The rest of the works (Patelli et al., 2022; Stojkoski et al., 2016 & Zaccaria et al., 2018) use different data sources to try to achieve a comprehensive perspective on the complexity between goods and services. As documented in the works themselves that perform this aggregation of data sources, this presents strong methodological problems (Stojkoski et al., 2016). So, the methodology and retrieval of the complexity index for services is a novel contribution to the economic complexity literature.

To clarify for the reader, the transaction flow considered hereafter corresponds to the sub-matrix from figure 1 detailed in figure 2. Note that the matrix is not square, so the products missing in the row are included in the columns and vice versa. In this case, there are only three products out of a total of 406.

		Products														
		Agriculture, forestry, fishing, and hunting	Mining	Utilities	Construction	Manufacturing	Wholesale trade	Retail trade	Transportation and warehousing	Information	Finance, insurance, real estate, rental, and leasing	Professional and business services	Educational services, health care, and social assistance	Arts, entertainment, recreation, accommodation, and food services	Other services, except government	Government
Products	Agriculture, forestry, fishing, and hunting															
	Mining															
	Utilities															
	Construction															
	Manufacturing															
	Wholesale trade															
	Retail trade															
	Transportation and warehousing															
	Information															
	Finance, insurance, real estate, rental, and															
	Professional and business services															
	Educational services, health care, and social assistance															
	Arts, entertainment, recreation, accommodation, and food services															
	Other services, except government															
	Government															
	Other															
	Scrap, used and secondhand goods															

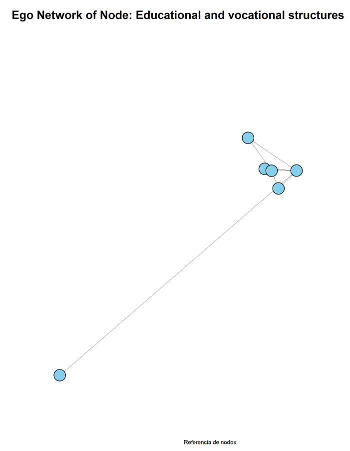
Figure 2: Supply table example for aggregate sectors. Own elaboration based on Horowitz and Planting (2009) and Young et al. (2014).

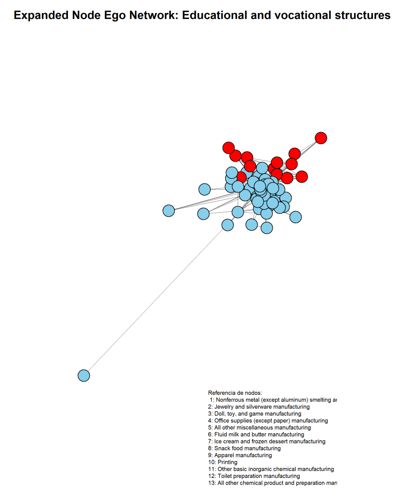
4 Results

First, it is necessary to understand the network structure of trade flows between industries. The network in different periods is all connected, with only 3 industries not trading with any other. This makes it impossible to determine, in principle, the complexity values for these cases.This high connectivity is good news for the proposed method. It is worth noting that without connectivity, such imputations would not be possible.

Another aspect to understand from the I-O flow matrix data is the relationship of service industries with the services themselves. For example, for the *Education and vocational structures* industry, there are direct relationships only with five other services. In this case, the proximity considered is with these industries (see Figure 3), but since these industries have no complexity information by definition, it is not possible to retrieve complexity information, as they all have missing data. However, an advantage of the method is that, having a distance matrix between all industries, there is a finite distance between the service and another type of industry. So, in this case, the imputation is between the service and those k nearest industries that are not categorized as services. Therefore, it is always possible to retrieve the nearest neighbour data under this condition. The larger the radius k of neighbours considered, the greater the coverage and the smaller the number of missing data. Clearly, this introduces a bias because not so close neighbours are being imputed as nearest neighbours. If we take a large k, the distance matrix itself accounts for this aspect and nuances it. Figure 3 shows this case.

Having understood this aspect, let's look at the values imputed by the method. Figure 4 shows the distribution of values by means of a box plot. It can be seen that the distribution of complexity is not sensitive to the chosen k. This shows the robustness of the method to the number of neighbours chosen. However, considering a very high k brings the problem that all cases are being placed in the same average, which removes variability from the imputations and ultimately leads to a trade-off between k and

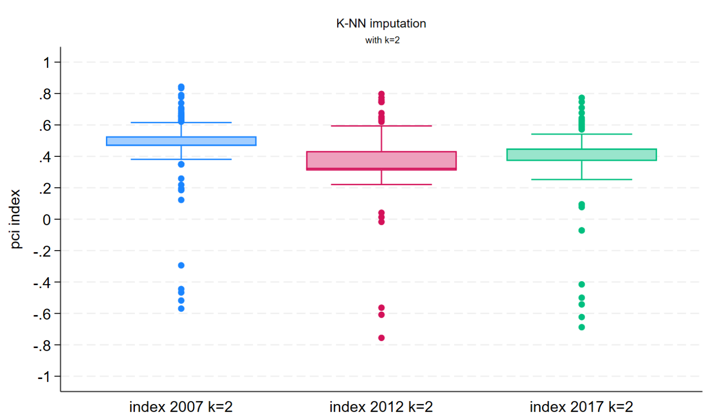




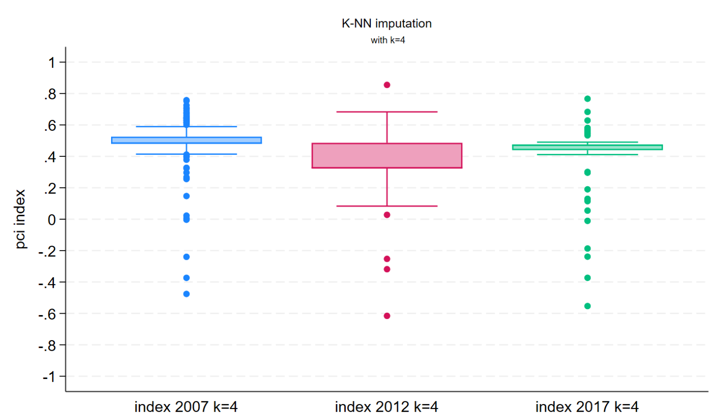
(a) Ego network (b) Ego network expanded

Figure 3: Ego networks of Educational services industry

variability. For this reason, the value of k is kept low.



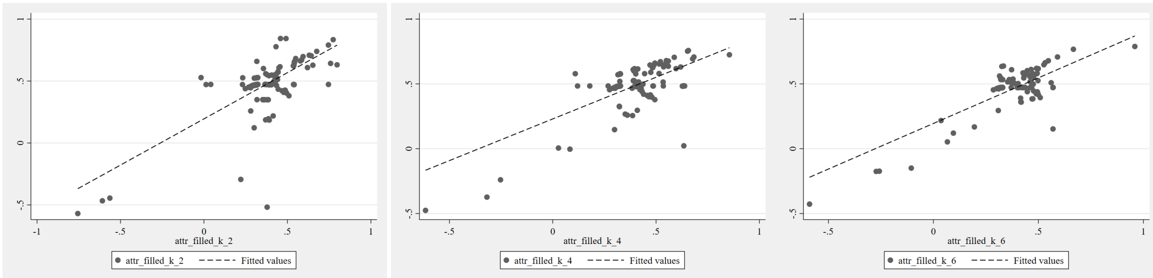
(a) 2 neighbors



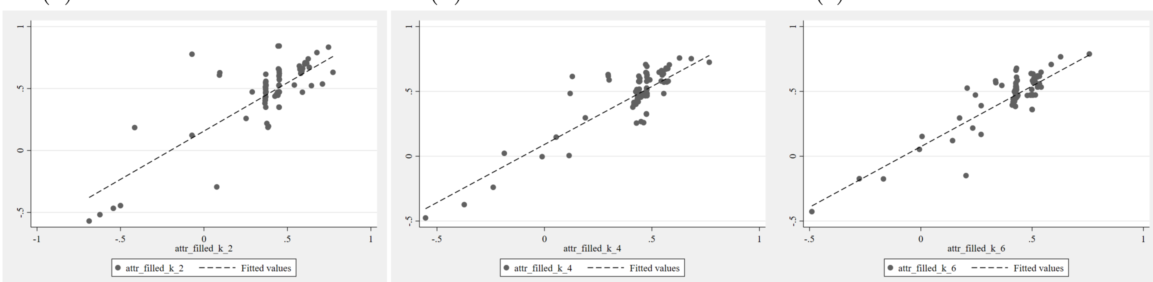
(b) 4 neighbors

Figure 4: Distribution of complexity index for services by number of k neighbors.

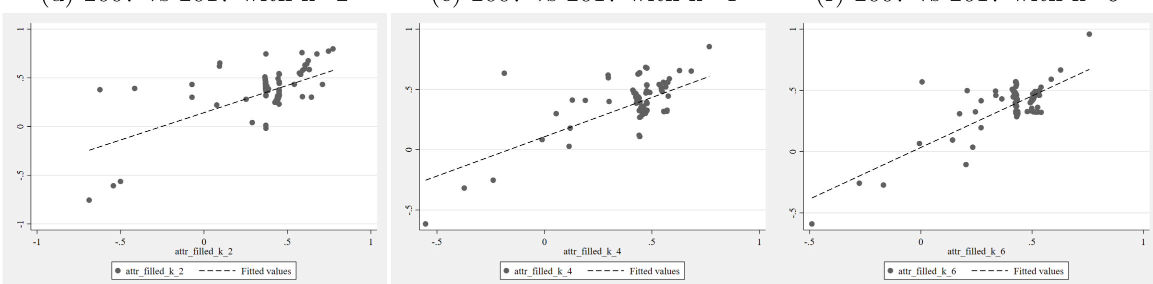
The information from the different trade matrices is then used to observe the stability over time of the imputed values. It is expected that there will be little variability between waves of information since the economy's internal trade flows do not change substantially over ten years. This is confirmed by the correlation between indices for each pair of periods compared.



