N.º 50
FEBRERO DE 2023

DOCUMENTOS
DEPARTAMENTO DE ECONOMÍA
UNIVERSIDAD DEL NORTE

Characterization of Colombian industries under Schumpeter's patterns of innovation

UNIVERSIDAD DEL NORTE
Vigilada Mineducación

Juan José Taborda Núñez
Jana Schmutzler

Serie Documentos n.º 50 • Febrero de 2023 • ISSN 0121-2346

Characterization of Colombian industries under Schumpeter's patterns of innovation

Juan José Taborda Nuñez
Jana Schmutzler

Serie Documentos, 50
Febrero de 2023

La serie *Documentos* del Departamento de Economía de la Universidad del Norte circula con el fin de difundir y promover las investigaciones realizadas en Uninorte, y también aquel resultado de la colaboración con académicos e investigadores vinculados a otras instituciones. Los artículos no han sido evaluados por pares, ni están sujetos a ningún tipo de evaluación formal por parte del equipo editorial.

UNIVERSIDAD DEL NORTE

Vigilada Mineducación

Universidad del Norte
Departamento de Economía
Apartado aéreo 1569
Barranquilla, Colombia

Characterization of Colombian industries under Schumpeter's patterns of innovation

by

Juan José Taborda-Núñez[1]

Jana Schmutzler[2]

Abstract

The literature has continuously examined the relationship between market structure and innovation. In particular, Joseph Schumpeter's Mark I and Mark II innovation patterns inquired on what type of firm drives innovation based on competitive structures. Schumpeter marks, together with innovation systems theory have been widely used to classify productive sectors across the world. This article aims to deploy an exercise of this nature for Colombia in two steps. First, exploit innovation-related metrics, like stability of innovation, technological opportunities and market concentration, through an unsupervised learning method. Second, take advantage of geographical data to map and compare the sectors of each group. Statistical analysis of the cluster reveals that each group's features align with Mark I and II archetypes. Moreover, geographical information shows patterns of agglomeration similar to those previously explored in the literature. Policy implications are discussed afterwards, concluding that the need of heterogeneous policy designs that tackle specific regional features is paramount for the country, and a potential Colombian policy instrument to channel these findings is suggested.

JEL Codes: C38, L60, O30

Author Key words: radical innovations, incremental innovations, market structure, k-means clustering, innovation systems.

[1]Universidad del Norte (Barranquilla, Colombia). Email: jtabordaj@uninorte.edu.co

[2]Business School, Universidad del Norte (Barranquilla, Colombia). Jackstädt Center of Entrepreneurship and Innovation Research, Bergische Universität Wuppertal (Wuppertal, Germany). Email: schmutzler@uninorte.edu.co

Acknowledgements

The idea for this project came from the Economics Of Innovation lecture, taught by Prof. Dr. Werner Bönte at the University of Wuppertal, where both Mark I and II were mentioned, together with several conceptual elements about innovation. Writing, modelling and presenting this project has been a result of all the teachings from the human talent within the Department of Economics at Universidad del Norte, to which I am indebted. I must give special recognition to Prof. Juan Perilla, Prof. Jose Ramos, and Prof. Jana Schmutzler for their mentorship during the ellaboration and presentation of my thesis.

Since this project is an improved version of my bachelor thesis, it has **not** received funding of any type.

Agradecimientos

La idea de este proyecto surgió del material de la clase "Economics of Innovation", dictada por el Prof. Dr. Werner Bönte en la Universidad de Wuppertal, donde se mencionaron tanto el Mark I como el II, junto con varios elementos conceptuales sobre la innovación. Escribir, modelar y presentar este proyecto ha sido el resultado de todas las enseñanzas del talento humano dentro del Departamento de Economía de la Universidad del Norte, con el cual estoy en deuda. Debo agradecer a los profesores Juan Perilla, José Ramos y Jana Schmutzler por su tutoría durante la elaboración y presentación de mi tesis.

Dado que este proyecto es una versión mejorada de mi tesis de pregrado, no ha recibido financiación de ningún tipo

1. Introduction

Who drives innovation within an industry? According to mainstream economics, a variety of factors influence on whether it is a small or established firm, among these, market structure, entry barriers, specialization, and patent legislation. The interaction of these factors is complex and multifaceted, and has spurred concepts like innovation systems, as they can concentrate or disperse efforts, disincentivize novel ideas in favor of established paradigms, place industries at the source of a supply chain, motivate cooperation, or foster the appropriation of gains from innovation.

The question of who drives innovation is both challenging and important, particularly for emerging countries seeking to develop. In this article, we will explore this by examining Joseph Schumpeter's Mark I and Mark II innovation patterns (Schumpeter, 1911; Schumpeter, 1942). Mark I contends that small firms are the primary drivers of innovation, while Mark II suggests that such role falls in large corporations. By examining these archetypes, we can gain a better understanding of the factors that contribute to innovation within different industries, laying the groundwork for potential policy implications.

According to Fontana et al. (2021), Mark I and Mark II innovation patterns have untapped potential for characterizing innovative activities within industries, and existing characterizations have been found to "*stand the test of time quite well*" (p. 19). Various studies have used these patterns to characterize industries at a national level, offering insights into innovation activities and providing a platform for innovation-related policy making (Breschi et al., 2000; Castellaci & Zheng, 2010; Malerba & Orsenigo, 1996).

However, these authors focus on developed economies, leaving aside characterization exercises in countries outside the global north, such as Colombia. On the other hand, while some studies have made cross-country comparisons (Malerba & Orsenigo, 1996; Breschi et al., 2000), regional differences have not been thoroughly explored. The role of regions in innovation has been recently acknowledged (Malerba & Breschi, 1997), and literature on the field has been emerging in the past years, pointing out that factors such as the type of knowledge, cultural diversity, wealth, institutional and geographical proximities shape much of the innovation process of a region (Gössling & Rutten, 2007; Boschma, 2005; Asheim et al., 2011). Such set of interactions has close similarities with Nelson's (1993) concept of Innovation Systems.

The limitations of existing literature place Colombia as a potential candidate for a study of this type. In one hand, as an emerging economy, it can add value to the understanding of the country's industrial landscape, specially in manufacturing, one of Colombia's most important industries, accounting for 10% of its Gross Domestic Product (GDP) and more than one-third of the economy's aggregated value chain (Arbelaez et al., 2021). On the other hand, the cultural influences of Colombia, reflected in strongly distinct features across regions (Melo, 2017), motivate a regional analysis of its relationship towards industrial, and more specifically, innovative activities.

That said, this document will utilize data from Colombia's statistic department, known as the *Departamento Administrativo Nacional de Estadística* (DANE). Specifically, we will utilize DANE's (2020) *Encuesta de Desarrollo e Innovación Tecnológica* (EDIT) survey and DANE's (2019) *Encuesta Anual Manufacturera* (EAM) survey for manufacturers to characterize industries as Mark I or Mark II according to their International Standard Industrial Classification code (ISIC, or CIIU in Spanish), with a focus on Section C: Manufacturing. Furthermore, borrowing the idea from past literature, we will exploit geographical information to provide additional insights, but adding novelty with a focus across Colombia's regions instead of cross-country comparisons. Finally, concepts from innovation systems will prove useful in the concluding parts of this work, as they have close relationship with the topic.

This article will proceed in the following order: First, an examination of existing literature to provide background and identify key theoretical elements. Second, a methodology outline for processing and analyzing the data obtained from DANE's surveys. Third, presentation of quantitative results, highlighting significant findings and trends. Finally, concluding with potential policy implications and how the results can inform future innovation-related policies.

2. Relevant background literature

The concept of innovation

To establish a clear understanding of the concept of innovation and innovative activities, it is important to acknowledge that these concepts are contested in the literature. For this research, we will adopt the definition provided by the Organization for Economic Co-operation and Development (OECD) OSLO Manual (2018). According to the OECD, innovation is a "*New or improved product or process (or a combination thereof) that significantly differs from the unit's previous products or processes and has been made available to potential users (product) or brought into use by the unit (process)*" (p. 20).

The OSLO Manual acknowledges the subjective nature of innovation, which is the subject of an ongoing literature debate. However, in practice, innovation can be considered objective if we focus on shared motives such as novelty, utility, and significant differences. Accordingly, the manual defines innovative activities as "*all developmental, financial, and commercial activities undertaken by a firm that is intended to result in an innovation for the firm*" (p. 20). Although the concept is firm-centric, it can be applied to markets and industries as well.

Two taxonomies of innovation are relevant to this study. Schumpeter (1942) emphasizes on radical innovations, which depart from known opportunities, create novelty, and destroy something old. This creates the concept of "creative destruction." Conversely, Kirzner (1973) emphasizes on incremental innovations, arguing that human alertness to new opportunities yields more profits when directed toward enhancements that do not disrupt the *status quo* and operate within the realm of known opportunities.

Schumpeter, markets and innovation

Market structure is a crucial factor for this study, and the literature has explored its impact on innovative activities. Loury (1979) found that innovation incentives can reach optimal levels in a market with limited competitive firms, even if the number of firms falls below the traditional "entry until profits are zero" rule. The author also contends that atomistic competition is the ideal market structure for innovation.

Mansfield (1963) identified a correlation between firm size, market concentration, and the share of innovative activities. The author holds three propositions: innovation costs are high and only larger firms can afford them, innovation requires a large scale for success and failure to balance out, and firms need sufficient market control to reap the benefits of their innovations. As a result, large firms tend to carry out a disproportionate share of innovations.

Raider's (1998) research suggests that a combination of market structure and adversity drives innovation. The author found that firms facing high pressure from market competition tend to innovate more, while markets without this phenomenon tend to innovate less. Raider's analysis incorporates a network market approach. Here, constraints imposed by buyers and sellers, together with extreme upstream and downstream competition, can determine of innovation.

Before delving deeper into the relationship between innovation taxonomies and market structure, it is worth reviewing Schumpeter's (1911; 1942) pioneering work on differentiated innovation archetypes. In his book, *The Theory of Economic Development* (1911), Schumpeter uses a metaphor to encapsulate his early thoughts on new combinations and who carries them: "*new combinations are, as a rule, embodied, as it were, in new firms* *which generally do not arise out of the old ones but start producing beside them; (...) in general, it is not the owner of stagecoaches who builds railways*" (p. 66). This idea would be coined as Schumpeter's Mark I, which posits that entrant firms are the drivers of innovation.

Schumpeter's ideas evolved in his 1942 book "*Capitalism, Socialism and Democracy,*" where he reflected on previous remarks, contrasting them with the continuous standard of living improvements during the era of unrestrained "*big business.*" Upon examining this phenomenon, Schumpeter discovered that recent productive improvements were largely attributable to large firms: "*(...) a shocking suspicion dawns upon us that big businesses may have had more to do with creating that standard of life than with keeping it down*" (p. 82). This perspective departs from Schumpeter's Mark I and becomes Mark II. Here, Schumpeter appoints tried and true, large corporations as the drivers of innovation.

Malerba & Orsenigo (1996) use the terms "*widening pattern*" and "*deepening pattern*" to describe Schumpeter's Mark I and Mark II, respectively. The former is characterized by fierce competition, low entry barriers, low appropriability, less technological opportunities and a large population of innovators, while the latter is defined by stable environments with large firms, high entry barriers, more technological opportunities, high appropriability and fewer players. Although these descriptions provide a useful starting point, we can further examine this relationship.

