UNIVERSIDAD
NACIONAL
DE COLOMBIA

Production networks and volatility in the Colombian manufacturing industry

Jennifer Taborda Martínez

Universidad Nacional de Colombia
Facultad de Ciencias Económicas
Bogotá Colombia
2021

Production networks and volatility in the Colombian manufacturing industry

Jennifer Taborda Martínez

Tesis de investigación presentada como requisito parcial para optar al título de:

Magister en Ciencias Económicas

Directora:

Doctora, Nancy Milena Hoyos Gómez

Universidad Nacional de Colombia

Facultad de Ciencias Económicas

Bogotá Colombia

2021

A Lorenzo y Giorgio

Agradecimientos

En primer lugar, quiero agradecer a la Facultad de Ciencias Económicas de la Universidad Nacional por acogerme para la realización este proyecto. A la Dra. Nancy Milena Hoyos por su guía y supervisión durante el proceso de investigación. Al grupo de apoyo administrativo de la Maestría en Ciencias Económicas, quienes estuvieron siempre prestos a guiar y apoyar cada paso de este camino. Una mención especial de agradecimiento al Dr. Jorge Armando Rodriguez Alarcón, Decano de la Facultad de Ciencias Económicas quien apoyó decididamente al grupo de estudiantes de reingreso en su sueño de culminar sus estudios de maestría en la Facultad. Igualmente, a los líderes de dicha iniciativa: Juan Carlos Cortes, Rafael Bernal y Delfín Soto, quienes aportaron su tiempo y entusiasmo para impulsar este proyecto. Quisiera también agradecer especialmente al Dr. Iván Darío Hernandez Umaña, Profesor de la Universidad de Ibagué y quien fue inicialmente mi supervisor. Su apoyo fue decisivo para el camino que siguió luego del inicio de mis estudios de maestría y en consecuencia para todas las pequeñas decisiones que hoy culminan en este trabajo de grado.

A mi familia y amigos que me han apoyado en cada paso de los muchos que me tomo culminar este proyecto, sin su amor y apoyo incondicional nada habría sido posible.

Resumen

Redes de producción y volatilidad en la industria colombiana

La volatilidad del producto es un indicador de la estabilidad de crecimiento, de la inversión y del empleo en una economía. Trabajos anteriores han mostrado que redes insumo-producto compuestas por algunos productos muy conectados y muchos otros con pocas conexiones entre sí, permiten la transmisión de choques sectoriales al resto de la economía, precisamente mediante la transferencia del choque entre cadenas de productos centrales que a la vez están conectados unos a otros, amplificando la magnitud del choque inicial y generando volatilidad agregada. Este artículo estudia la estructura de la red insumo-producto para Colombia entre 1982 y 2012, para entender su papel en la generación de volatilidad agregada del producto industrial. Para ello, se construyen una serie de redes insumo-producto a nivel de bienes, basadas en datos de la Encuesta Anual Manufacturera a nivel de establecimiento. La riqueza de los datos permite estudiar las propiedades estructurales de la red, mediante la caracterización de las secuencias de primer y segundo orden de la centralidad de los productos en la red. Se observa que la red de insumo-producto en la industria manufacturera en Colombia está compuesta por algunos pocos productos centrales que proveen de insumos a muchos sectores y que se encuentran conectados entre sí, y muchos otros productos que tienen una importancia menor en la provisión de insumos. Este tipo de estructura heterogénea tiene un impacto en la generación de volatilidad agregada que es 3,3 veces mayor al impacto que tendría una red insumo-producto en la cual la participación de todos los sectores es homogénea. En conclusión, la red intersectorial tiene un papel muy importante en la propagación de choques sectoriales y en consecuencia en la generación de volatilidad del PIB industrial.

Palabras clave: Volatilidad agregada, redes de producción, diversificación, tablas insumo-producto.

Códigos JEL: D57, D85, E32, L14

Abstract

Production networks and volatility in the Colombian manufacturing industry

Output volatility is a useful indicator to assess the stability growth, investment, and employment of any economy. Previous research has shown that an input-output structure dominated by few sectors, enables the transmission of sector-level shocks via central activities to the rest of the economy, amplifying the effect of microeconomic shocks into aggregate output volatility. This paper studies the structure of the industrial input-output network in Colombia between 1982 and 2012 to understand its role as a source of industrial output volatility. We build a series of unique input-output networks at the product level, based on industrial survey data on production at the plant level for Colombia. The richness of the data allows us to study the structural properties of the network by characterizing the distribution of first- and second-order degree sequences. The impact of the intersectoral network in the propagation of sector-level shocks into output volatility in the manufacturing industry is quite central. The input-output industrial network in Colombia is composed of few very central inputs providers connected among them and many other products with smaller importance as input suppliers. Such heterogeneous structure amplifies the impact of the intersectoral network on output volatility 3.3 times on average, versus the impact that a balanced structure would have on aggregate volatility.