(a) 2007 vs 2012 with k=2 (b) 2007 vs 2012 with k=4 (c) 2007 vs 2012 with k=6



(d) 2007 vs 2017 with k=2 (e) 2007 vs 2017 with k=4 (f) 2007 vs 2017 with k=6

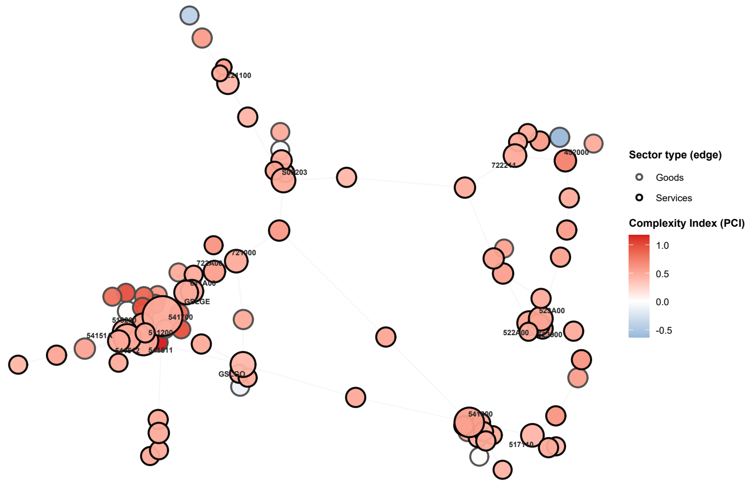


(g) 2012 vs 2017 with k=2 (h) 2012 vs 2017 with k=4 (i) 2012 vs 2017 with k=6

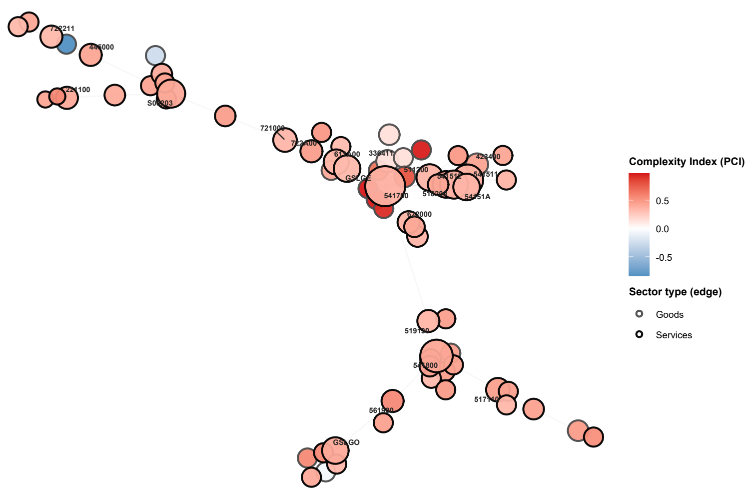
Figure 5: Correlations of complexity index for services by number of k neighbors.

4.1 The network

First, we will take advantage of the network structure of commercial relationships to observe what the structure is like. To do this, the main links in the matrix were filtered, leaving 10% of the heaviest links. This is done in order to visually observe the structure of the matrix at the core of the main transactions. What we want to highlight about these networks (see figure 6) is as follows: The main core of links is characterized by two results. One is that the nodes are red, meaning they are more complex, and the other is that the types of nodes are mostly services. This supports the idea of previous work (Hidalgo et al., 2007), which asserts that complexity occurs at the center. It also confirms the fact that current economies are characterized by the importance of services.



(a) 2007



(b) 2012

(c) 2017

Figure 6: I-O matrices, main core. Filter for the top 10% of links

Likewise, in the first row for each year of Table (3), we observe that the input–output (I–O) network has sustained a decline in density as well as in the average degree per node. Beyond this observation—which indicates a reduction in the number of linkages and, indirectly, in the complexity of the network structure—it is of interest to examine the relationship between the network structure, services, and economic complexity.

If we examine the sub-networks, progressively expanding the number of linkages, a clear pattern emerges. As connections are extended from the strongest to the weakest, the share of services declines as a larger proportion of the network is considered. The same occurs with complexity. The share of services falls from 90% to nearly 30% in all years. Average complexity declines from 0.47 to 0.40 in 2007, from 0.42 to 0.36 in 2012, and from 0.48 to 0.37 in 2017. In summary, as we move closer to the core of the network, it becomes more complex and more service-intensive.

Table 3: Network statistics according to percentile of links considered

Year	Percentile	N	L	D	$\bar{d}$	Mean pci	% services
	0	396	5403	0.04	27	0.40	31
	50	394	2693	0.02	13	0.40	32
	60	388	2144	0.01	11	0.40	32
2007	70	373	1594	0.01	8	0.39	33
	80	284	959	0.01	6	0.41	38
	85	209	605	0.01	5	0.41	50
	90	93	215	0.02	4	0.48	72
	95	8	12	0.21	3	0.47	88
	0	396	5079	0.03	25	0.36	31
	50	396	2523	0.02	12	0.36	31
	60	387	2020	0.01	10	0.35	32
	70	359	1482	0.01	8	0.35	34
2012	80	274	887	0.01	6	0.35	40
	85	191	531	0.01	5	0.38	53
	90	70	157	0.03	4	0.41	76
	95	12	23	0.17	3	0.42	92
	0	396	4330	0.03	21	0.37	31
	50	396	2459	0.02	12	0.37	31
	60	390	1964	0.01	10	0.36	31
	70	356	1438	0.01	8	0.36	34
2017	80	261	842	0.01	6	0.35	43
	85	187	508	0.01	5	0.38	56
	90	72	162	0.03	4	0.45	81
	95	14	27	0.15	3	0.48	93

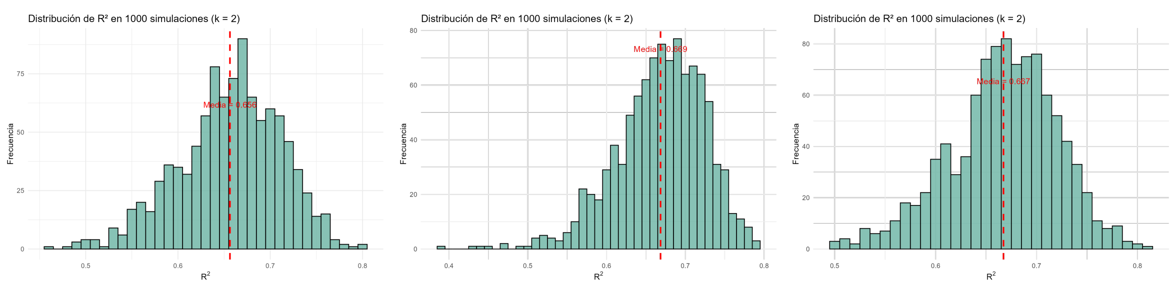
In the theory of input–output network formation, it has been shown theoretically how central nodes emerge, or what Oberfield (2018) refers to as "star sectors." These sectors arise when there are high values of α (the elasticity of the buyer's output with respect to the supplier's input), low marginal costs in these sectors, or when such costs are offset by extremely high productivity. In this regard, it is observed that services are, on average, more complex, which may be correlated with productivity. Moreover, given that services tend to have lower fixed costs compared to goods, the probability of the emergence of star sectors is higher in service-type sectors. This, in turn, confirms the empirical results in light of the theoretical findings mentioned above.

4.2 Learning complexity

Using the k-NN algorithm, it is possible to learn the index of missing values from the distance matrix between sectors. Note that by having a complete distance matrix that collects all network information, it is possible to retrieve the values of all nodes that have at least one neighbor with a non-missing value of the PCI attribute.

But one question to consider before proceeding is which inter-sector relationship matrix to use. One possibility is to use the supply matrix (SM), another is the total requirements matrix (TRM). In turn, when using the algorithm, one could use only the purchases of the sectors (the upper part of the diagonal), only the sales (the lower part), or the matrix as reported. Of all these variants, the R^2 of the training sample is initially reported, and the matrix with the highest R^2 is chosen.

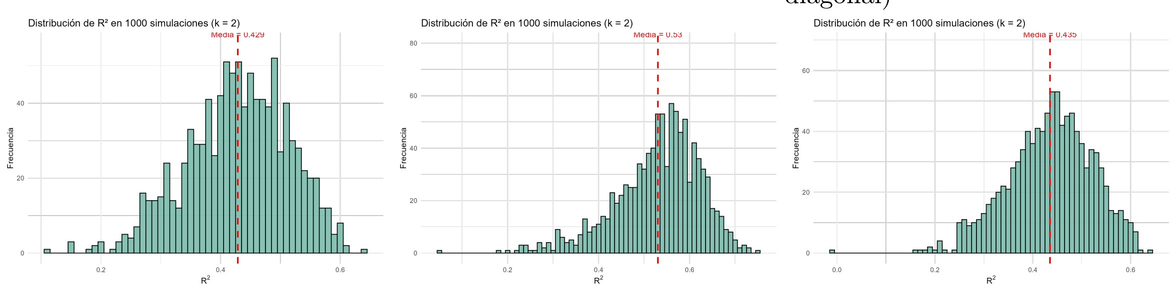
Of the 1000 samples extracted, the results (see figure 7) show a strong consistency between the real value and the estimated one for the supply matrix. This is why it was chosen to work with this matrix, as the average R^2 is almost 0.7 and, in some cases, higher than 0.8.



(a) Using SM with k=2

(b) Using SM with k=2 (values below the diagonal)

(c) Using SM with k=2 (sum values below and above the diagonal)



(d) Using SM with k=2

(e) Using SM with k=2 (values below the diagonal)

(f) Using SM with k=2 (sum values below and above the diagonal)

Figure 7: Distribution of R^2 for the 1000 simulations.