Two propositions can help us understand the relationship between innovation and market structure. Firstly, according to Gilbert (2006), the incentive for a firm to innovate is the difference in profits it can earn by doing so. Secondly, Arrow (1962) argues that innovation incentives are higher in competitive markets since existing monopoly power discourages new ideas, as the incumbent firm already profits from its operations. Baumol (2004) further supports this view, stating that "*major breakthroughs have tended to come from small new enterprises*" (p. 10). Thus, risky innovations that seek to disrupt markets are Mark I related.

Baumol (2004) continues: "*(. . .) while the invaluable incremental contributions that multiply capacity and speed, and increase reliability and user-friendliness have been the domain of the larger firms*" (p. 10). This observation helps us define the market structure that is conducive to incremental innovations. Considering Arrow's replacement effect and Gilbert's statement on innovation incentives, we can argue that incremental innovations suit concentrated markets. While it is true that entry barriers and high resource endowments discourage new entrants, when it comes to innovation, monopolies have less incentives as it risks their existing profits. Thus, incremental innovation align with Mark II.

While it may look like both archetypes are opposites, Shapiro (2012) notes that they share underlying principles: both consider contestable markets and potential profits as drivers of innovation. In summary, the literature suggests that Mark I relates to dispersed markets, where small firms lead radical innovations, resulting in a highly competitive, turbulent, and unstable environment. In contrast, Mark II archetypes are observed in concentrated market structures, such as monopolies or oligopolies, where large firms focus on incremental innovations, resulting in a stable but less competitive market. In such markets, new ideas are disincentivized as they pose a threat to the dominant position of incumbents.

The state of characterization exercises

Based on this, we can pinpoint a first prominent Schumpeterian characterization exercise in Malerba & Orsenigo (1996) article. There, Schumpeterian archetypes were identified across 49 technological classes of six countries: the US, Japan, Germany, France, the UK, and Italy. The study found that innovative activities vary across different technological classes. Moreover, innovation patterns within each technological class were similar in cross-country comparison, suggesting that knowledge-base features play a significant role. It is also important to remember that these were the authors that suggested the deepening and widening patterns exposed earlier.

Breschi et al. (2000) provide further evidence that aligns with earlier findings by Malerba & Orsenigo (1996), suggesting that patterns of innovative activities are similar across countries in the same technological class. Utilizing patent data from the United Kingdom, Germany, and France, as well as measures built from the *Policy, Appropriability and Competitiveness for European Enterprises* (PACE) questionnaire, the study reveals that Mark II is associated with high degrees of cumulativeness and appropriability, while Mark I is associated with low degrees of cumulativeness and appropriability. What is more, the authors suggest that a pattern of innovation is the result of multiple technological regimes, which at the same time, are the result of particular combinations of technological opportunities, appropriability of innovations, cumulativeness of technical advances, and properties of the knowledge base.

Fontana et al. (2012) builds upon the work of Breschi et al. (2000) by utilizing similar measures in their econometric exercise. Their findings maintain Mark I as the disperse and turbulent archetype, while Mark II is characterized as stable and concentrated. Landström & Schön (2010) make further distinction between these two, while noting some similarities between them. Notably, both patterns highlight the importance of innovation in economic development, and both require the capitalist to assume risk. This highlights the role of risk-taking in driving innovation.

Schumpeterian archetypes are not uncontested in the literature. Pavitt's (1984) taxonomy offers a classification based on the requirements of its users, the sources of its technologies, and the degree of appropriability in four different patterns: Supplier-dominated sectors; scale-intensive industries employing process and product innovation; specialized suppliers that market technology to other firms; finally, science-based, knowledge industries with a high degree of appropriability and tailored towards exploring innovative technological breakthroughs.

Archibugi (2001) has identified close links between Pavitt's taxonomy and Kondratiev's long waves of capitalist development. Castellacci (2008), in contrast, connects it to the sectors that fueled the growth of advanced economies during the Fordist era. Schumpeter's Mark I and II, on the other hand, are anchored in the dynamics of an industry life cycle. Mark I characterizes the early stage of an industry, while Mark II refers to a more mature, late-stage phase (Malerba, 2005, as found in Maleki & Rosiello, 2014).

The selection of either Schumpeter's or Pavitt's framework depends on the scope and objectives of the research. Within the literature, Fontana et al. (2021), Breschi et al. (2000), Castellaci & Zheng (2010), Malerba & Orsenigo (1996), and Corrocher et al. (2007) have focused on Schumpeter's patterns. Breschi et al. (2000) drew upon elements from Malerba & Orsenigo (1996) but narrowed their study to three countries and omitted control variables for the effects of a country's national innovation system. Corrocher et al. (2007) aimed to distinguish sectors focused on information and communications technologies (ICT), while Castellaci & Zheng (2010) focused their efforts on industry-level productivity characterizations.

In contrast, authors like Van Dijk (2000), Marsili & Verspagen (2002), and Leiponen & Drejer (2007) advocated for an approach utilizing Pavitt's taxonomy. Van Dijk (2000) demonstrated that certain groups within Dutch manufacturing do not fit neatly into either of Schumpeter's archetypes. Similarly, Leiponen & Drejer (2007) observed a comparable phenomenon within Danish industries. Meanwhile, Marsili & Verspagen (2002) refined Pavitt's taxonomy by introducing a new classification proposed by one of the authors.

Focusing on those using Schumpeterian archetypes, they have developed measures to quantify them. For example, Malerba & Orsenigo (1996) approach, and eventually, Breschi et al. (2000) work tested variables such as concentration, entry, firm size, hierarchy of innovators and turbulence of markets. Malerba & Orsenigo (1996) approach also explored a relation between appropriability and technological opportunities. On the other hand, Maleki et al. (2018) used patenting growth rate as a measure for technological opportunities. Similarly, Fontana et al. (2021) inquires into technological opportunities, but also on accumulativeness and appropriability of knowledge. Finally, Castellaci & Zheng (2010), productivity analysis approach focuses on appropriability and technological opportunities, discriminating by mode and source.[3]

Innovation systems and the role of regions

When multiple countries are involved, such as Germany, the UK, France, the US, and others, authors recognize specific country differences. Malerba & Orsenigo (1996) acknowledged the historical evolution and country-specific effects of national firms, using binary variables to capture their causal effect in their study. Similarly, Breschi et al. (2000) employed a similar method to capture country-specific effects. The question arises, how can we explain these differences? This is where innovation systems play a role.

Innovation systems (IS) encompass a set of interactions that facilitate, generate, transform, and disseminate knowledge within a specific territory. Nelson (1993) was among the first authors to introduce this concept, coined as National Innovation Systems (NIS). Malerba (2002, 2003, 2005) expanded it to Sectoral Innovation Systems (SIS), while Asheim & Gertler (2006) examined Regional Innovation Systems (RIS), a notion that was later reinforced by authors such as D'Allura et al. (2012).

Innovation systems are intertwined with institutional, urban, and spatial economics. Firms tend to locate near sources of labor, financing, and raw materials, and in areas with favorable transport conditions. Non-tangible interactions like economies of scale, institutional coverage, and spillover effects also play a role. Schrempf et al. (2012) identify two main interactions within an IS: between firms and government offices, and between firms and higher education institutions. Policy incentives, regional governance structures, and informal institutions like culture also affect industry localization.

Given the research's scope, it is pertinent to detail the role of regions and innovation. Breschi & Malerba (1997) made relevant contributions to this topic, pointing out that the features of a technological regime have varying impacts on firms located in different geographical areas and are not uniformly distributed. Technological regimes (and the subsequent patterns it creates) impact the environment around them through mechanisms like modifying the source of opportunities for innovation, increasing the concentration of innovators, defining the means of knowledge transmission, and the extent of positive

[3] We shall find appropriability modes by the names of "conventional" and "non-conventional" protection mechanisms later on in this article.

knowledge externalities.

Breschi & Malerba (1997) summarize their findings in two core ideas. One is that innovators are geographically concentrated when there are conditions of high technological opportunity, high appropriability, and high firm accumulativeness, fitting the Mark II archetype. The other idea is that innovators are geographically dispersed when there is low opportunity, low appropriability, and low firm accumulativeness, something that fits the Mark I innovation archetype.

The case of Colombia

Finally, making an emphasis on Colombia, Schumpeter's theory has been applied in some articles (Umaña-Aponte et al., 2013; Marroquín, 2010; Arroyo-Mina & Guerrero, 2018; Langebaek-Rueda & Vásquez, 2007). However, is not used to characterize industries. Although there have been attempts to develop sector-specific entrepreneurship and innovative profiles (Cerón et al. 2010; Ovallos-Gazabón & Amar-Sepúlveda, 2014), they do not employ Schumpeterian patterns of innovation.

Colombia is considered a peripheral economy, and according to the Dependence Theory (Prebisch, 1950) and empirical evidence (Arezki et al., 2013), markets in peripheral economies function differently than those in developed and core economies. In peripheral economies, low-value goods flow outwards and high-value goods flow inwards through imports, while in developed and core economies, the opposite occurs.

A problem may arise if we characterize peripheral economies under core economies frameworks. This is why Pavitt's Taxonomy may be non-suitable, as it inquires on industrial features proper of advanced economies, like science-based sectors and specialized high-tech. However, Schumpeterian archetypes do not have this problem. Both archetypes find what type of firm drives innovation through common market metrics like concentration, firm size, or stability of innovation. Moreover, per Malerba (2005) observation, it may shed light on the dynamic of an industry, whether it is in an early stage, or it is entering a late phase of maturity.

Characterizing peripheral economies using frameworks designed for core economies can present problems. Pavitt's Taxonomy, for instance, focuses on industrial features typical of advanced economies such as science-based sectors and specialized high-tech industries, unsuitable for peripheral economies. In contrast, Schumpeterian archetypes use common market metrics like concentration, firm size, and innovation stability. Additionally, as observed by Malerba (2005, as found in Maleki & Rosiello, 2014), Schumpeterian archetypes can provide insights into the dynamics of an industry, including its stage of maturity.

3. Methodology

For this study, we use a cross-section combination of DANE's (2020) EDIT survey and DANE's (2019) EAM survey, both conducted in 2018. The EAM survey is a national census of firms that meet the classification criterion of having at least ten employees and sales exceeding 517 million Colombian Pesos[4]. A "*Numero de Orden*" (NORDEMP) number is used to identify firms across surveys, simplifying the merging of databases.

In addition, EAM includes geographic data using DANE's "*DIVIPOLA*" classification, which assigns a unique number to each department in Colombia. Although this is not necessary for the initial cluster analysis, we will use it after to assign each firm within industries a geographic location. This will allow for geographically specific analyses.

The EDIT survey provides information on the innovative activities undertaken by firms and their impact on their industrial activities. As stated in DANE's (2020) methodological overview, the survey is based on a theoretical framework that largely follows the OECD's methodological guidelines, with a focus on the innovation-specific OSLO Manual. Annex 1 presents the universe of study for the EDIT survey at a four-digit ISIC level, based on DANE's Methodology (2018, p. 7)[5].

Since EDIT samples EAM sectors instead of individual firms, not all industries in EAM are included in the EDIT dataset. However, this can be resolved by performing an inner join of both databases in Rstats, using their ISIC code and NORDEMP registrations as keys. The resulting database has 6405 observations. Based on exposed background literature, we can quantify a Schumpeterian pattern of innovation using three dimensions: stability of innovation, technological opportunities, and market concentration. We will use similar variables from both surveys that correspond to these dimensions, as listed in the following table:

[4]Firms that meet the sales requirement but not the employee requirement are also included in the census

[5]As an exception, ISIC codes 202 and 210 select specific four-digit industries within them

Table 1: Relevant variables for the cluster

Dimension	Concept	Variable	Description
Stability	Amount of radical innovations	I1R4C2N	Goods and services that are new or different from those existing in the market, and were introduced in the 2017-2018 period.
	Amount of incremental innovations	I1R4C2M	Goods and services that are improvements to existing goods or services in the market, and were introduced in the 2017-2018 period.
Concentration	Total sales	I3R2C1	Income or operational sales, both local and foreign, perceived by the firm between 2017 and 2018. In thousands of Colombian Pesos.
	Total spending on innovative activities	II1R10C2	Total investment by the firm on innovative activities for the 2017-2018 period.
	Total employees	PERTOTAL	Permanent amount of employees within a firm. Includes owners, temporal employees and permanent roster.
	Total output	PRODBIND	Firm output accounting for the value of goods and services at the end of the production process, excluding intermediate goods. In thousands of Colombian Pesos.
Technological Opportunities	Possession of conventional protection mechanisms valid until 2018	VI1R8C2	Prior to the 2017-2018 period, and valid until the end of said year, how many conventional protection mechanisms were in possession of the firm. Ex: Patents, IP, Copyright, Trademarks.

Obtention of conventional protection mechanisms between 2017-2018	VI2R8C2	Between 2017 and 2018, how many conventional protection mechanisms did the firm acquire.
Usage of non-conventional protection mechanisms	VI3R5C2	Between 2017 and 2018, how many non-conventional protection mechanisms did the firm acquire. Ex: Non-Disclosure Agreement, industrial secrets, high complexity on design.

Source: Own elaboration based on DANE's (2019; 2020)

After running these variables in Rstats, three industry-level dimensions are constructed: Concentration (CON), Technological Opportunities (TO), and Stability (STA). Below, we provide a comprehensive explanation of each dimension along with its proposed way to measure it:

1 CON: The measurement technique utilized by CON is based on the research of Malerba and Orsenigo (1996) and utilizes the Herfindahl-Hirschman Index (HHI, or HH Index) to measure the concentration of innovative activities and firm size.

In addition to these conventional market concentration variables, CON also employs a variable representing investment in innovative activities and another measuring the number of employees in each firm, enabling analysis of whether labor for innovative activities is concentrated within one particular firm. To ensure uniformity between each index, variations are smoothed out through the use of a geometric mean, as follows:

$$CON = (HH_{ms} * HH_{msa} * HH_{lsd} * HH_{ss})^{1/4} \tag{1}$$

In the above equation, HH is a Herfindahl-Hirschman index, with each subindex representing a particular measure. Specifically, ms denotes Market Share, msa denotes Innovative Activities, lds denotes Labor Demand Share, and ss denotes Supply Share. Based on the literature, it is anticipated that low values of CON will be associated with Mark I, whereas high values will associate with Mark II.

2 *TO*: Constructed based on Maleki et al. (2018) approach, which employs patent growth rate. However, relying solely on patents is inadequate, as more mechanisms are used. Additionally, due to the incomparability of EDIT surveys, conducting a continuous growth rate analysis is not feasible. To overcome this challenge, a solution is devised by calculating the proportion of new protection mechanisms relative to registered mechanisms such as:

$$TO = \frac{PM_{1718} + NCPM_{1718}}{PM} \tag{2}$$

In the equation, PM_{1718} denotes the industry-wide aggregate of all protection mechanisms acquired between 2017 and 2018. Similarly, $NCPM_{1718}$ represents the same for non-conventional protection mechanisms such as Non-Disclosure Agreements, Industrial Secrets, and Complex Design. Lastly, PM encompasses all protection mechanisms that remain valid until the end of the year in question (2018).

Technological opportunities are closely connected with appropriability. A high value of technological opportunities indicates a high level of appropriability, whereas a low value suggests the opposite. Based on the literature, Mark II industries are likely to have high values, while Mark I industries tend to have lower amounts.

3 *STA*: Stability is closely related to the entry and exit of firms in the innovative market, necessitating a dynamic analysis. Due to the incomparability of EDIT surveys, conducting such an analysis is not possible. As an alternative, we will utilize the number of radical and incremental innovations introduced by firms to construct an industry-level measure. The literature suggests that radical and incremental innovations can provide insight into the stability of an industry. As Baumol (2004, p. 10) notes, "major breakthroughs have tended to come from small new enterprises, while the invaluable incremental contributions (...) have been the domain of larger firms". Hence, to quantify that stability approach, the following mathematical form is proposed:

$$STA = Sr - Si \tag{3}$$

In the previous equation, Sr is calculated as the ratio of the number of radical innovations to the total number of innovations, Si is calculated as the ratio of the number of incremental innovations to the total number of innovations. Following Baumol's observation, we expect that Mark I will have a high share of radical innovations, while Mark II will have the opposite.

The literature has suggested additional measures, such as entry and exit rates and changes in the ranking of top innovative firms. However, data availability issues arise when attempting to measure these elements in Colombia, as the data may be incomplete, incompatible, or unavailable. It is important to note that the NORDEMP registration is a fundamental basis for differentiating firms in this study. Therefore, data obtained from sources without this data column cannot be compared since there is no reference point to determine which information belongs to which firm.

To ensure the reliability of the concentration measure, additional filters were applied to the data. For instance, in cases where some industries reported zero spending on innovative activities, the concentration measure would result in zero, leading to unreliable results. This was particularly problematic for industries with less than 20 observations. Therefore, the minimum number of observations required per industry was set at 20. For a detailed overview of the industries included in the study and their respective results in each dimension, please refer to Annex 2.

The data is analyzed using unsupervised machine learning techniques, specifically principal component analysis (PCA) and k-means clustering. The Lloyd algorithm is used with 10 repetitions and two groups. As explained by MacKay (2003), k-means calculates a mean for each group and assigns industries to a group through the following mechanism:

1 *Assignation phase*: Assign each observation to the group with the closest mean (by Euclidean distance). Each generated group is composed by a centroid that serves as reference point for the following steps.

2 *Update phase*: Group parameters adjust to match the means of the data points.

3 *Repetition phase*: Assignation and Update phases repeat until cluster positions do not change anymore.

All information is standardized prior to the cluster calculation. This is made to prevent problems that may arise due to non-normalized euclidean distance, such as those noted by Martinez et. al. (1999):

1 Distance sensibility: To avoid overestimation by extreme values, it is important to standardize the Euclidean distance. In our dataset, which consists of i points, each point x_i has a coordinate μ_i. To standardize the data, we follow the definition of standardization:

$$\mu_i := \sigma \frac{x_i}{\sigma_i} \tag{4}$$

Where σ is an hypothetical measurement of standardization. If we consider that equation (4) normalized a data point by its standard deviation, then we have to hold true that the variance of the normalized dataset will be 1, as required by the properties of standard distributions. This stems from the variance property:

$$\begin{aligned} Var(\alpha X) &= Var(X)\alpha^2 \\ Var(\frac{1}{\sigma}X) &= \frac{Var(X)}{\sigma^2} \\ Var(\frac{X}{\sigma}) &= \frac{\sigma^2}{\sigma^2} \\ &= 1 \end{aligned} \tag{5}$$

With this in mind, the formulae of an Euclid distance d between two vectors of points q and p:

$$d = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{6}$$

Will adopt the form

$$d = \sqrt{\sum_{i=1}^{n}(\sigma\frac{x_{iq}}{\sigma_{iq}} - \sigma\frac{x_{ip}}{\sigma_{ip}})^2} \tag{7}$$

The resulting distance will be a d value of σ standard units. Thus, an initial standardization of the data yields a normalized Euclid distance. In other words, if we compute Euclidean distances on hypothetical u' normalized coordinate vectors of our data, the resulting distances will be measured in units of standard deviations. This solves the distance sensibility problem.

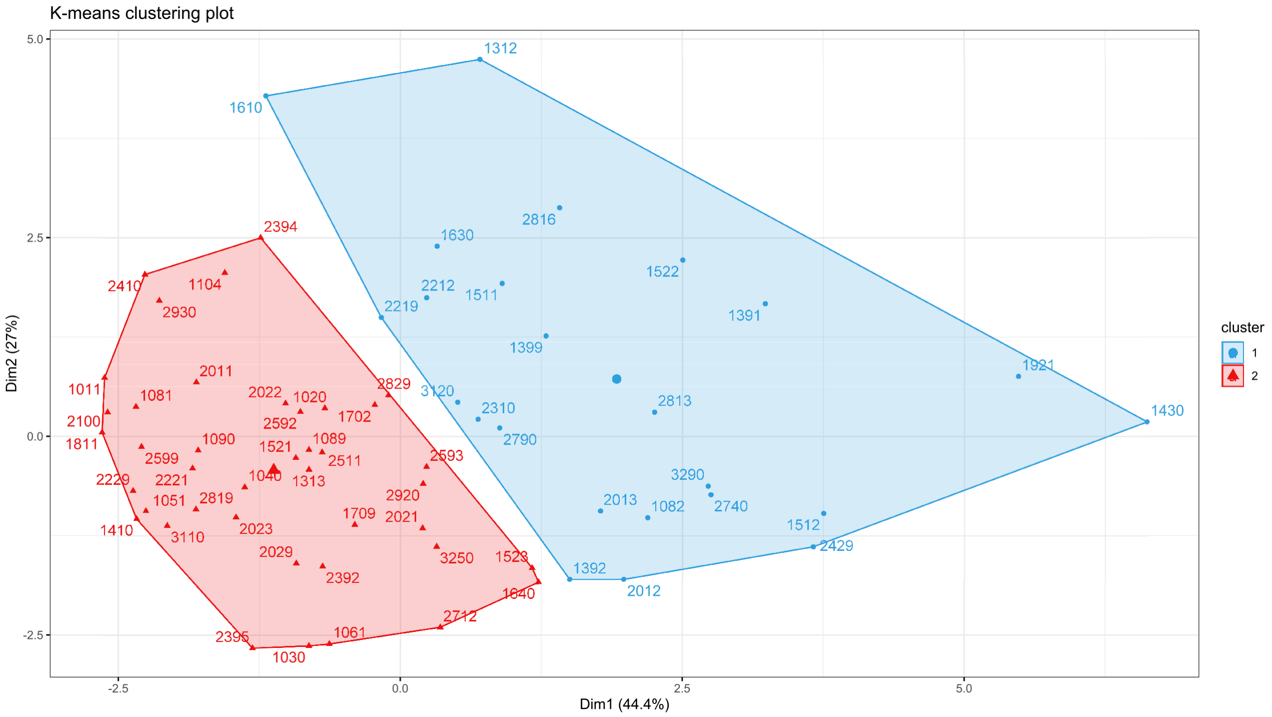
2 Redundancy by correlation: If variables are similar in construction, correlation may provide redundancy, and thus, overestimation of the distances. However, a previous calculation of Pearson correlation coefficients between TO and STA, TO and CON, and CON with STA, yielded very low $\rho = 0.12$, $\rho = 0.08$ and $\rho = -0.17$ values respectively. Thus, redundancy is low and the problem is not present.

4. Results

4.1. The Cluster

The figure below displays the cluster after employing both learning methods.First, the PCA dimensional reduction algorithm is employed with our three dimensions (*STA, TO, CON*) to reduce the data to a traditional two-axis form, capturing 44.4% and 27% of the variation in the original dataset with shadow variables, *Dim1* and *Dim2*. Then, k-means clustering was employed and the results were plotted.