In summary, the outputs of simulations of complexity for goods are shown in table 4. The first rows are those whose values present the highest minimum values of R^2. In general, we see that the SM is better than TRM. Furthermore, in the SM matrix, the algorithm applied to the original matrix and the plus matrix are the better approximations for the index. Moreover, they are the approximations with the smallest standard deviation but also maintain very high values of R^2 in mean, median, and maximum value. Taking into account all these reasons, we choose the original SM to analyse the results. We choose a non arbitrary k number of neighbours whose R2 mean is maximum. Furthermore, a non-arbitrary k number of neighbours is chosen, which is the one that maximises the average R^2

$$\arg\max_k \mathbb{E}\left[R_k^2\right] \tag{9}$$

k=9 in this case. So from now on, the analyses will be carried out for the complexity of services with 9 neighbours. For descriptive statistics, we sometimes use 4, 6, or 9 neighbours interchangeably because the correlation is really high.

Table 4: R^2 distribution by network

Network	**mean**	**median**	**sd**	**min**	**max**
Supply-original	0.656	0.660	0.055	0.461	0.801
Supply-lower	0.669	0.674	0.056	0.389	0.793
Supply-upper	0.678	0.682	0.058	0.375	0.821
Original TR	0.429	0.433	0.085	0.112	0.637
TR-lower	0.530	0.542	0.094	0.055	0.752
TR-min	0.420	0.429	0.107	0.000	0.663
TR-upper	0.296	0.300	0.099	-0.021	0.542
Use	0.385	0.392	0.105	-0.000	0.628

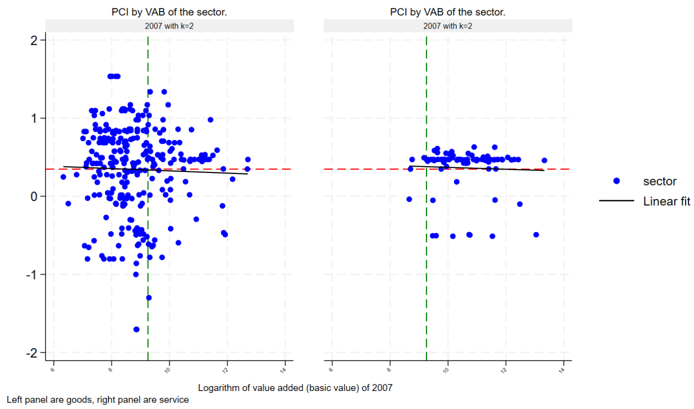
From the information collected, it stands out that most of the PCI index has positive values. This shows a high level of complexity for the services compared to the PCI of goods. In general, the results of the index are expected (see tables 10 to 18 from 5). The results for *Consultancy* as well as *Legal services* show a high relative complexity in 2007 but change to the bottom in 2012 and then come back to a middle place. Relative refers to the fact that although they are positive values, the maximum complexity for no services is 1.66. However, they are still considered highly complex, as they exceed the average of the index, which is 0.32. On the other hand, the least complex refer to basic services, personal care, and activities related to agriculture. Low complexity values are also found with historical industry activities such as mining. This is true in some cases, but the order of industries differs in each network.

There are striking results, such as cases of high complexity when they clearly do not seem to be (*Death and care services*, *Gambling*). These strange cases occur because those sectors focus on final consumption. To see this, we define an industry focused on final consumption as the ratio $\frac{\text{household consumption}}{\text{Total use of products}}$ bigger than 90%. The industries that hold the definition are shown in table 5. These cases are excluded from the rest of the analysis.

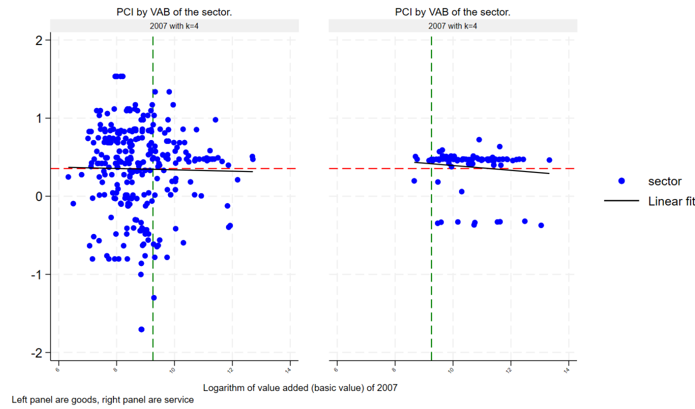
Table 5: Final consumption industries

Industries		
All other retail	Gambling industries (except casino hotels)	Nursing and community care facilities
Amusement parks and arcades	Gasoline stations	Offices of dentists
Apparel manufacturing	General merchandise stores	Offices of other health practitioners
Building material and garden equipment and supplies dealers	Grantmaking, giving, and social advocacy organizations	Offices of physicians
Child day care services	Health and personal care stores	Outpatient care centers
Clothing and clothing accessories stores	Home health care services	Owner-occupied housing
Community food, housing, and other relief services, including vocational rehabilitation services	Hospitals	Personal care services
Death care services	Individual and family services	Private households
Direct life insurance carriers	Junior colleges, colleges, universities, and professional schools	Religious organizations
Doll, toy, and game manufacturing	Motor vehicle and parts dealers	Residential mental health, substance abuse, and other residential care facilities
Elementary and secondary schools	Museums, historical sites, zoos, and parks	Tenant-occupied housing
Food and beverage stores	Newspaper publishers	

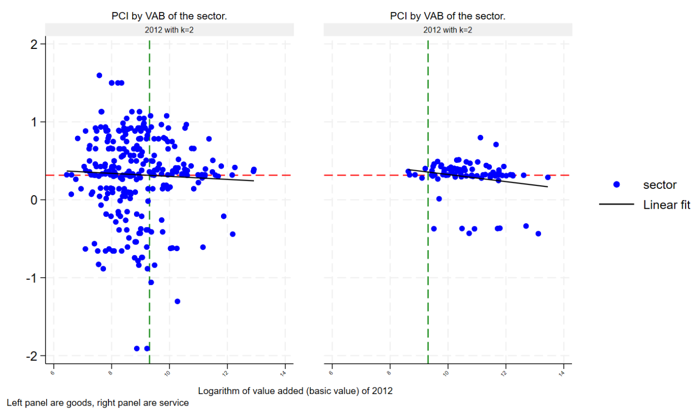
To safeguard the method and correct for these exceptions, the GDP information of the industry is used, which in this case corrects, for example, the fact that a service is complex when in reality it does not appear to be so, or does not weigh in the economy. For example, “Death and care services" in 2007 present a high complexity but with a low GDP (see 8), so the industry is being overemphasised.