Figure 1: Resulting k-means cluster for Colombian manufacture



Source: Own elaboration based on DANE (2019;2020) databases.

Cluster results show two groups. Cluster Group 2 (CG2), coloured red, has 5192 observations, while Cluster Group 1 (CG1), coloured blue, has 794 observations. Initial descriptive statistics for both groups are shown in the following table.

Table 2: Initial descriptive statistics of the two groups

	Cluster Group 2			*Cluster Group 1*		
	TO	*CON*	*STA*	*TO*	*CON*	*STA*
Max	3.500	1023.11	0.586	6.000	2702.96	1.000
Min	0	160.28	-1.000	0.047	1068.11	-1.000
Mean	0.522	572.22	-0.294	0.983	1507.72	-0.345
Std Dev	0.651	262.98	0.414	1.496	416.08	0.550

Source: Own elaboration

Regarding CG2, technological opportunities exhibit a wide range of values, with a representative value of 0.52 and a standard deviation of 0.65. Market concentration values in CG2 tend to be smaller, ranging from 160 to 1023 on the HHI, with a representative value of 1507. The standard deviation across concentration stands at 262 index points.

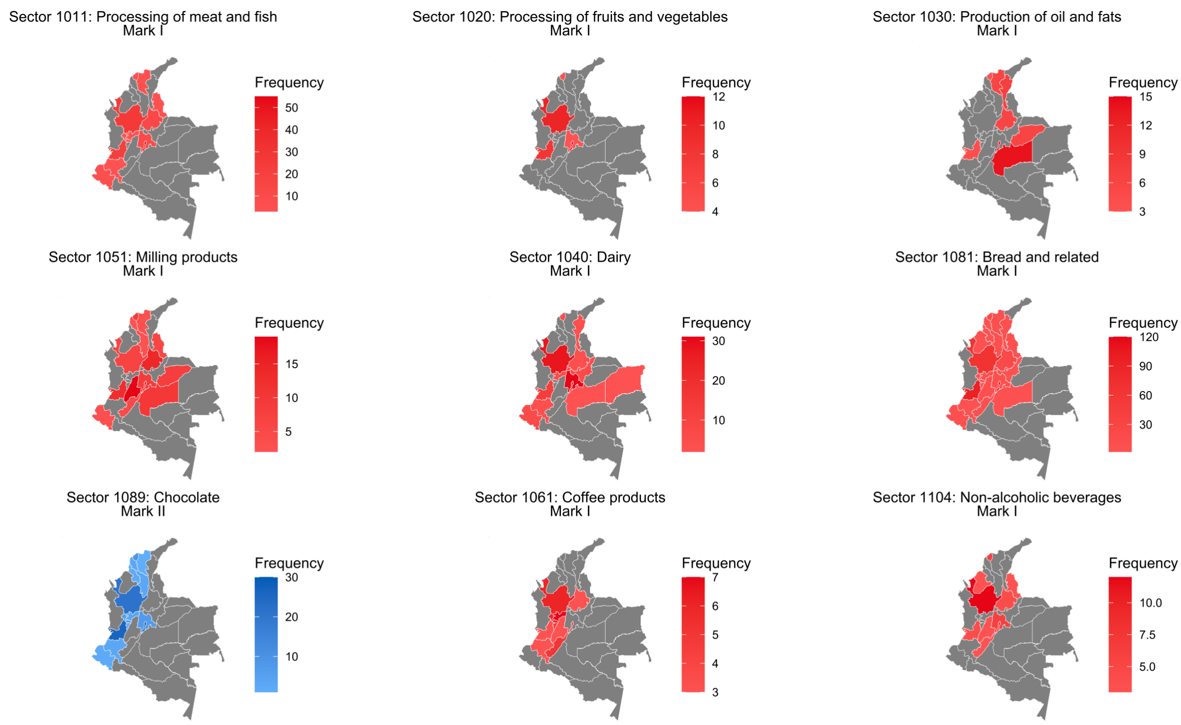
Furthermore, stability of innovation reports values between 0.586 and -1, with a standard deviation of 0.414 and a mean of -0.294.

In contrast, CG1 exhibits larger values of technological opportunities compared to CG2, with a mean near 1 and a standard deviation of 1.496. Additionally, HHI reports a significant increase, with values ranging from 1068 to 2702 within the HHI and a representative value of 1507. Finally, the stability of innovation on average shows slightly higher values towards incremental innovations.

4.2. The Maps

This section discusses the results obtained from mapping the DIVIPOLA classification contained within the EAM survey, not captured by the k-means clustering. These findings will be complemented later on by incorporating theoretical elements from innovation systems. To begin with, the spatial localization of industries related to food products is shown in the figure below.

Figure 2: Spatial distribution of firms elaborating food products



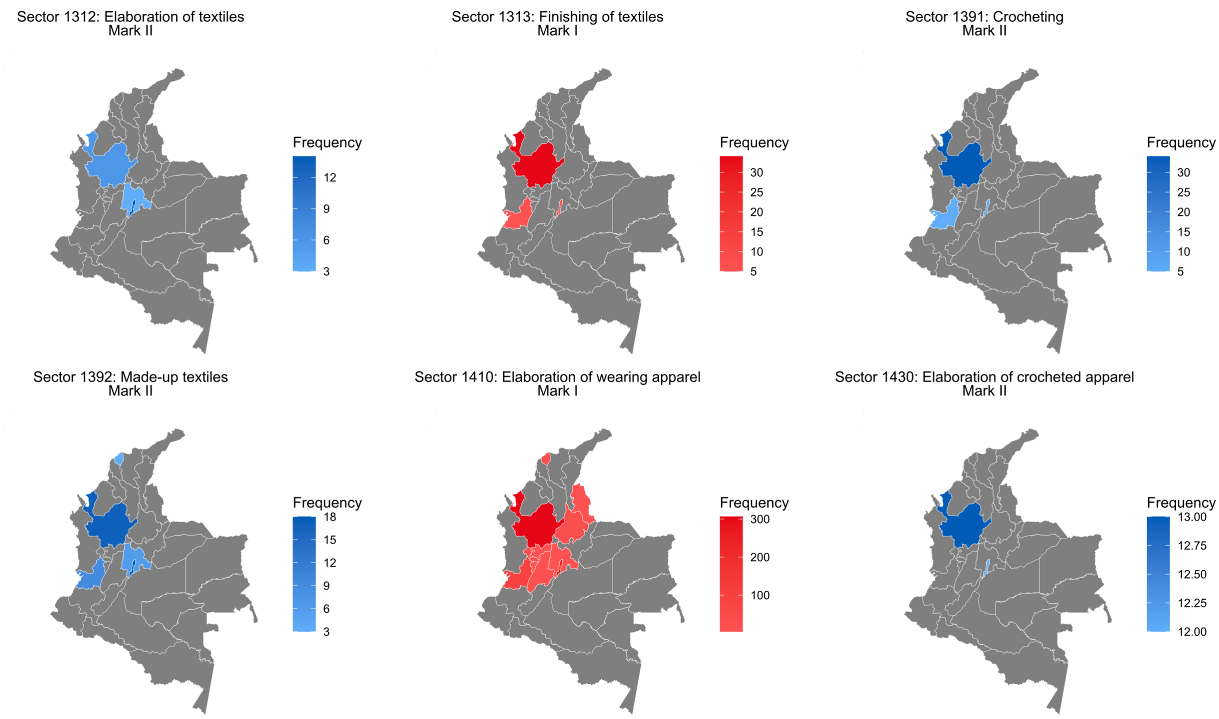
Source: Own elaboration

Both the Andean and Caribbean regions exhibit the highest number of firms. Although Atlántico and Magdalena are the most frequently occurring regions in the Caribbean, they fall behind the Andean Region, where Bogotá[6], Antioquia, and Cundinamarca have the majority of firms. Among them, Antioquia holds the densest clusters across sectors.

In the case of a predominantly Mark I segment, spatial distribution spread across several departments, which is evident in sectors such as oil and fats (1030), dairy (1040), milling products (1051), and bread (1081). On the other hand, the only Mark II industry (chocolates, 1089) spans across both the Caribbean and Andean regions, without major breaks in the distribution. However, it has only two dense centers located in Cauca Valley and Antioquia.

For textile products, the following figure shows maps narrower distributions. Across all sectors in Mark II, and one in Mark I, dense agglomerations appear near important regions in central Colombia. Only the elaboration of wearing apparel (1410) is somewhat disperse, but even its dispersion falls mostly within the Andean Region.

Figure 3: Spatial distribution of firms in the textile segment



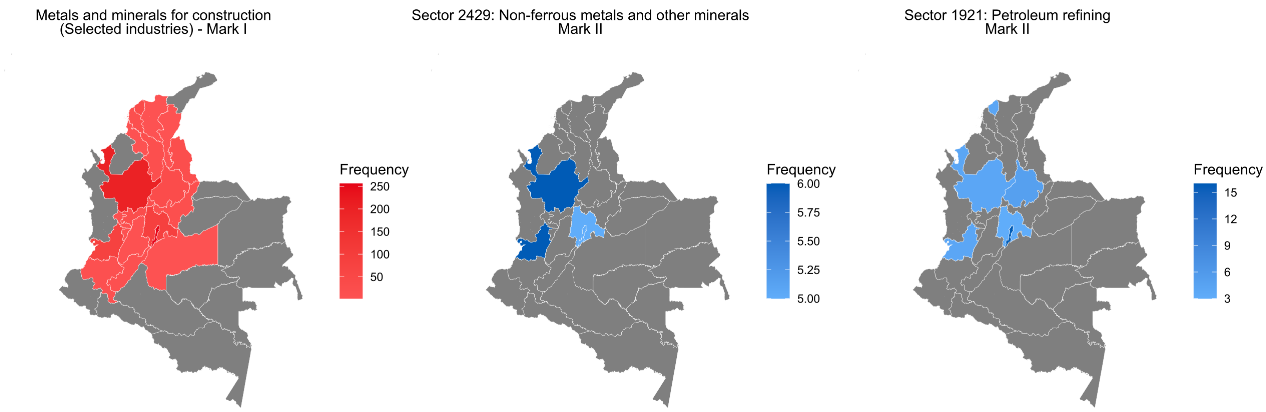
Source: Own elaboration

Moving forward, the next figure displays the spatial distribution of industries involved in the production of petroleum, metal, or minerals[7].

[6]According to Colombia's political-administrative partition, Bogotá is a first-order territory with its own administrative autonomy. Despite serving as the capital of Cundinamarca, its population size and administrative faculties necessitate distinction.

[7]Note that the map for metals does not specify the two-digit ISIC sector. This decision aligns with EDIT's methodology, which selects specific four-digit sectors, which can be found in Annex 1

Figure 4: Spatial distribution of firms processing metals, minerals and petroleum



Source: Own elaboration

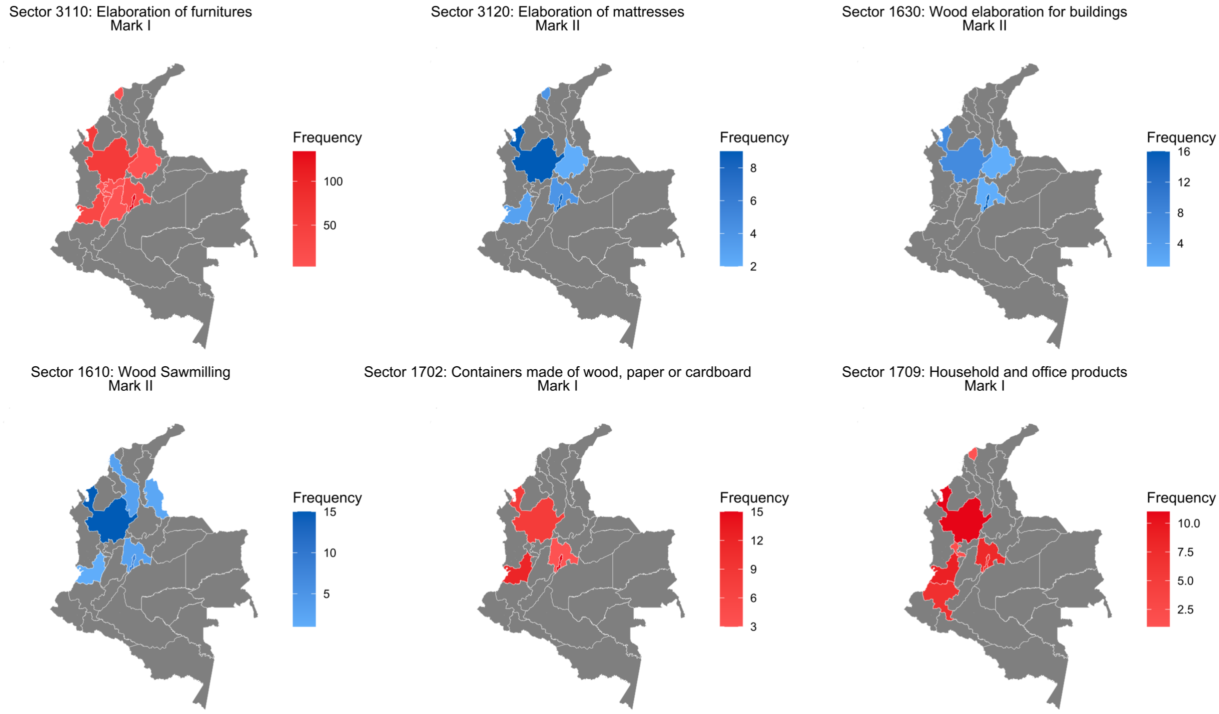
Regarding ferrous metals and minerals commonly used in construction and foundries, there is an extensive distribution across the country. Notably, previous maps (e.g., maps for 1011 and 1081) have revealed a similar distribution that extends across both the Caribbean and Andean regions. These industries have demonstrated the most extensive spatial distributions throughout the study, and all of them belong to the Mark I segment.

However, this is not the case for more specialized metals and minerals (2429), which are very concentrated. Antioquia, once again configures itself as the main region for agglomerations, with Cauca Valley, Cundinamarca and the capital city following. Finally, the last segment to analyse is the elaboration wood, cardboard and paper products, together with furnitures and mattresses. The following figure presents their spatial distribution. This is not the case for more specialized metals and minerals (2429), which demonstrate a high level of concentration. Once again, Antioquia emerges as the primary region, with Cauca Valley, Cundinamarca, and the capital city following closely.

Finally, the last segment to be analyzed is the elaboration of wood, cardboard, and paper products, along with furniture and mattresses. The spatial distribution of these industries is presented in the following figure.

These results show a mixed picture. For instance, the elaboration of containers (1702) displays a spatial distribution typically seen only in Mark II until now. On the other hand, wood sawmilling (1618) exhibits a dispersion characteristic of Mark I, the elaboration of furniture (3110), Mark I, distributes across the central Andean region like if it was Mark II. Finally, the elaboration of mattresses (3120), Mark II, is concentrated in major centers of interest, consistent with previous findings.

Figure 5: Spatial distribution of firms elaborating household furnitures, wood, paper and cardboard products



Source: Own elaboration

Antioquia remains the primary national cluster regardless of the archetype. The Caribbean has fallen behind as we move forward in segments, with only Atlántico displaying firms, in contrast to previous maps where Magdalena, Bolivar or Cordoba were colored. The trend seems to be towards centralism around the Andean region, with one or two followers elsewhere. Notably, there is yet to be any discernible pattern in the Orinoco or Amazonas regions.

5. Analysis of results

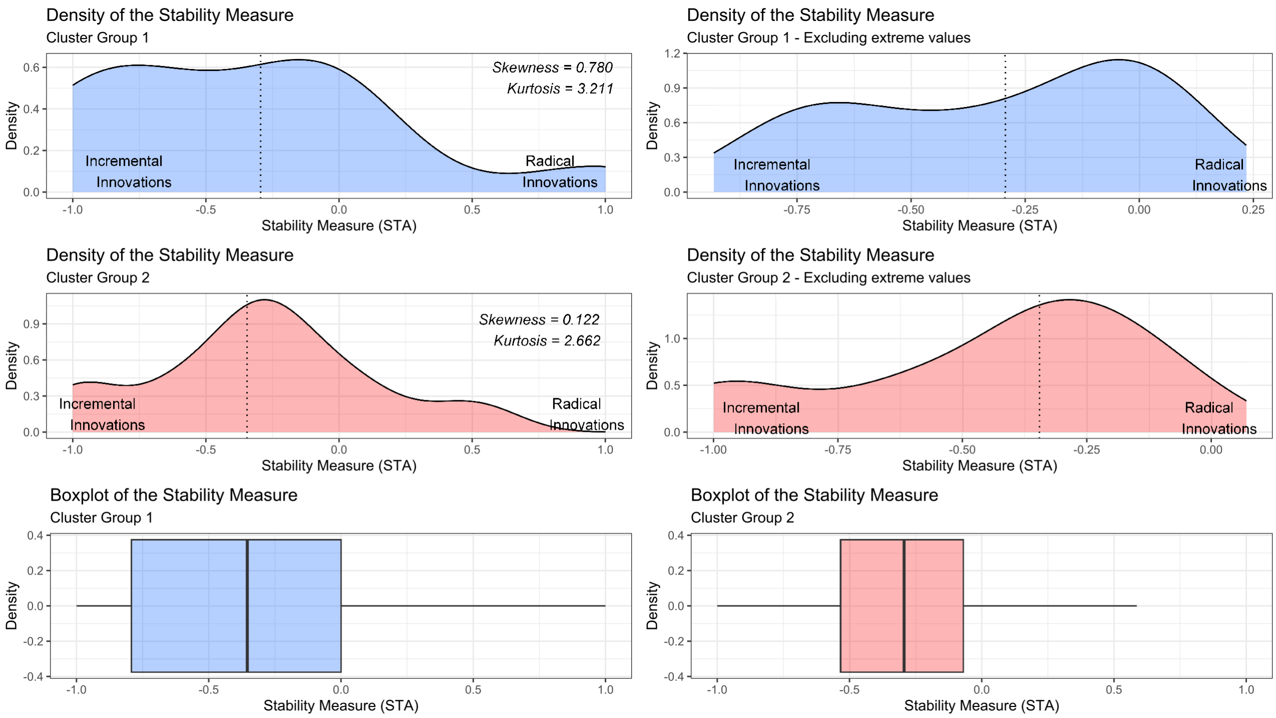
5.1. Inferential analysis

Further statistical tools may shed more light into the nature of both groups and its relationship with innovation archetypes. Based on data distribution as per Figure 2, CG2 has lower stability values when compared to CG1, the density plot for CG1 shows a prominent number of observations between 0 and -1, which based on the measure employed, means that CG1 has many industries with a predominance of incremental innovations in their aggregates. This remark is reinforced by a highly positive skew of the distribution.

CG2, also distributes below 0, but its differentiating factor is its lower density when compared to CG1 across said interval. Even accounting for non-extreme values, results remain consistent. Skewness and kurtosis moments further support this idea.

We may conclude that CG1 industries tend to be dominated by larger shares of incremental innovations in their aggregate, while industries in CG2 do not show this tendency, instead, they have smaller differences between radical and incremental shares rather than disproportions towards one of the two types. Box plots support this conclusion, considering that CG1 left box is wider than its CG2 equivalent.

Figure 6: Distribution of both Groups in the Stability dimension.

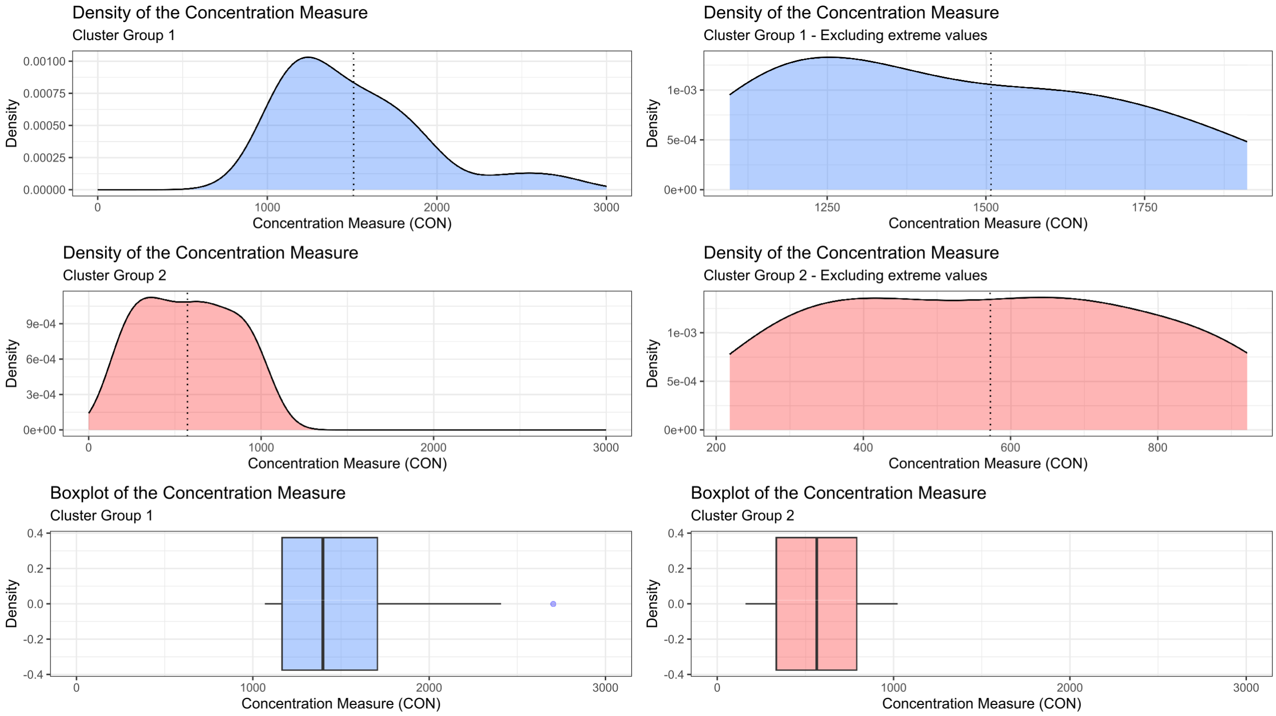


Source: Own elaboration

Within the concentration dimension, density plots in the next figure show a clear difference between both groups. CG2 distribution finds its peak near 500 and reaches 1000 HHI points at most, but CG1 is above 1000, with a long-tailed right side until 2700. This evidence persists after excluding extreme values.

Box plot of both groups provides a different visualization of this phenomenon, but confirms that CG1 market concentration is, on average, much larger than industries on CG2. All this statistical insight points out to one conclusion: CG1 industries are highly concentrated, while CG2 ones are not.

Figure 7: Distribution of both Groups in the Concentration dimension.