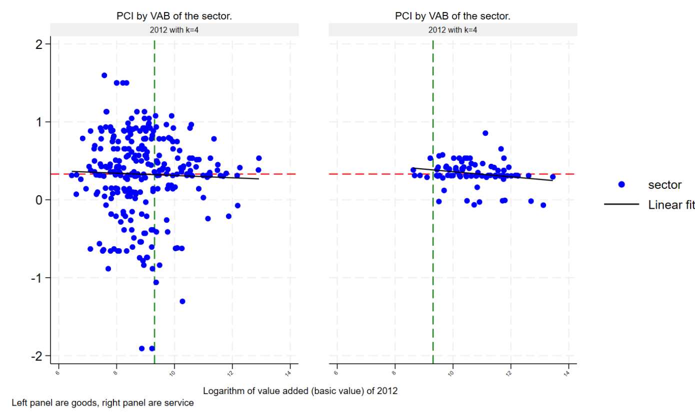
(a) 2007 with k=2



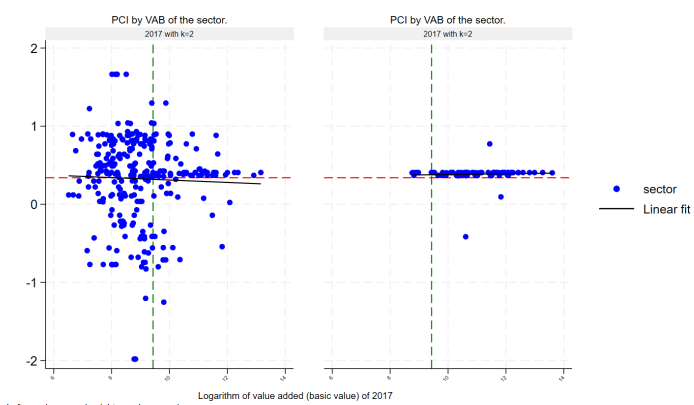
(b) 2007 with with k=4



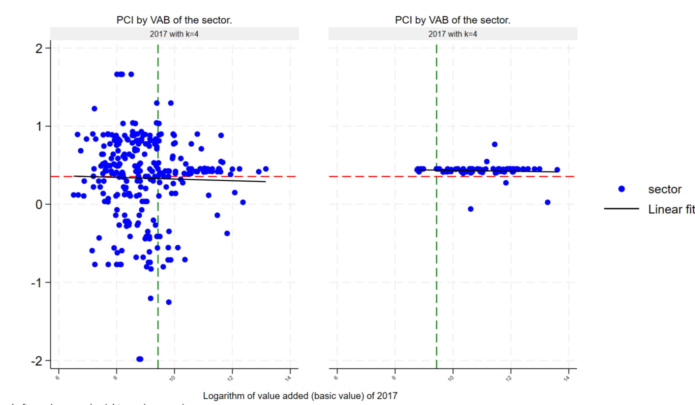
(c) 2012 with k=2



(d) 2012 with k=4



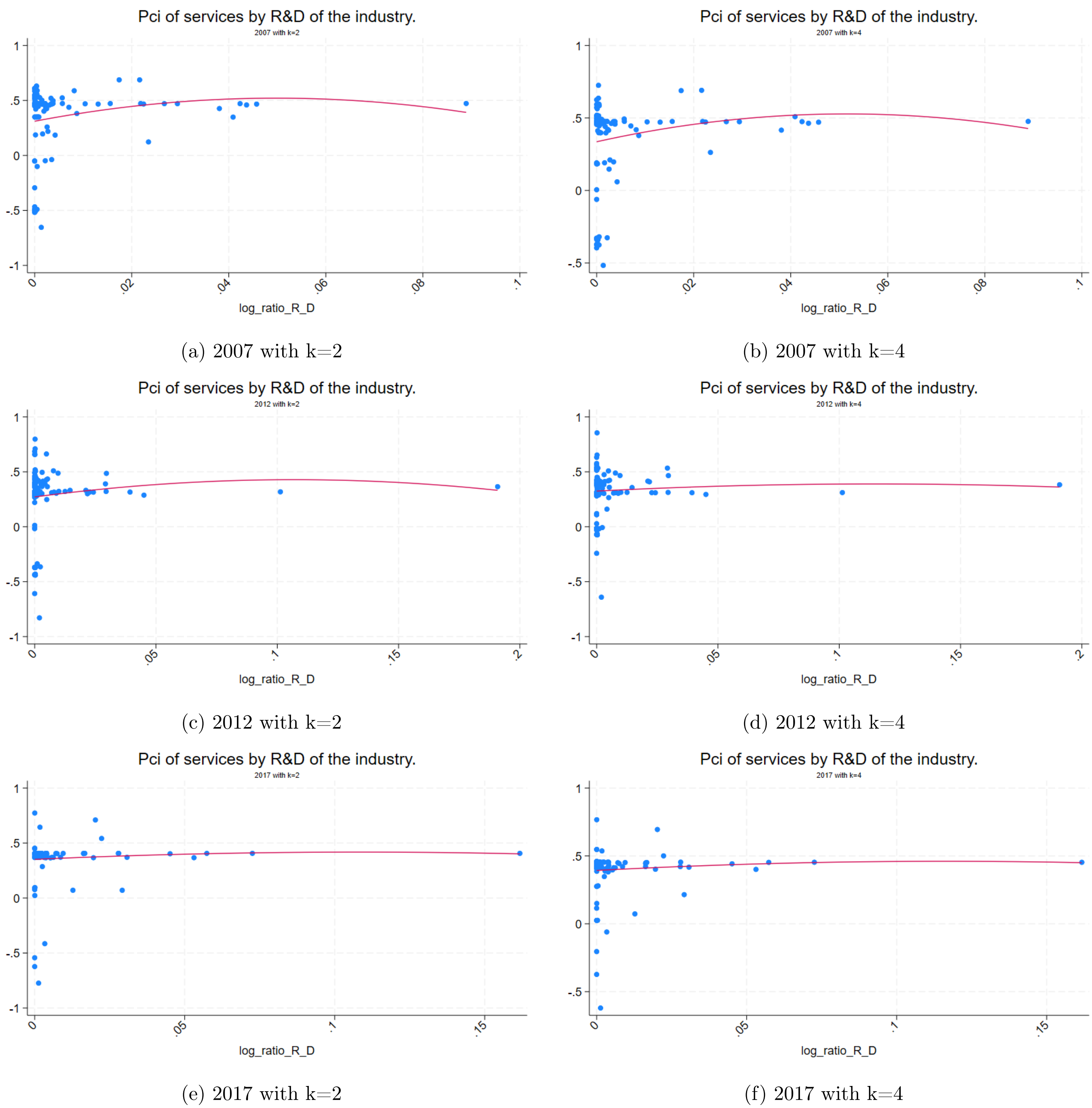
(e) 2017 with k=2



(f) 2017 with k=4

Figure 8: Mapping the complexity by GDP industry, by k.

As validation of robustness, we use the information on R&D investment by sectors extracted from the BEA. The complexity correlates with R&D investment in sectors, which further proves that complexity captures the sectors that invest the most in knowledge. Investing in knowledge would be indirectly linked to capacity building. As we can see in the figure 9, the relationship is not negative and is sometimes quadratically positive. This is not a clear demonstration, but it is an indication that, although the indicator is not clearly positive in its relationship with investment in knowledge, at least it does not have striking results contrary to the theory.



(a) 2007 with k=2

(b) 2007 with k=4

(c) 2012 with k=2

(d) 2012 with k=4

(e) 2017 with k=2

(f) 2017 with k=4

Figure 9: Distribution of PCI of services by number of k neighbors, years and R+D.

4.3 Services are more complex

To enable a comparison with previous studies, we analyze the average complexity of service sectors. Stojkoski et al. (2016) show that service sectors are more complex than goods. In our case, a similar pattern is observed. The descriptive statistics in Table (10) indicate that services exhibit higher average complexity and lower volatility than goods.

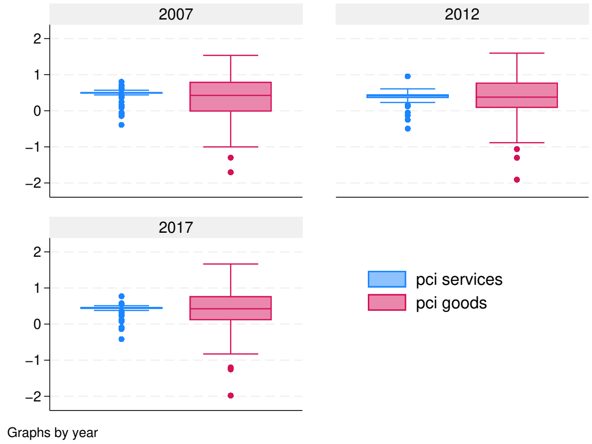


Figure 10: Distribution of Economic Complexity Index by year

In order to see the statistical difference, the difference between sectors is observed by the fact of being services as in Stojkoski et al. (2016).

$$PCI_{i,t} = \Delta\mu_s D_s + \eta_t + \eta_i + e_{i,t} \tag{10}$$

where $e_{i,t}$ is the error, $\Delta\mu_s$ the additional effect of being services and η_i is the sector fixed effect and η_t fixed effect per time.

Table 6: Panel regressions: mean difference between goods and services

	(1) pci
services	0.111*** (0.0335)
year=2007	0 (.)
year=2012	-0.0390*** (0.00700)
year=2017	-0.0295*** (0.00679)
Constant	0.363*** (0.0335)
Observations	1191

Standard errors in parentheses
Fixed effects: Yes (by sector)
Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Growth of GDP by states and PCI

In this section, we use the GDP of each state in the USA to test the traditional hypothesis that complexity determines growth. But, what type of complexity matters for growth?

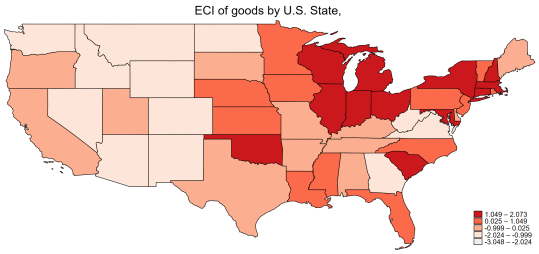
The main goal is to see, in the mean, what drives the economic growth of each state. In the traditional works, the analysis is by countries. In this sense, the standard approach found a correlation between growth and complexity. But this complexity metric is summarized in the ECI, which measures the complexity of the country. Here, the analysis is the same using the complexity by state of the USA. For that purpose, we use the complexity index of each good and service. Then, the revealed comparative advantage (RCA) in each state for each good and service is used to construct the new complexity index by state.

$$k_{state} = \frac{1}{M_{state}} \sum_{p} M_{state,p} PCI_p \quad (11)$$

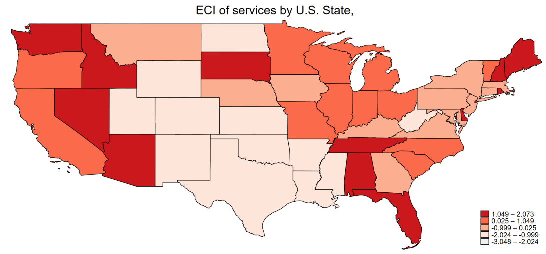
$$ECI_{state} = \frac{k_{state} - mean(k_{state})}{\sigma(k_{state})} \quad (12)$$

Where $M_{state,p}$ is the RCA of each product p in the state.

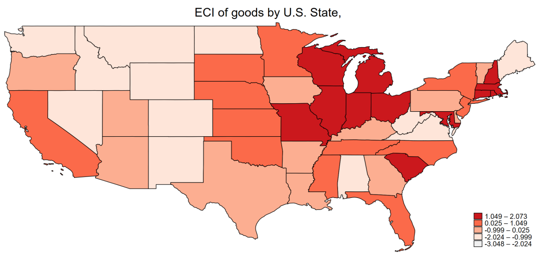
The service complexity indicator captures the complexity of the West Coast (see the right column of figure 11), which appears to be home to the largest technology hubs. The major cities in technology are related to San Francisco, Boston, Durham, and Austin (Clarke et al., 2016; Moretti, 2012). This measure captures a complexity that is not apparent when looking at goods alone (see the left column comparing with the right column, Figure 11).