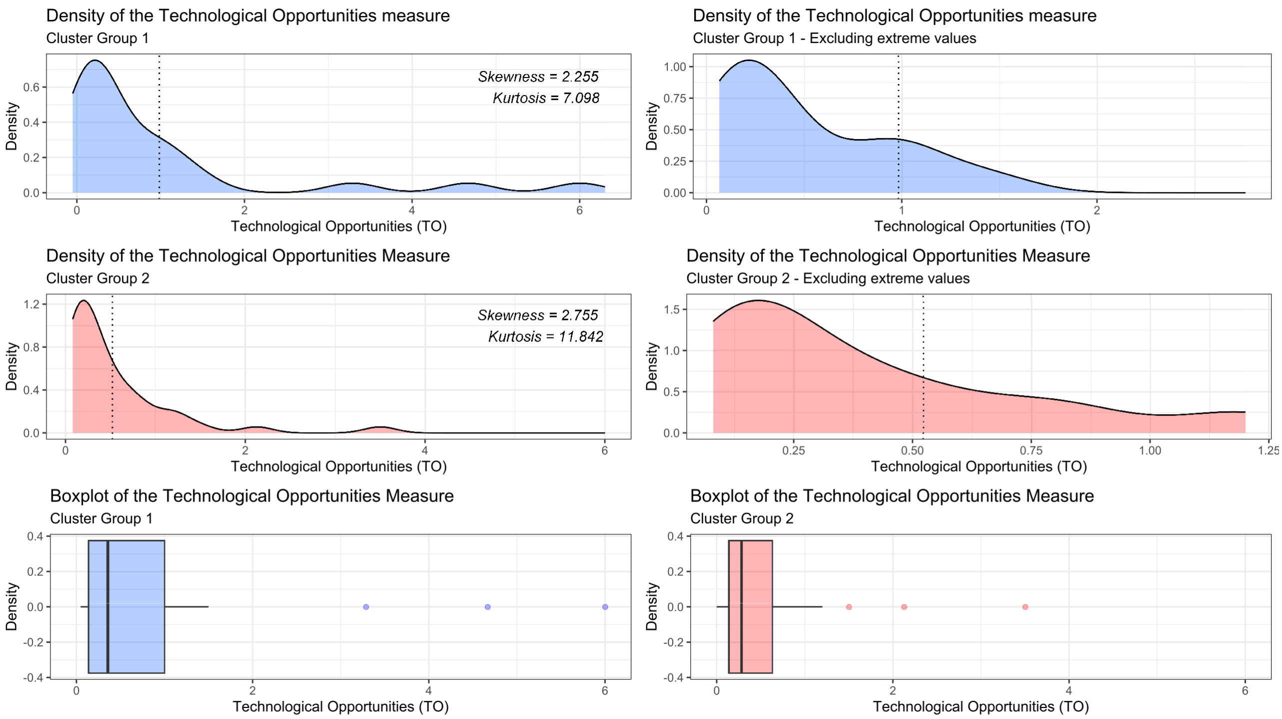
Source: Own elaboration

Finally, the behaviour of technological opportunities shows that both groups distribute mostly near zero. However, statistical moments show that skewness in CG2 is larger than CG1, which may lead us to infer that their registry of new protection mechanisms relative to the existing ones is low, which in turn may indicate low appropriability. Kurtosis in CG2 is also higher, which indicates that the TO measure industries classified on CG2 is highly concentrated around its mean.

In the case of CG1, both moments are lower. Thus, CG1 industries register a higher proportion of protection mechanisms compared to CG2, which may indicate an overall higher level of appropriability. Box plots reinforce this idea, as CG1 right box is wider. Expanding this inference exercise by excluding extreme values, we find a similar result. CG2 distributes mostly between 0 and 1, while CG1 spans wider, from 0 to 2.

Let us go over again our three discoveries. First, we found a trend in CG1 towards high shares of incremental innovations, while CG2 tends to be more balanced between shares. Second, CG1 market concentration is clearly larger in average than CG2. Finally, CG1 industries register more protection mechanisms relative to the existing ones compared to CG2. Per this inferential analysis, and based on reviewed theory, we can characterize CG1 industries under the Mark II pattern, while features of CG2 industries gravitate towards the Mark I Archetype.

Figure 8: Distribution of both Groups in the Technological Opportunities dimension.



Source: Own elaboration

This shows consistency with reviewed literature and theory. Malerba & Orsenigo (1996) concentration approach is consistent in this work, and including novel measures has not shown distorted results. The measure based on Maleki et al. (2018) form for technological opportunities behaves similarly. Finally, proposing a novel measure for stability, based on Baumol (2004), yielded intriguing yet coherent results. In short, we have one of the groups with highly concentrated, with several registrations of protection mechanisms and high share of incremental innovations. On the other hand, there is a disperse group, with less protection mechanisms, and high share of radical innovations.

5.2. Analysis of segments

Due to its position as a periphery economy, food products, commodities (e.g., petroleum, metal and minerals) and first-generation manufacturing (textiles, clothing, footwear) are strategic sectors in Colombia, while advanced manufactures and technological appliances (computers, chipsets, telecommunications) are not. Therefore, we will use these three crucial segments as starting point. In that sense, the relevant ISIC codes for this analysis are production of meat and fish products (101), general food products (102, 103, 104, 105, 108, 110), coffee products (106), petroleum (192), metals and non-metallic minerals (239, 242, 251, 259, 241, 243), furnitures (311, 312), wood, paper and cardboard products (161, 163, 164, 170), and textiles (131, 139, 141, 143).

For food products, nine sub sectors are Mark I. Only one exception in this segment arises, cacao and chocolates (1082) find itself in a Mark II environment. The main Mark I trend implies that flagship industries in Colombia like Coffee elaboration, which sustained its early economic growth in the 20th Century (Luzardo-Luna, 2019), has dispersed efforts, low market concentration and less technological opportunities reflected in less appropriability.

In the case of early manufactures like textiles, most sub sectors are Mark II. Among these, weaving of textiles (1312), made-up articles for home and related (1392), knitting and crocheting (1391, 1430), and other textiles (1399). However, the elaboration of wearing apparel (1410) using several types of materials (wool, leather, fabric, lace, et cetera) behaves as a Mark I industry. Similarly, finishing textiles (1313) through dyeing, dressing, drying or bleaching, behaves as Mark I. It is true that Mark II spans across several sectors, however, manufacture and finishing of clothing in general do not follow this trend.

Another early manufacture has mixed results, while the elaboration of furnitures (3110) in general has features of a Mark I industry, elaboration of mattresses (3120) behaves as a Mark II industry. Something similar occurs in the segment of wood, cardboard and paper products. Sawmilling of wood (1610) and elaboration of wood components for building (1630) behave as Mark II, but the elaboration of wood (1640), cardboard or paper containers (1702), and other manufactures of the same concern, like household and office products (1709) operate under Mark I environments.

Finally, commodities like petroleum have only one sub sector in the study, which is the manufacture of several fuel types used in modern industrial activities (1921), making it a crucial sub sector today. This study has found it to be Mark II . On the other hand, metals and non-metallic minerals have a Mark I trend. Manufacture of clay (2392), cement (2394), concrete (2395), basic iron and steel (2410), structural metals (2511), hand tools (2592), metal coating (2593), among others (2599) are Mark I. However, production and refining of non-ferrous metals like nickel, copper, lead, chrome, manganese, zinc, aluminium, among others (2429) fall into the Mark II category. Note that sub sectors who treat conventional metals or minerals (iron, steel, clay, cement, et cetera) have, but more specialized metals tend to be more concentrated.

5.3. Spatial analysis and innovation systems

In this section, we will argue about the potential set of interactions that may shape innovation systems and spatial localization within those regions. The maps presented earlier show hints of localization behavior similar to those pointed out by Breschi & Malerba (1997, with Mark II industries concentrating in fewer departments when compared to Mark I. In this part, let us look at the most notable findings in terms of localization and the potential interactions behind.

An interesting segment to kick off this analysis is the food products one. Note that sectors like dairy (1040), together with meat and fish (1011), agglomerate close to needed inputs, which could be valleys for livestock, or coasts for fish. A similar phenomenon occurs with coffee products (1061), which distributes around Colombia's coffee region (*Eje cafetero*). Localization can also be influenced by access to major labour centers. To give an example, note that the distribution of manufactured products using fruits or vegetables (1020) coincides with the location of major cities (Cali in Cauca Valley, Barranquilla in Atlántico, Medellín in Antioquia and the zone near Bogotá), where labour supply for manufactures is abundant.

Cultural differences also seem to be important. Historically, as presented by Melo (2017), culture has strongly shaped the country's regions, with rural zones being great suppliers of agriculture, mining or fishing resources, inhabited by low-income households traditionally dedicated to said activities (e.g, laborers, small farmers, fishers). On the other hand, urban centers leaned towards early manufactures and services, with higher wages that allowed the rise of a small middle class, which technified itself through education, informal ties and entrepreneurship. Whats more, urban centers laid the groundwork for innovation, as it benefits from knowledge spillovers and agglomeration economies.

These elements may be behind of Antioquia's position as the department with most firms locating here. Antioquia's position echoes with its 19th and 20th century economic history. As noted by Ocampo (2015) and Luzardo-Luna (2019), hard-currency profits from coffee exports, human capital formation, cultural informal ties between citizens, railway infrastructure, and other resource endowments consolidated Antioquia as an industrial powerhouse, with domestic textile companies like Coltejer (*Colombiana de tejidos*) and Fabricato (*Hilados y tejidos del hato*). In the light of these elements, results suggest that these effects may have had a long term impact on the department, serving as a differentiating factor from other departments. In fact, the textile legacy of the region, in the form of a large spatial concentration of firms, can be seen in all of the six analyzed sectors in the textile segment.

Institutional factors and transport routes are also important. The decision to locate on certain regions also depends on nearby access to transport hubs. To name a few in Colombia, *El Dorado* cargo terminal in Bogotá, Buenaventura port in Cauca Valley, Magdalena river and Caribbean ports like Barranquilla or Cartagena. Similarly, access to government authorities, rule of law and cultural differences between highly urbanized regions and mostly rural ones affect this decision.

This is evident in one of Colombia's most strategic sector (1921, petroleum), spatial distribution shows close proximity to major cities and urban centres like Bogotá in Cundinamarca, but with emphasis on shipping routes and transport. Take for example Cauca Valley and Atlántico, both have petroleum refineries in their territory, and major ports to ship petroleum to international markets. Similarly, Santander and Antioquia's proximity to the Magdalena river allows for ease in transport.

An interesting contrast is the one between non-ferrous metals and minerals (2429) versus metals and minerals commonly used in construction activities. Due to DANE's methodology, sectors from 241 to 243 are aggregated in one map, and since said sectors have different scopes, distribution can be more spread than if we map each sector separately. However, based on DANE's document on ISIC (2022a), most of them share the usage of iron, steel and common minerals, which are widely used in many buildings, and thus, not only access to labour or transport routes influences, but also the potential demand they can find in the construction sectors. These interactions between suppliers and consumers, together with an easier knowledge diffusion due to wide use of it foster a disperse environment, where suppliers localization decision considers potential buyers in the construction sector.

But non-ferrous metals and minerals (2429) such as aluminium, zinc, lead are not as widely used as those mentioned earlier. What's more, treating such minerals required special processes due to particular chemical features, and have more specialized uses, such as in batteries, airplanes, metal coatings, welds or pipelines, that have a narrow set of buyers and demand a more technified use of knowledge when compared to metals and minerals commonly used in construction. This results in an environment where efforts and suppliers concentrate in specific regions of the country, seeking to exploit close distance relationships, access existing buyers and exploit knowledge spillovers.

6. Conclusions and implications

This article aimed to characterize Colombian industries using Schumpeterian Mark I and II patterns through a k-means clustering algorithm. We were limited to certain measures and non-dynamic variables due to data availability and comparability across DANE's databases. Nevertheless, the study found what type of firm drives innovation on each inquired industry. CG1, the less dense group, has been labeled as Mark II, thus, larger firms drive innovation. On the other hand, CG2, the densest group, gravitates toward Mark I, placing small firms as the drivers of innovation. This contributes to an already existing body of literature about characterization exercises, but made in an emerging economy as Colombia.

The differences in quantity between groups shows that Colombian manufacture is more of a Mark I than Mark II. Sectors like groceries are predominantly characterized as Mark I, including Colombia's flagship coffee sector. Sectoral analysis found that textiles are predominantly Mark II, but its most traditional sector, the elaboration and finishing of wearing apparel, has a Mark I environment. On the other hand, oil refining, one of the most important commodities in the modern world, is a Mark II industry in Colombia, but the set of industries supplying the construction sector are Mark I.

We noted that, in general, Mark II industries are less disperse, concentrating in major cities and regions (e.g, Barranquilla, Cali, Bogotá and Medellín), while Mark I ones spread across several regions. This reinforces what was said in the literature review about the features of concentration and dispersion of innovators within a geographical space, and contributes to the discussion of regional innovation in the country. Antioquia positions itself as the main point of interest in our study, perhaps due to long term effects of its industrial history, but the effect of cultural differences and institutions can not be underestimated. What was found within Antioquia reinforces reviewed theoretical elements about innovation systems. Specifically, how cultural, economic and institutional pre-requisites such as transport access, or rule of law together with interactions between agents such as informal ties, knowledge spillovers and consumer-buyer relationships foster innovation and are a differentiating factor from other regions.

Policy implications revolve around the need for heterogeneity in design, echoing with what was said about innovation systems in the theoretical framework and results. It is important to consider specific sectoral relationships, usage of goods, strategic relevance to the country, historical features, presence (or absence) of clusters, nearby infrastructure and other agglomeration dynamics. A suitable policy instrument for these findings can be the Colombian Science Ministry PEDCTI (*Planes Estratégicos Departamentales* *en Ciencia, Tecnología e Innovación*) report, a context-wise diagnosis instrument that assesses strengths, vulnerabilities and potential opportunities of each region in terms of innovation, science and technology.

For example, sectors whose output is widely used as input in other activities, like petroleum refining and supplies for construction, where a change in that sector may spread to the entire market demands carefulness, as an inadequate innovation policy or change into the legal framework may negatively disrupt activities. In contrast, sectors where the output is a final good (e.g, tooling workshops or groceries), the improvement final user and producer relationship can be very beneficial. On the other hand, historical significance of some food products (e.g, Colombian coffee) may benefit from policies that foster producer-foreign relationships, together with improvements on market access through transport.

As the literature continues to explore the relationship between market structure and innovation through characterizations, new findings contribute to a more robust and informed discussion in policy making around the nature of innovation in each industry and leave the door open for potential complements.

Even if these results are relevant and coherent from a theoretical viewpoint, further quantitative research is encouraged. Let us not forget that the variables employed in this study were static in nature due to data limitations, and the analyses made on innovation systems and regions relied heavily on theoretical elements instead of its own quantitative evidence (aside from the DIVIPOLA classification).

Therefore, adding dynamism to variables and employing econometric methods could be a good starting point in the future. Similarly, developing quantitative measures to analyze the set of interactions within Colombian innovation systems could enrich the discussion and set the way forward for more comprehensive characterizations.

References

Arbeláez, M. A., Becerra, A., & Benítez, M. (2021). Contribución fiscal y tributación efectiva de la industria manufacturera en Colombia. Retrieved August 25, 2022, from http://www.repository.fedesarrollo.org.co/handle/11445/4071

Archibugi, D. (2001). Pavitt's Taxonomy Sixteen Years On: A Review Article. *Economics of Innovation and New Technology*, *10*(5), 415–425. https://doi.org/10.1080/10438590100000016

Arezki, R., Hadri, K., Loungani, P., & Rao, Y. (2013). *Testing the Prebisch-Singer Hypothesis Since 1650: Evidence from panel techniques that allow for multiple breaks* (tech. rep. No. 124). Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford. Retrieved May 6, 2021, from https://ideas.repec.org/p/oxf/oxcrwp/124.html

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609–626). Princeton University Press. Retrieved September 21, 2022, from https://www.nber.org/books-and-chapters/rate-and-direction-inventive-activity-economic-and-social-factors/economic-welfare-and-allocation-resources-invention

Arroyo-Mina, J. S., & Guerrero, D. (2018). Schumpeterian Behavior in a CPR Game: Experimental Evidence from Colombian Fisheries Under TURF's Management. *Mediterranean Journal of Social Sciences*, *9*(4), 109. Retrieved August 15, 2022, from https://www.mcser.org/journal/index.php/mjss/article/view/10260

Asheim, B. T., Boschma, R., & Cooke, P. (2011). Constructing Regional Advantage: Platform Policies Based on Related Variety and Differentiated Knowledge Bases [Publisher: Routledge _eprint: https://doi.org/10.1080/00343404.2010.543126]. *Regional Studies*, *45*(7), 893–904. https://doi.org/10.1080/00343404.2010.543126

Asheim, B. T., & Gertler, M. S. (2006). *The Geography of Innovation: Regional Innovation Systems*. Oxford University Press. Retrieved February 17, 2022, from http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286805.001.0001/oxfordhb-9780199286805-e-11

Baumol, W. J. (2004). Entrepreneurial Enterprises, Large Established Firms and Other Components of the Free-Market Growth Machine. *Small Business Economics*, *23*(1), 9–21. https://doi.org/10.1023/B:SBEJ.0000026057.47641.a6

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment [Publisher: Routledge _eprint: https://doi.org/10.1080/0034340052000320887]. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Breschi, S., & Malerba, F. (1997). Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics, and Spatial Boundaries. In *Systems of Innovation: Technologies, Institutions and Organizations* (pp. 130–156).

Breschi, S., Malerba, F., & Orsenigo, L. (2000). Technological Regimes and Schumpeterian Patterns of Innovation. *The Economic Journal*, *110*(463), 388–410. https://doi.org/10.1111/1468-0297.00530

Castellacci, F. (2008). Technological paradigms, regimes and trajectories: Manufacturing and service industries in a new taxonomy of sectoral patterns of innovation. *Research Policy*, *37*(6-7), 978–994. https://doi.org/10.1016/j.respol.2008.03.011

Castellacci, F., & Zheng, J. (2010). Technological regimes, Schumpeterian patterns of innovation and firm-level productivity growth. *Industrial and Corporate Change*, *19*(6), 1829–1865. https://doi.org/10.1093/icc/dtq051

Cerón, C. A., Alcazar, F. L., & M, J. J. G. (2010). Caracterización emprendedora de los empresarios en los Valles de Tundama y Sugamuxi. Boyacá (Colombia). *Pensamiento & Gestión*, (29), 163–189. Retrieved August 29, 2022, from https://www.redalyc.org/articulo.oa?id=64619990008

Corrocher, N., Malerba, F., & Montobbio, F. (2007). Schumpeterian patterns of innovative activity in the ICT field. *Research Policy*, *36*(3), 418–432. https://doi.org/10.1016/j.respol.2007.01.002

D'Allura, G., Galvagno, M., & Mocciaro Li Destri, A. (2012). Regional Innovation Systems: A Literature Review. Retrieved November 19, 2022, from https://papers.ssrn.com/abstract=2180119

Departamento Administrativo Nacional de Estadistica DANE. (2019). Encuesta Anual Manufacturera – EAM - 2018. Retrieved January 13, 2022, from https://microdatos.dane.gov.co/index.php/catalog/650/study-description

Departamento Administrativo Nacional de Estadistica DANE. (2020). Encuesta de Desarrollo e Innovación Tecnológica - EDIT - Industria - 2017 - 2018. Retrieved January 13, 2022, from https://microdatos.dane.gov.co/index.php/catalog/651/datafile/F38

Departamento Administrativo Nacional de Estadistica DANE. (2022a). Clasificación Industrial Internacional Uniforme de todas las actividades económicas. https://www.dane.gov.co/files/sen/nomenclatura/ciiu/CIIU_Rev_4_AC2022.pdf

Departamento Administrativo Nacional de Estadistica DANE. (2022b). Codificación Divipola. Retrieved November 18, 2022, from https://geoportal.dane.gov.co/geovisores/territorio/consulta-divipola-division-politico-administrativa-de-colombia/

Departamento Administrativo Nacional de Estadistica DANE & Dirección de Metodología y Producción Estadística. (2018). Metodología General Encuesta de Desarrollo e Innovación Tecnológica en la industria manufacturera - EDIT. Retrieved September 16, 2022, from https://microdatos.dane.gov.co/catalog/651/related_materials

Fontana, R., Martinelli, A., & Nuvolari, A. (2021). Regimes reloaded! A reappraisal of Schumpeterian patterns of innovation, 1977–2011. *Journal of Evolutionary Economics*, *31*(5), 1495–1519. https://doi.org/10.1007/s00191-021-00735-6

Fontana, R., Nuvolari, A., Shimizu, H., & Vezzulli, A. (2012). Schumpeterian patterns of innovation and the sources of breakthrough inventions: Evidence from a dataset of R&D awards. *Journal of Evolutionary Economics*, *22*(4), 785–810. https://doi.org/10.1007/s00191-012-0287-z

Gilbert, R. (2006). Looking for Mr. Schumpeter: Where Are We in the Competition–Innovation Debate? *Innovation Policy and the Economy*, *6*, 159–215. Retrieved September 21, 2022, from https://www.jstor.org.ezproxy.uninorte.edu.co:2048/stable/25056183

Gössling, T., & Rutten, R. (2007). Innovation in Regions [Publisher: Routledge _eprint: https://doi.org/10.1080/09654310601078788]. *European Planning Studies*, *15*(2), 253–270. https://doi.org/10.1080/09654310601078788

Kirzner, I. M. (1978). *Competition and Entrepreneurship*. University of Chicago Press. Retrieved September 21, 2022, from https://press.uchicago.edu/ucp/books/book/chicago/C/bo27304815.html

Landström, H., & Schön, L. (2010). Industrial Renewal and Entrepreneurship in Sweden: A Structural Cycle Explanation. *Historical Foundations of Entrepreneurship Research*. Retrieved August 14, 2022, from https://www.elgaronline.com/view/edcoll/9781847209191/9781847209191.00029.xml

Langebaek-Rueda, A., & Vásquez, D. M. (2007). *Determinantes de la actividad innovadora en la industria manufacturera colombiana* (tech. rep.). Banco de la República. Bogotá, Colombia. Retrieved August 29, 2022, from https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/5451/be_433.pdf

Leiponen, A., & Drejer, I. (2007). What exactly are technological regimes?: Intra-industry heterogeneity in the organization of innovation activities. *Research Policy*, *36*(8), 1221–1238. https://doi.org/10.1016/j.respol.2007.04.008

Loury, G. C. (1979). Market Structure and Innovation. *The Quarterly Journal of Economics*, *93*(3), 395–410. https://doi.org/10.2307/1883165

Luzardo-Luna, I. (2019). *Colombia's slow economic growth: From the nineteenth to the twenty-first century*. Palgrave Macmillan.