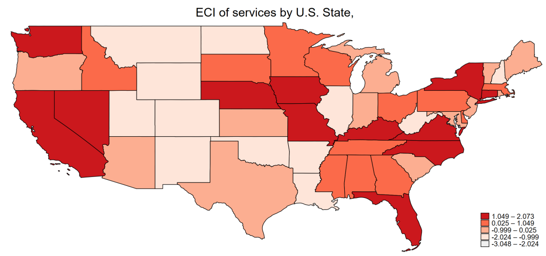
(a) ECI goods — year 2007



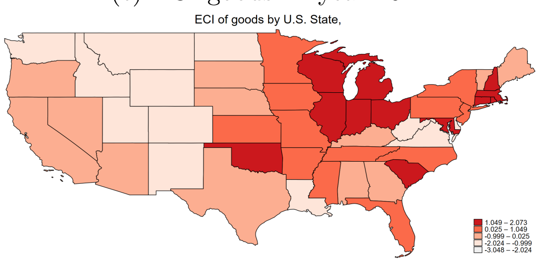
(b) ECI services — year 2007



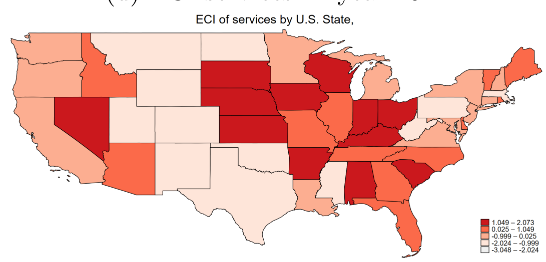
(c) ECI goods — year 2012



(d) ECI services — year 2012



(e) ECI goods — year 2017



(f) ECI services — year 2017

Figure 11: ECI by State and year

The econometric specifications consist of two. One includes all states, all industries, and the three periods. The other specification is a typical growth equation with one observation per State per year; in this case, the complexity is summarized in the mean of each type.

$$GDP_{growth_{i,t}} = \alpha + \beta_s \, eci^S_{i,t-1} + \beta_g \, eci^G_{i,t-1} + \phi \, \ln(GDP_{i,t-1}) + \boldsymbol{\delta}^\top \mathbf{X}_{i,t-1} + \gamma_i + \tau_t + \varepsilon_{i,t} \quad (13)$$

If the coefficient β_s is positive, this means that the complexity of the product i determines the growth. The hypothesis is confirmed with a positive β_s

4.5 Estimations: FE state

Following the pioneering work of Hidalgo and Hausmann (2009), a positive relationship between the complexity of goods and growth is expected. When services are added, previous work does not expect major changes. However, in our case, initial explorations show a positive and significant effect on growth prediction by the ECI for services rather than goods (see table 7). This confirms our hypothesis and rejects the postulate of not using measures of service complexity because they are redundant.

Table 7: Panel regressions: Real Growth and ECI

	(1) growth	(2) growth	(3) growth
rgpd pc t-1	0.653***	0.555***	0.574***
	(0.102)	(0.128)	(0.139)
ECI of goods t-1	0.0173		0.0111
	(0.0118)		(0.0137)
ECI of services t-1		0.0131***	0.0117**
		(0.00451)	(0.00544)
Pers income pc real t-1	-0.0000150***	-0.0000133**	-0.0000143**
	(0.00000502)	(0.00000527)	(0.00000570)
PCE t-1 pc	-14.99	-23.75**	-21.95**
	(12.13)	(12.03)	(10.75)
PCE real serv t-1 pc	16.91	29.71*	28.41*
	(16.77)	(16.03)	(14.93)
Population t-1	-2.27e-09**	-2.09e-09**	-2.43e-09**
	(1.09e-09)	(9.44e-10)	(9.69e-10)
Constant	2.798***	2.431***	2.497***
	(0.435)	(0.562)	(0.588)
Observations	100	100	100

Standard errors in parentheses
Fixed effects: Yes (by state)
Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The summary of the results of the different specifications is shown in Table (8). The results show a positive effect in all specifications for the ECI, with significance only in one at the 10% level. Services then show a positive and significant effect with fixed effects but not with time controls. The effect is significant, considering that the average growth is 0.0138 (1.4 per cent) and that ECI_s ranges between 1.820459 and -1.88775. A one-unit increase in the services ECI is associated with an approximately 1 percentage-point increase in GDP growth.

Specification	ECI of goods t-1	ECI of services t-1
controls in levels — FE state	0.0173	0.0131***
controls in levels, both ECI — FE state	0.0111	0.0117**
controls in logs — FE state	0.0195*	0.0112**
controls in logs, both ECI — FE state	0.0137	0.0094
controls in levels — FE state & time	0.01112	-0.0208
controls in levels, both ECI — FE state & time	0.013	-0.0236
controls in logs — FE state & time	0.0132	-0.0211
controls in logs, both ECI — FE state & time	0.0147	-0.0229

Table 8: Regression coefficients of eci
See the whole output regression in Annex 5

If we look at some studies that analyse this type of relationship with economic growth, the relationship that is expected to be found is weak and ambiguous, mainly because they are observing short-term growth trajectories (Ourens et al., 2013). Following this idea, we will explore various ways of measuring the relationship with different specifications: fixed effects by territory, fixed effects by territory with demand controls, and the same including time controls.

4.6 Estimations: First difference

The above specifications present fixed effects by state and by state and year, which is necessary to remove unobservable heterogeneity due to characteristics intrinsic to the states or to shocks over time. However, degrees of freedom are lost, which affects the power of the estimates. In this regard, we seek to mitigate this weakness by taking advantage of the panel structure. To do so, we take first differences and observe the effect of a variation in the ECI on growth.

$$log(GDP_{i,t}) = \alpha + \beta_s\, ECI^S_{i,t-1} + \beta_g\, ECI^G_{i,t-1} + \phi\, \ln(GDP_{i,t-1}) + \boldsymbol{\delta}^\top \mathbf{X}_{i,t-1} + \gamma_i + \tau_t + \varepsilon_{i,t} \tag{14}$$

Difference equation (14) increases degrees of freedom:

$$\Delta log(GDP_{i,t}) = \alpha + \beta_s\, \Delta ECI^S_{i,t-1} + \beta_g\, \Delta ECI^G_{i,t-1} + \phi\, \Delta \ln(GDP_{i,t-1}) + \boldsymbol{\delta}^\top \Delta\mathbf{X}_{i,t-1} + \tau_t + \varepsilon_{i,t}$$

The focus is once again on ECI^S in this case, we observe how the acceleration of ECI affects the first difference of the log of GDP.: $\boldsymbol{\beta_s} = \frac{\boldsymbol{\partial \Delta log(GDP_{i,t})}}{\boldsymbol{\partial \Delta ECI^S_{i,t-1}}}$. Under this new specification, the positive effect of services on growth is reconfirmed.

The effects of services confirm their importance, but the effect of the complexity of goods does not have significant effects. While this is striking in relation to complexity theory and empirical evidence, many works for the United States have documented a process of deindustrialization characterised by a large divergence between cities and regions, which has widened the growth gap between the most productive regions via technology, such as the cities mentioned above, to the detriment of traditionally industrial regions (Clarke et al., 2016; Moretti, 2012). This has been partly driven by the process of opening up and local disinvestment towards regions with higher marginal productivity of labour and capital, such as China. Although the effect has not been negative on employment (Magyari et al., 2017), it has led to a reallocation. For example, Bloom et al. (2024) find that the process of imports from China resulted in a larger reallocation from manufacturing to services jobs; almost 40% of jobs have been reallocated from industry to services (research, management, and wholesale.). This process of deaccumulation of capabilities may explain the weak result of goods complexity for this period in the United States.

Table 9: Panel regressions: Real Growth and ECI

	(1) Log real gdp pc t in dif	(2) Log real gdp pc t in dif
$\boldsymbol{ECI^G}$ t-1 in dif	0.00144	0.00204
	(0.0156)	(0.0157)
$\boldsymbol{ECI^S}$ t-1 in dif	0.0343**	0.0348**
	(0.0154)	(0.0161)
Pers income pc real t-1 in dif	0.601***	0.617***
	(0.154)	(0.155)
Log PCE t-1 in dif	0.787*	0.827*
	(0.397)	(0.436)
Log PCE real serv t-1 in dif	-0.216	-0.271
	(0.363)	(0.450)
Log Population t-1 in dif	-0.894***	-0.905***
	(0.201)	(0.202)
year=3	0	0
	(.)	(.)
Log real gdp pc t-1 in dif		0.0161
		(0.0622)
Constant	0.0867**	0.0895**
	(0.0375)	(0.0420)
Observations	50	50

Standard errors in parentheses
Fixed effects: Yes (by state)
Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.7 Estimations: Spillovers in dependent and independent variables

A new specification is used to correct for spatial correlation bias between states in the United States. The Spatial Durbin Model is used for this purpose, which allows this bias to be corrected using the following specification:

$$\boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{W X \theta} + \boldsymbol{\epsilon} \tag{15}$$

Where W is a spatial correlation matrix that measures the distance between each of the states of the United States. To measure the distance, the average of the coordinates of each polygon is used, i.e., a point close to the centroid of each state. $\boldsymbol{Wy}$ measures the spatial dependence of the dependent variable; in this case, growth. In the same way, $\boldsymbol{WX}$ measures the spatial dependence of the regressors. If $\boldsymbol{\rho}$ is significantly different from zero, the growth of one state influences the other, the same as with $\boldsymbol{\theta}$ in the case of independent variables.

The model can be rewritten as follows:

$$\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{\rho W})^{-1}(\boldsymbol{X\beta} + \boldsymbol{WX\theta}) \tag{16}$$

Where it can be seen that marginal effects are no longer interpreted directly because there are feedback effects. For a variable $\boldsymbol{x_i}$, the effect it has on $\boldsymbol{y}$ is as follows:

$$\frac{\boldsymbol{\partial y}}{\boldsymbol{\partial x_i}} = (\boldsymbol{I} - \boldsymbol{\rho W})^{-1}(\boldsymbol{\beta_i I} + \boldsymbol{\theta_i W}) \tag{17}$$

The results show a spatial dependence in the complexity of services (0.0697), which means that an increase in the complexity of services has multiplier or spillover effects in neighbouring states. When looking at the total effect of the parameters, as should be done in this type of modelling, a positive and significant (0.0450) long-term total effect (considering direct and indirect effects) is observed. The total effect is explained by the spillover generated by increased service complexity for neighbouring states. This generates a spillover in growth through the services of neighbouring states. This reconfirms the positive effect of services on growth, but this time only through indirect spillover effects to other States.

Table 10: Spatial Durbin Model – Main, Spatial, and Long-Run Effects

	Coef.	Std. Err.	Sig.
Main effects			
$\ln(\text{GDP pc})_{t-1}$	0.8359	0.0756	***
ECI^g_{t-1}	0.0106	0.0183	
ECI^s_{t-1}	-0.0167	0.0168	
Personal income pc (real)	-0.0000169	0.0000033	***
Spatially lagged regressors (W × X)			
$W \ln(\text{GDP pc})_{t-1}$	0.6765	0.8577	
$W\ ECI^g_{t-1}$	-0.0774	0.1304	
$W\ ECI^s_{t-1}$	0.0697	0.0285	**
W Personal income pc	0.0000366	0.0000209	*
Spatial parameter			
ρ	-0.1966	0.4379	
variance of error			
σ^2_e	0.0003266	0.0000668	***
Long-run Direct Effects			
$\ln(\text{GDP pc})_{t-1}$	0.8333	0.0789	***
ECI^g_{t-1}	0.0105	0.0183	
ECI^s_{t-1}	-0.0156	0.0116	
Personal income pc (real)	-0.0000171	0.0000033	***
Long-run Indirect Effects			
$\ln(\text{GDP pc})_{t-1}$	0.4182	0.6696	
ECI^g_{t-1}	-0.0543	0.1674	
ECI^s_{t-1}	0.0607	0.0271	**
Personal income pc (real)	0.0000346	0.0000276	
Long-run Total Effects			
$\ln(\text{GDP pc})_{t-1}$	1.2515	0.6822	*
ECI^g_{t-1}	-0.0438	0.1699	
ECI^s_{t-1}	0.0450	0.0262	*
Personal income pc (real)	0.0000175	0.0000283	

Observations	100
Groups	50
Time periods	2
Estimation	Spatial FE, robust SEs
Log-pseudolikelihood	259.3791

Notes: Spatial Durbin Model estimated with `xsmle`, spatial fixed effects, balanced panel. Robust standard errors clustered at state level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusion

The economy has reallocated resources from agriculture to manufacturing and subsequently to services. This reallocation has shifted the economy toward services driven by technology, retail, research, legal services, accounting, and managerial tasks, all of which entail new capabilities aimed at generating value within these service sectors. The services sector accounts for the largest share of value measured by GDP in modern economies. Under these stylized facts, it is necessary to reformulate the approach to complexity—traditionally measured through goods—toward the economy as a whole. This work contributes to the literature on complexity by adding evidence of the implications of using information about services that is not found in the original index. Such an approach contributes to a better understanding of the new economy led by the service sector.

The study offers new evidence on the measurement of complexity in service sectors by introducing a novel machine learning method. It allows us to "learn" about missing information—in this case, the complexity of services. Based on the idea of homophily across sectors, missing values are imputed using the network of intersectoral trade flows. This methodology enables us to provide evidence on the distribution of complexity values within services. These two contributions are novel within the complexity literature.

The results show that the complexity of the services sector is not redundant information relative to goods complexity. In fact, the complexity of services is on average significantly higher than that of goods by 0.11 units. The services sector is relevant in explaining the dynamics of new value creation in the economy. In fact, it is shown in this work to constitute the core of the input–output network of the U.S. economy. The core of the network is led by complex services. This finding is supported by several specifications that validate the importance of complex services for growth and the economic structure of the United States. For the main estimates, a one-unit increase in the $\boldsymbol{ECI^S}$ is associated with approximately a 1 percentage point increase in the growth rate. This does not imply disregarding the importance of goods, as there is complementarity between many goods and services, and both types of sectors reinforce one another. However, the results for the 2007–2017 period in the United States demonstrate that neglecting services leads to serious omissions.

This work provides empirical evidence on the importance of services in the economic structure from a capabilities-based perspective, thereby confirming earlier theories that have documented the structural transformation of modern economies. Although this stylized fact is well known, it has not been widely incorporated into the complexity literature.

This study does not delve deeply into theory or the potential mechanisms explaining this finding; however, it outlines possible reasons, such as differences in costs and fixed investments that may be influencing the results—leaving room for further research. Another limitation concerns scalability, as global input–output matrices (OECD, 2024) include fewer sectors compared to the data used in this study. Moreover, heterogeneity within services must be examined in future work—for example, identifying which sectors are key drivers of growth and how they interact with manufactured and agricultural goods.

Another important aspect to consider is the heterogeneity of services and their particular nature; that is, they may have the capacity to be more ubiquitous and have a different cost structure, which suggests that future research should further examine whether alternative network representations may better capture the specific characteristics of services. Future research should investigate how services behave within the structure of the economy compared to goods and how these differences affect the measurement of complexity.

An additional aspect that goes beyond the scope of this research, but which is related to the identification of growth equations, is that complexity, although a concept in itself, can also be considered an instrument for estimating the effects of variables associated with capabilities, innovation, or knowledge creation. This idea is not developed here, but it represents a possible avenue for future research at the intersection of econometric literature on growth and economic complexity.

Finally, we acknowledge the limited number of periods covered by the research and the intrinsic limitations of growth equations. Beyond these weaknesses, the results provide consistent evidence of the importance of services and their developed capabilities as an engine of economic growth.

References

Antonelli, C., Crespi, F., & Quatraro, F. (2022). Knowledge complexity and the mechanisms of knowledge generation and exploitation: The european evidence. *Research Policy*, *51*(8), 104081.

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., o'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450. https://doi.org/https://doi.org/10.1016/j.respol.2021.104450

Bartelme, D., Lan, T., & Levchenko, A. A. (2024). Specialization, market access and real income. *Journal of International Economics*, *150*, 103923.

Batista, G., & Monard, M. C. (2002). A study of k-nearest neighbour as an imputation method. *Soft Computing Systems: Design,. . .*

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, *81*(2), 608–650.

Bloom, N., Handley, K., Kurmann, A., & Luck, P. A. (2024). *The china shock revisited: Job reallocation and industry switching in us labor markets* (tech. rep.). National Bureau of Economic Research.

Cardoso, B.-H. F., Catela, E. Y. d. S., Viegas, G., Pinheiro, F. L., & Hartmann, D. (2023). Export complexity, industrial complexity and regional economic growth in brazil. *arXiv preprint arXiv:2312.07469*.

Clarke, G., Martin, R., & Tyler, P. (2016). Divergent cities? unequal urban growth and development. *Cambridge Journal of Regions, Economy and Society*, *9*(2), 259–268. https://doi.org/10.1093/cjres/rsw011

Domini, G. (2022). Patterns of specialization and economic complexity through the lens of universal exhibitions, 1855-1900. *Explorations in Economic History*, *83*, 101421.

Durlauf, S. N. (2009). The rise and fall of cross-country growth regressions. *History of Political Economy*, *41*(Suppl_1), 315–333.

Gao, J., & Zhou, T. (2018). Quantifying china's regional economic complexity. *Physica A: Statistical Mechanics and its Applications*, *492*, 1591–1603.

Hastie, T., Tibshirani, R., Friedman, J., et al. (2009). The elements of statistical learning.

Hausmann, R., & Hidalgo, C. A. (2010, October). *Country diversification, product ubiquity, and economic divergence* (CID Working Paper No. 201). Center for International Development, Harvard Kennedy School.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., & Simoes, A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. Mit Press.

Herrendorf, B., Rogerson, R., & Valentinyi, Á. (2013). Two perspectives on preferences and structural transformation. *American Economic Review*, *103*(7), 2752–2789. https://doi.org/10.1257/aer.103.7.2752

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, *3*(2), 92–113.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10570–10575. https://doi.org/10.1073/PNAS.0900943106

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487.

Hidalgo, C. A., & Stojkoski, V. (2025). The theory of economic complexity. *arXiv preprint arXiv:2506.18829*.

Horowitz, K. J., & Planting, M. A. (2009). Concepts and methods of the us input-output accounts. *Bureau of Economic Analysis*.

Koch, P. (2021). Economic complexity and growth: Can value-added exports better explain the link? *Economics Letters*, *198*, 109682. https://doi.org/https://doi.org/10.1016/j.econlet.2020.109682

Li, N., Meleshchuk, S., Yin, Q., Zhao, D., & Zymek, R. (2025). *Bilateral trade in services: Insights from a new research dataset* (tech. rep.). International Monetary Fund.

Lo Turco, A., & Maggioni, D. (2022). The knowledge and skill content of production complexity. *Research Policy*, *51*(8), 104059.

Magyari, I., et al. (2017). Firm reorganization, chinese imports, and us manufacturing employment. *Columbia University*.

Mao, Z., & An, Q. (2021). Economic complexity index and economic development level under globalization: An empirical study. *Journal of Korea Trade*, *25*(7), 41–55.

Mealy, P., Farmer, J. D., & Teytelboym, A. (2019). Interpreting economic complexity. *Science advances*, *5*(1), eaau1705.

Mishra, S., Tewari, I., & Toosi, S. (2020). Economic complexity and the globalization of services. *Structural Change and Economic Dynamics*, *53*, 267–280.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Oberfield, E. (2018). A theory of input–output architecture. *Econometrica*, *86*(2), 559–589.

OECD. (2024). Inter-country input-output (icio) tables, 2024 release.

Ortiz-Ospina, E., Beltekian, D., & Roser, M. (2018). Trade and globalization. our world in data.

Ourens, G., et al. (2013). *Can the method of reflections help predict future growth?* IRES.

Patelli, A., Pietronero, L., & Zaccaria, A. (2022). Integrated database for economic complexity. *Scientific Data*, *9*(1), 628.

Pérez-Balsalobre, S., Llano Verduras, C., & Díaz-Lanchas, J. (2019). *Measuring subnational economic complexity: An application with spanish data* (tech. rep.). JRC Working Papers on Territorial Modelling and Analysis.

Poncet, S., & De Waldemar, F. S. (2013). Export upgrading and growth: The prerequisite of domestic embeddedness. *World Development*, *51*, 104–118.