1. The Particular Colombian Case in Latin America: A Singular Path with the

Same Results – 2. The Price of the Regeneration, 1870-1914: How Colombia Missed the Belle Époque – 3. The Take-Off, 1914-1929: Coffee, Railways and Regional Divergence – 4. The Liberal Republic, 1930-1945: Overcoming the Great Depression, the Rise of Interventionism and Economic Slowdown – 5. The Import Substitution Era, 1945-1980: The Consolidation of Interventionism, Financial Repression and the Slow Way to Industrialisation – 6. The Lost Decades, 1980-2000: External Debt, Structural Reforms and a Deep Financial Crisis – 7. Commodities-Driven Growth, 2001-2018: The Colombian Miracle.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge University Press.

Maleki, A., & Rosiello, A. (2014). The dynamic effects of knowledge base complexity on Schumpeterian patterns of innovation: A case study of the upstream petroleum industry.

Maleki, A., Rosiello, A., & Wield, D. (2018). The effect of the dynamics of knowledge base complexity on Schumpeterian patterns of innovation: The upstream petroleum industry. *R&D Management*, *48*(4), 379–393. https://doi.org/10.1111/radm.12251

Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy*, *31*(2), 247–264. https://doi.org/10.1016/S0048-7333(01)00139-1

Malerba, F. (2003). Sectoral Systems and Innovation and Technology Policy. *Revista Brasileira de Inovação*, *2*(2), 329–375. https://doi.org/10.20396/rbi.v2i2.8648876

Malerba, F. (2005). Sectoral systems of innovation: A framework for linking innovation to the knowledge base, structure and dynamics of sectors. *Economics of Innovation and New Technology*, *14*(1-2), 63–82. https://doi.org/10.1080/1043859042000228688

Malerba, F., & Orsenigo, L. (1996). Schumpeterian patterns of innovation are technology-specific. *Research Policy*, *25*(3), 451–478. https://doi.org/10.1016/0048-7333(95)00840-3

Mansfield, E. (1963). Size of Firm, Market Structure, and Innovation. *Journal of Political Economy*, *71*(6), 556–576. Retrieved October 13, 2022, from https://www.jstor.org.ezproxy.uninorte.edu.co:2048/stable/1828440

Marroquín, G. A. (2010). Is Joseph Schumpeter's theory of economic development still useful? The case of a semirural community in Colombia. In D. C. Wood (Ed.), *Economic Action in Theory and Practice: Anthropological Investigations* (pp. 77–98, Vol. 30). Emerald Group Publishing Limited. Retrieved August 15, 2022, from https://doi.org/10.1108/S0190-1281(2010)0000030007

Marsili, O., & Verspagen, B. (2002). Technology and the dynamics of industrial structures: An empirical mapping of Dutch manufacturing. *Industrial and Corporate Change*, *11*(4), 791–815. https://doi.org/10.1093/icc/11.4.791

Martinez, J., La Roca, F., Martinez, I., Martinez, R., & Mendez, S. (1999). CONTENEDOR HIPERMEDIA DE ESTADÍSTICA APLICADA A LAS CIENCIAS ECONÓMICAS Y SOCIALES. Retrieved November 18, 2022, from https://www.uv.es/ceaces/intro/pro/8.htm

Melo, J. O. (2017). *Historia mínima de Colombia*. Colegio de Mexico. https://doi.org/10.2307/j.ctv1nhm21

Nelson, R. R. (1993). *National Innovation Systems: A Comparative Analysis*. Oxford University Press.

Ocampo, J. A. (2017). *Historia económica de Colombia*. Fondo de Cultura Economica.

OECD & Eurostat. (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation* (4th). Retrieved April 12, 2021, from https://www.oecd-ilibrary.org/science-and-technology/oslo-manual-2018_9789264304604-en;jsessionid=82Pf8iyig0VUix25LgMRN_ka.ip-10-240-5-20

Ovallos-Gazabón, D. A., & Amar-Sepúlveda, P. A. (2014). Perfil innovador de la industria manufacturera colombiana. Caso del sector metalmecánico de Barranquilla. *Revista Ingenierías Universidad de Medellín*, *13*(25), 115–136. https://doi.org/10.22395/rium.v13n25a8

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, *13*(6), 343–373. https://doi.org/10.1016/0048-7333(84)90018-0

Prebisch, R. (1950). The economic development of Latin America and its principal problems. Retrieved January 28, 2023, from https://repositorio.cepal.org/handle/11362/29973

Raider, H. J. (1998). Market Structure and Innovation. *Social Science Research*, *27*(1), 1–21. https://doi.org/10.1006/ssre.1997.0608

Schrempf, B., Kaplan, D., & Schroeder, D. (2012). *National, Regional, and Sectoral Systems of Innovation – An overview* (tech. rep. No. 2.2). European Union Community Research. Retrieved November 19, 2022, from https://www.progressproject.eu/wp-content/uploads/2013/12/Progress_D2.2_final.pdf

Schumpeter, J. (1911). *Theory of Economic Development* (1st ed.).

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Routledge.

Shapiro, C. (2012). Competition and Innovation Did Arrow Hit the Bull's Eye? In *The Rate and Direction of Inventive Activity Revisited* (pp. 361–410). University of Chicago Press. Retrieved September 21, 2022, from http://www.bibliovault.org/BV.landing.epl?ISBN=9780226473031

Umaña-Aponte, M., Estupiñan, F., & Duque, C. (2013). *Innovation and productivity in services : An impact evaluation of Colciencias funding programs in Colombia* (Working Paper). Centro de Investigaciones Económicas (CINVE), Montevideo, UY. Retrieved August 15, 2022, from https://idl-bnc-idrc.dspacedirect.org/handle/10625/52552

van Dijk, M. (2000). Technological regimes and industrial dynamics: The evidence from Dutch manufacturing. *Industrial and Corporate Change*, *9*(2), 173–194. https://doi.org/10.1093/icc/9.2.173

A. Annex 1: Universe of study for the EDIT survey.

ISIC code	Economic activity
101	Processing and preservation of meat and fish
102	Processing and preservation of fruits, vegetables and tubers
103	Production of oils and fats
104	Dairy Processing
105	Production of milling products, starches and their derivatives
106	Elaboration o f coffee products
107	Sugar and panela processing
108	Manufacture of other foodstuffs
109	Preparation of prepared feedingstufss for animals
110	Beverage production
131	Spinning, weaving and finishing of textiles
139	Manufacture of other textiles
141	Manufacture of clothing
143	Manufacture of knitted and crocheted articles
151	Tanning and retanning of hides and manufacture of travel goods
152	Footwear manufacturing
161	Sawing, waxing and impregnation of wood
162	Manufacture of sheets of wood for plating, boards and panels
163	Manufacture of wooden parts and pieces
164	Manufacture of wooden containers
169	Manufacture of other wood products
170	Manufacture of paper and cardboard
181	Printing activities and related services
190	Coking, oil refining and fuel mixing
201	Manufacture of basic chemicals and their products
203	Manufacture of synthetic and artificial fibres
221	Manufacture of rubber products
222	Manufacture of plastic products
231	Manufacture of glass and glass products
239	Manufacture of non-metallic mineral products
242	Basic precious and non-ferrous metal industries
251	Manufacture of metal products for structural use
259	Manufacture of other products made of metal

260	Manufacture of computer, electronic and optical products
270	Manufacture of electrical appliances and equipment
281	Manufacture of machinery and equipment for general use
282	Manufacture of machinery and equipment for special use
291	Manufacture of motor engines and their engines
292	Manufacture of bodies for motor vehicles
293	Manufacture of parts, auto parts and vehicle accessories
300	Manufacture of other types of transport equipment
311	Furniture manufacturing
312	Manufacture of mattresses and box springs
321	Manufacture of jewellery and related articles
323	Manufacture of articles and equipment for the practice of sport
324	Manufacture of games, toys and headbutts
325	Manufacture of medical and dental instruments, apparatus and materials
329	Other manufacturing industries
330	Maintenance and repair of metal products, machinery and equipment
2021	Manufacture of pesticides or chemicals for agricultural use
2022	Manufacture of paints, varnishes and similar coatings
2023	Manufacture of soaps, detergents and perfume
2029	Manufacture of other chemicals
2100	Manufacture of pharmaceutical products and medicinal chemicals
241 to 243	Metal foundry

B. Annex 2: Industry-level results

Three-Digit ISIC	Four-Digit ISIC	TO	CON	STA	Mark
101	1011	0.157	305.89	0.234	I
102	1020	0.353	815.72	-0.111	I
103	1030	0.121	396.99	-0.933	I
104	1040	0.136	488.88	-0.294	I
105	1051	0.094	198.09	-0.279	I
106	1061	1.192	412.75	-1	I
108	1081	0.136	334.15	0.099	I

	1082	0.062	1389.42	-0.784	II
	1089	0.172	674.37	-0.217	I
109	1090	0.14	423.24	-0.12	II
110	1104	0.109	790.62	0.467	I
131	1312	0.308	1753.45	0.926	II
	1313	0.818	656.25	-0.333	I
139	1391	0.5	1904.26	-0.143	II
	1392	0.404	1156.52	-1	II
	1399	0.102	1504.79	-0.077	II
141	1410	0.252	160.28	-0.313	I
143	1430	0.048	2702.96	-1	II
151	1511	1.5	1291.6	0	II
	1512	0.158	1847.33	-1	II
152	1521	0.092	651.43	-0.25	I
	1522	0.211	1913.61	0	II
	1523	1.5	1017.81	-1	I
161	1610	1	1093.55	1	II
163	1630	6	1105.45	0	II
163	1640	0	1023.11	-1	I
170	1702	0.766	933.74	-0.176	I
	1709	0.054	716.54	-0.534	I
181	1811	0.866	193.94	0	I
192	1921	1.25	2407.87	-0.667	II
201	2011	0.425	528.27	0.091	I
	2012	0.195	1316.55	-1	II
	2013	0.138	1281.94	-0.692	II
202	2021	0.086	866.59	-0.613	I
	2022	0.114	740.74	-0.048	I

	2023	0.324	414.61	-0.402	I
	2029	0.205	499.38	-0.622	I
210	2100	0.1	259.39	0.107	I
221	2212	3.286	1068.11	0	II
	2219	1	1081.54	0.111	II
222	2221	0.16	386.37	-0.19	I
	2229	1.179	164.87	-0.258	I
231	2310	0.283	1186.85	-0.333	II
239	2392	0.28	544.95	-0.647	I
	2394	0.129	919.09	0.586	I
	2395	0.585	231.47	-0.905	I
241	2410	0.165	583.96	0.545	I
242	2429	0.133	1690.52	-1	II
251	2511	1.2	628.54	-0.371	I
259	2592	2.125	693.44	-0.2	I
	2593	0.299	974.42	-0.44	I
	2599	0.448	285.49	-0.07	I
271	2712	0.408	750.81	-1	I
274	2740	0.92	1626.86	-0.818	II
279	2790	0.821	1168.34	-0.375	II
281	2813	0.429	1633.96	-0.5	II
	2816	4.667	1405.11	0	II
	2819	0.505	317.84	-0.353	I
282	2829	3.5	886.41	-0.333	I
292	2920	0.804	920.49	-0.5	I
293	2930	0.632	563.31	0.412	I
311	3110	0.531	219.11	-0.382	I

312	3120	0.099	1125.81	-0.273	II
325	3250	0.256	887.84	-0.714	I
	3290	0.066	1528.87	-0.667	II