Stojkoski, V., & Kocarev, L. (2017). The relationship between growth and economic complexity: Evidence from southeastern and central europe.

Stojkoski, V., Utkovski, Z., & Kocarev, L. (2016). The impact of services on economic complexity: Service sophistication as route for economic growth. *PloS one*, *11*(8), e0161633. https://doi.org/10.1371/journal.pone.0161633.

Teixeira, F. O., Missio, F. J., & Dathein, R. (2022). Economic complexity, structural transformation and economic growth in a regional context: Evidence for brazil. *PSL Quarterly Review*, *75*(300), 63–79.

Temple, J. R. (2021). Growth econometrics. In *Oxford research encyclopedia of economics and finance*.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for dna microarrays. *Bioinformatics*, *17*(6), 520–525.

West, D. (1996). *Introduction to graph theory, prentice hall, inc., upper saddle river, nj.*

Young, J. A., Howells, T. F., Strassner, E. H., & Wasshausen, D. B. (2014). Supply-use tables for the united states. *Survey of Current Business*, *94*(5), 1–8.

Zaccaria, A., Mishra, S., Cader, M. Z., & Pietronero, L. (2018). Integrating services in the economic fitness approach. *World Bank Policy Research Working Paper*, (8485).

Annex

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Full-service restaurants	0.625	0.581	0.508
Automotive repair and maintenance (including car washes)	0.449	0.569	0.492
Federal electric utilities	0.575	0.546	0.479
Monetary authorities and depository credit intermediation	0.580	0.548	0.476
State and local government (hospitals and health services)	0.452	0.477	0.475
Other state and local government enterprises	0.451	0.477	0.475
Pipeline transportation	0.451	0.477	0.475
Photographic services	0.450	0.476	0.475
Scenic and sightseeing transportation and support activities	0.450	0.476	0.475
Truck transportation	0.449	0.475	0.475
Insurance agencies, brokerages, and related activities	0.450	0.476	0.475
Insurance carriers, except direct life	0.450	0.476	0.475
State and local government (other services)	0.449	0.475	0.475
Dry-cleaning and laundry services	0.449	0.475	0.475
Natural gas distribution	0.447	0.473	0.474
Warehousing and storage	0.450	0.443	0.474
Couriers and messengers	0.449	0.443	0.474
Transit and ground passenger transportation	0.449	0.443	0.474
Postal service	0.446	0.472	0.473
Limited-service restaurants	0.441	0.466	0.437
Satellite, telecommunications resellers, and all other telecommunications	0.441	0.466	0.437
Other amusement and recreation industries	0.440	0.465	0.437
Computer systems design services	0.442	0.468	0.437
State and local government (educational services)	0.439	0.464	0.437
Directory, mailing list, and other publishers	0.437	0.463	0.437
Other ambulatory health care services	0.444	0.470	0.437
State and local government passenger transit	0.444	0.470	0.437
Accounting, tax preparation, bookkeeping, and payroll services	0.445	0.471	0.437

Table 11: Top 2017

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Waste management and remediation services	0.445	0.471	0.437
Legal services	0.445	0.471	0.437
All other miscellaneous professional, scientific, and technical services	0.446	0.471	0.437
Residential maintenance and repair	0.446	0.472	0.436
Architectural, engineering, and related services	0.447	0.472	0.436
Specialized design services	0.447	0.473	0.436
Other federal government enterprises	0.433	0.457	0.436
Accommodation	0.447	0.473	0.436
Facilities support services	0.448	0.474	0.436
Management of companies and enterprises	0.448	0.474	0.435
Other computer related services, including facilities management	0.448	0.474	0.435
Employment services	0.449	0.475	0.435
Internet publishing and broadcasting and Web search portals	0.449	0.475	0.435
Custom computer programming services	0.449	0.444	0.435
Other financial investment activities	0.450	0.476	0.434
News syndicates, libraries, archives and all other information services	0.450	0.476	0.434
Securities and commodity contracts intermediation and brokerage	0.450	0.476	0.434
Office administrative services	0.449	0.444	0.434
Other durable goods merchant wholesalers	0.443	0.470	0.434
Software publishers	0.450	0.476	0.434
Commercial and industrial machinery and equipment repair and maintenance	0.444	0.471	0.434
Scientific research and development services	0.450	0.444	0.434
Other educational services	0.451	0.476	0.433
Personal and household goods repair and maintenance	0.451	0.477	0.433
Air transportation	0.371	0.441	0.433
Management consulting services	0.425	0.448	0.433
Travel arrangement and reservation services	0.371	0.443	0.433
Advertising, public relations, and related services	0.452	0.477	0.433

Table 12: Middle 2017

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Other personal services	0.371	0.443	0.432
Veterinary services	0.371	0.443	0.432
Water transportation	0.371	0.443	0.432
General and consumer goods rental	0.371	0.441	0.432
Periodical Publishers	0.371	0.441	0.432
Support activities for agriculture and forestry	0.371	0.440	0.432
Promoters of performing arts and sports and agents for public figures	0.371	0.440	0.432
Investigation and security services	0.371	0.439	0.432
Spectator sports	0.370	0.438	0.432
Book publishers	0.370	0.436	0.431
Independent artists, writers, and performers	0.370	0.437	0.431
Other support services	0.370	0.436	0.431
Funds, trusts, and other financial vehicles	0.370	0.436	0.431
Automotive equipment rental and leasing	0.370	0.436	0.431
State and local government electric utilities	0.370	0.435	0.431
Wireless telecommunications carriers (except satellite)	0.370	0.434	0.430
Wired telecommunications carriers	0.370	0.434	0.430
Performing arts companies	0.369	0.432	0.429
Business support services	0.369	0.430	0.428
Services to buildings and dwellings	0.368	0.425	0.426
Radio and television broadcasting	0.367	0.418	0.421
Data processing, hosting, and related services	0.367	0.416	0.420
Environmental and other technical consulting services	0.365	0.411	0.417
Grocery and related product wholesalers	0.383	0.450	0.329
Other support activities for mining	-0.415	-0.011	0.072
Wholesale electronic markets and agents and brokers			
Federal general government (nondefense)			
Civic, social, professional, and similar organizations			

Table 13: Bottom 2017

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
State and local government electric utilities	0.404	0.627	0.554
Funds, trusts, and other financial vehicles	0.401	0.628	0.554
Other support services	0.399	0.629	0.554
Investigation and security services	0.386	0.632	0.553
Promoters of performing arts and sports and agents for public figures	0.384	0.633	0.553
Water transportation	0.375	0.635	0.553
Other personal services	0.365	0.638	0.552
Commercial and industrial machinery and equipment repair and maintenance	0.492	0.683	0.526
Full-service restaurants	0.676	0.588	0.490
Environmental and other technical consulting services	0.509	0.494	0.489
Other educational services	0.376	0.534	0.486
General and consumer goods rental	0.370	0.398	0.481
Spectator sports	0.379	0.402	0.479
Independent artists, writers, and performers	0.390	0.407	0.478
Performing arts companies	0.417	0.422	0.478
Radio and television broadcasting	0.487	0.473	0.474
Data processing, hosting, and related services	0.486	0.472	0.471
Book publishers	0.445	0.440	0.455
Services to buildings and dwellings	0.441	0.437	0.453
Wired telecommunications carriers	0.427	0.428	0.449
Wireless telecommunications carriers (except satellite)	0.421	0.424	0.448
Periodical Publishers	0.427	0.428	0.446
Insurance agencies, brokerages, and related activities	0.541	0.537	0.446
Business support services	0.406	0.415	0.444
News syndicates, libraries, archives and all other information services	0.361	0.394	0.438
Veterinary services	0.366	0.397	0.438
Support activities for agriculture and forestry	0.013	0.120	0.435
Travel arrangement and reservation services	0.323	0.327	0.434

Table 14: Top 2012

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Scenic and sightseeing transportation and support activities	0.318	0.325	0.434
Securities and commodity contracts intermediation and brokerage	0.353	0.390	0.434
Other financial investment activities	0.352	0.390	0.434
Automotive equipment rental and leasing	0.418	0.422	0.430
Air transportation	0.317	0.324	0.425
Federal electric utilities	0.544	0.490	0.422
Insurance carriers, except direct life	0.543	0.489	0.421
Management of companies and enterprises	0.315	0.323	0.421
Personal and household goods repair and maintenance	0.538	0.485	0.417
Advertising, public relations, and related services	0.537	0.484	0.416
Monetary authorities and depository credit intermediation	0.535	0.482	0.414
State and local government (hospitals and health services)	0.323	0.327	0.413
Other state and local government enterprises	0.322	0.327	0.412
Pipeline transportation	0.322	0.327	0.412
Couriers and messengers	0.320	0.326	0.411
Legal services	0.308	0.319	0.410
Photographic services	0.319	0.325	0.410
Automotive repair and maintenance (including car washes)	0.306	0.318	0.410
Warehousing and storage	0.319	0.325	0.410
Transit and ground passenger transportation	0.317	0.324	0.408
Truck transportation	0.317	0.324	0.408
Architectural, engineering, and related services	0.309	0.320	0.408
Dry-cleaning and laundry services	0.315	0.323	0.406
State and local government (other services)	0.314	0.323	0.406
Natural gas distribution	0.309	0.320	0.401
Postal service	0.307	0.319	0.399
Office administrative services	0.356	0.392	0.373
Scientific research and development services	0.320	0.326	0.371

Table 15: Middle 2012

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Custom computer programming services	0.314	0.323	0.366
Employment services	0.322	0.327	0.350
Other computer related services, including facilities management	0.321	0.326	0.350
Facilities support services	0.320	0.326	0.349
Specialized design services	0.318	0.325	0.348
Software publishers	0.318	0.325	0.348
Computer systems design services	0.316	0.324	0.347
Internet publishing and broadcasting and Web search portals	0.316	0.323	0.347
Accommodation	0.309	0.320	0.343
All other miscellaneous professional, scientific, and technical services	0.308	0.319	0.342
Accounting, tax preparation, bookkeeping, and payroll services	0.306	0.318	0.341
Satellite, telecommunications resellers, and all other telecommunications	0.305	0.318	0.341
Waste management and remediation services	0.302	0.316	0.339
State and local government passenger transit	0.302	0.316	0.339
Other ambulatory health care services	0.302	0.316	0.339
Residential maintenance and repair	0.293	0.310	0.334
Limited-service restaurants	0.289	0.308	0.332
Other amusement and recreation industries	0.289	0.307	0.331
Other durable goods merchant wholesalers	0.280	0.302	0.331
State and local government (educational services)	0.288	0.307	0.331
Directory, mailing list, and other publishers	0.280	0.301	0.325
Other federal government enterprises	0.267	0.291	0.317
Management consulting services	0.248	0.275	0.301
Grocery and related product wholesalers	0.369	0.351	0.132
Other support activities for mining	0.391	0.083	0.123
Wholesale electronic markets and agents and brokers			
Federal general government (nondefense)			
Civic, social, professional, and similar organizations			

Table 16: Bottom 2012

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Full-service restaurants	0.739	0.705	0.619
Commercial and industrial machinery and equipment repair and maintenance	0.843	0.708	0.610
Federal electric utilities	0.660	0.647	0.570
Insurance carriers, except direct life	0.653	0.642	0.566
Monetary authorities and depository credit intermediation	0.623	0.626	0.553
Advertising, public relations, and related services	0.621	0.625	0.553
Office administrative services	0.601	0.601	0.545
Insurance agencies, brokerages, and related activities	0.650	0.640	0.540
News syndicates, libraries, archives and all other information services	0.350	0.525	0.534
Air transportation	0.526	0.575	0.531
Software publishers	0.528	0.578	0.531
Automotive repair and maintenance (including car washes)	0.524	0.572	0.531
Photographic services	0.472	0.484	0.530
Securities and commodity contracts intermediation and brokerage	0.350	0.525	0.520
Other financial investment activities	0.350	0.524	0.520
General and consumer goods rental	0.559	0.518	0.516
Business support services	0.548	0.515	0.515
Other educational services	0.349	0.514	0.515
Spectator sports	0.551	0.514	0.514
Independent artists, writers, and performers	0.543	0.511	0.513
Services to buildings and dwellings	0.512	0.498	0.505
Performing arts companies	0.493	0.484	0.498
Other computer related services, including facilities management	0.472	0.484	0.491
Water transportation	0.472	0.484	0.491
Support activities for agriculture and forestry	0.472	0.484	0.491
Investigation and security services	0.471	0.484	0.491
Promoters of performing arts and sports and agents for public figures	0.471	0.484	0.491
Other support services	0.470	0.483	0.490

Table 17: Top 2007

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Funds, trusts, and other financial vehicles	0.470	0.483	0.490
State and local government electric utilities	0.469	0.483	0.490
Architectural, engineering, and related services	0.467	0.482	0.488
Postal service	0.467	0.482	0.488
State and local government passenger transit	0.467	0.482	0.488
Custom computer programming services	0.467	0.482	0.488
Other ambulatory health care services	0.467	0.482	0.488
Satellite, telecommunications resellers, and all other telecommunications	0.468	0.482	0.488
Residential maintenance and repair	0.466	0.481	0.488
Legal services	0.466	0.481	0.488
All other miscellaneous professional, scientific, and technical services	0.466	0.481	0.488
Accounting, tax preparation, bookkeeping, and payroll services	0.466	0.481	0.488
Accommodation	0.469	0.483	0.488
Waste management and remediation services	0.465	0.480	0.488
Management of companies and enterprises	0.469	0.483	0.488
Dry-cleaning and laundry services	0.470	0.483	0.488
Internet publishing and broadcasting and Web search portals	0.471	0.484	0.488
Computer systems design services	0.471	0.484	0.487
Scientific research and development services	0.471	0.484	0.487
Truck transportation	0.471	0.484	0.487
Transit and ground passenger transportation	0.471	0.484	0.487
Other amusement and recreation industries	0.462	0.478	0.487
Specialized design services	0.471	0.484	0.487
Warehousing and storage	0.472	0.484	0.487
Facilities support services	0.472	0.484	0.487
Personal and household goods repair and maintenance	0.472	0.484	0.487
Couriers and messengers	0.472	0.484	0.487
Other personal services	0.472	0.484	0.486

Table 18: Middle 2007

CommodityDescription	PCI k=2	PCI k=4	PCI k=9
Limited-service restaurants	0.460	0.476	0.486
Employment services	0.472	0.484	0.486
Veterinary services	0.473	0.484	0.486
Pipeline transportation	0.473	0.484	0.486
Other state and local government enterprises	0.473	0.484	0.486
Travel arrangement and reservation services	0.473	0.484	0.486
State and local government (educational services)	0.459	0.476	0.486
State and local government (hospitals and health services)	0.473	0.484	0.486
Other durable goods merchant wholesalers	0.446	0.466	0.486
Directory, mailing list, and other publishers	0.457	0.474	0.485
Other federal government enterprises	0.449	0.467	0.480
Periodical Publishers	0.520	0.466	0.478
Automotive equipment rental and leasing	0.494	0.455	0.476
Management consulting services	0.438	0.455	0.470
Book publishers	0.434	0.421	0.440
Data processing, hosting, and related services	0.427	0.415	0.434
Radio and television broadcasting	0.425	0.414	0.433
Natural gas distribution	0.468	0.482	0.426
Wireless telecommunications carriers (except satellite)	0.501	0.459	0.416
Wired telecommunications carriers	0.498	0.457	0.415
Scenic and sightseeing transportation and support activities	0.472	0.484	0.414
State and local government (other services)	0.469	0.324	0.410
Environmental and other technical consulting services	0.381	0.379	0.395
Grocery and related product wholesalers	0.187	0.267	0.083
Other support activities for mining	0.185	-0.003	-0.052
Federal general government (nondefense)			
Wholesale electronic markets and agents and brokers			
Civic, social, professional, and similar organizations			

Table 19: Bottom 2007

Annex: regressions

Table 20: Panel regressions: Real Growth and ECI

	(1) growth	(2) growth	(3) growth
rgpd pc t-1	0.642*** (0.103)	0.557*** (0.130)	0.580*** (0.140)
eci of goods t-1	0.0195* (0.0112)		0.0137 (0.0133)
Pers income pc real t-1(log)	-0.827*** (0.240)	-0.723*** (0.258)	-0.779*** (0.280)
PCE real t-1(log) pc	-0.450 (0.456)	-0.740* (0.414)	-0.664* (0.392)
PCE services real t-1(log) pc	0.374 (0.413)	0.634* (0.366)	0.598* (0.348)
Population t-1(log)	-0.0131 (0.0105)	-0.00588 (0.0102)	-0.0117 (0.00959)
eci of services t-1		0.0112** (0.00491)	0.00940 (0.00578)
Constant	10.89*** (3.289)	9.418*** (3.387)	10.30*** (3.789)
Observations	100	100	100

Standard errors in parentheses

Fixed effects: Yes (by state)

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Panel regressions: Real Growth and ECI

	(1) growth	(2) growth	(3) growth
rgpd pc t-1	0.572***	0.570***	0.585***
	(0.137)	(0.122)	(0.131)
eci of goods t-1	0.0112		0.0130
	(0.0136)		(0.0139)
Pers income pc real t-1	-0.0000146***	-0.0000144***	-0.0000153***
	(0.00000563)	(0.00000521)	(0.00000552)
PCE t-1 pc	-25.32**	-30.48***	-28.86***
	(10.43)	(11.80)	(10.95)
PCE real serv t-1 pc	33.56**	39.49**	38.58***
	(14.29)	(15.44)	(14.85)
Population t-1	-2.37e-09**	-1.81e-09*	-2.19e-09**
	(9.80e-10)	(1.06e-09)	(1.04e-09)
year=2	0	0	0
	(.)	(.)	(.)
year=3	-0.0317**	-0.0822**	-0.0854**
	(0.0123)	(0.0333)	(0.0333)
eci of services t-1		-0.0208	-0.0236
		(0.0144)	(0.0148)
Constant	2.522***	2.574***	2.631***
	(0.572)	(0.535)	(0.560)
Observations	100	100	100

Standard errors in parentheses

Fixed effects: Yes (by state)

,

Robust standard errors in parentheses

*** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$**

Table 22: Panel regressions: Real Growth and ECI

	(1) growth	(2) growth	(3) growth
rgpd pc t-1	0.575***	0.580***	0.597***
	(0.140)	(0.125)	(0.134)
eci of goods t-1	0.0132		0.0147
	(0.0132)		(0.0134)
Pers income pc real t-1(log)	-0.790***	-0.786***	-0.838***
	(0.277)	(0.252)	(0.270)
PCE real t-1(log) pc	-0.773**	-0.970**	-0.900**
	(0.380)	(0.409)	(0.403)
PCE services real t-1(log) pc	0.711**	0.869**	0.839**
	(0.330)	(0.357)	(0.351)
Population t-1(log)	-0.0104	-0.00221	-0.00855
	(0.00953)	(0.0111)	(0.0103)
year=2	0	0	0
	(.)	(.)	(.)
year=3	-0.0260*	-0.0774**	-0.0777**
	(0.0135)	(0.0329)	(0.0339)
eci of services t-1		-0.0211	-0.0229
		(0.0143)	(0.0145)
Constant	10.45***	10.26***	11.09***
	(3.735)	(3.300)	(3.624)
Observations	100	100	100

Standard errors in parentheses

Fixed effects: Yes (by state)

Robust standard errors in parentheses

*** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$**

documentos de trabajo

Ciencias Sociales
Universidad de la República
URUGUAY

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Facultad de Ciencias Sociales
Universidad de la República
Constituyente 1502 - 2410-6720
comunicacion@cienciassociales.edu.uy
www.cienciassociales.edu.uy