Keywords: Aggregate volatility, production networks, diversification, input-output linkages.

JEL codes: D57, D85, E32, L14

Content

List of figures

Pág.

List of tables

Pág.

Introduction

Economic stability is one of the key variables of macroeconomic well-being in an economy. It is known that aggregate output volatility, or in other words, large variations on the growth rate of aggregate output (measured by the standard deviation of the growth rate) are associated with low economic growth (Badinger, 2010; Ramey & Ramey, 1995), low investment (Joshua Aizenman & Marion, 1999), and poverty and inequality (Breen & Garcia-Penalosa, 2005; Laursen & Mahajan, 2005). Output volatility hits harder to developing countries (Hakura, 2007; Sahay & Goyal, 2006) with long-lasting consequences on human capital accumulation and long-term growth, in the case of large economic downturns (Joshua Aizenman & Pinto, 2005; Calderon & Yeyati, 2009). At the sector level, changes in output are followed almost by a proportional change in employment (Jaimovich et al., 2013).

Recent research has shown that the structure of the production network of an economy, plays an important role in the propagation of microeconomic shocks into increased output volatility (Acemoglu et. al 2012). A non-homogeneous structure of the production network (one in which some sectors are more important suppliers than others to the rest of the economy) can act as a transmission channel of idiosyncratic shocks into aggregate volatility. Therefore, to assess the effect of microeconomic shocks on the medium- and long-term vulnerability of the economy, it is necessary to understand the structural properties of the underlying I-O linkages that support it.

In the current scenario, when the world faces the economic consequences of the supply and demand shocks produced by COVID 19, this kind of analysis is of the utmost importance. One of the most immediate consequences of COVID 19 is the disruption of value chains and the destruction of firms and jobs (Baldwin & Weder di Mauro, 2020; del Rio-Chanona et al., 2020). It is natural to anticipate a change in the structure of the production network, especially in the manufacturing sector where the market for intermediates is particularly important. The consequences of such changes on the aggregate volatility of industrial output are particularly worrisome in terms of employment and long-term economic performance, given the role of the manufacturing industry as an engine of growth and a source for export diversification and sophistication (OECD/World Trade Organization, 2019; Szirmai, 2012).

Empirical exercises have covered so far the United States, France, Belgium, Japan, Ecuador, Turkey, and Brazil and all of them use databases covering several countries at a high level of sector aggregation (Carvalho & Tahbaz-Salehi, 2019; Romero et al., 2018). The use of aggregated data clouds the perception of heterogeneity at the sector level, but due to data availability, most of the evidence relies on the two-digit level of aggregation of the product classification. This research proposal intends to cover that gap by exploiting Industry Survey data at the 7-digit level for Colombia to answer two main research questions: Which are the structural properties of the industrial production network in Colombia? and what is the role of the production network in the propagation on microeconomic shocks into industrial output volatility?

Given the evidence found for other countries and the level of data disaggregation used in this exercise, we will test two main hypotheses:

- Industrial output volatility in Colombia is influenced by the heterogeneous structure of the production network. The existence of products with a greater centrality as inputs suppliers in the network facilitates the propagation of microeconomic shocks into aggregate industrial volatility.

- Between 1982 and 2012 the I-O industrial sector in Colombia has become more heterogeneous despite the increase in the number of products in the network, which has increased the vulnerability of the industry for microeconomic shocks to increase aggregate volatility.

Input-output structure and output volatility

For a long time, the role of the production network on the propagation of microeconomic shocks into aggregate fluctuations was dismissed by the literature. Classic macroeconomics argues that idiosyncratic shocks at the industry level have a negligible effect on aggregate volatility because aggregate output converges to its mean at a large rate. This has been called the “diversification argument”. The argument relies on the law of large numbers and states that sectoral shocks in an economy with n number of sectors, average out at a rate $\sqrt{n}$. In that sense, in a diversified economy (with a large n), microeconomic shocks fade out fast and have only a marginal effect on aggregate volatility, the explanation is known as the “diversification argument” (Lucas, 1977).

To understand the *diversification argument* lets state some basic definitions. A network is in its simplest form, a collection of points joined together in pairs by lines. In the network literature, a point is referred to as a *node* and a line is referred to as an *edge* (Newman, 2018). In the case of a *production network* in its most disaggregated version, a node is a product (e.g. cotton, woven fabrics of cotton, T-shirts), and an edge is an amount of money that represents the value needed from product A to produce product B. In other words, a production network summarizes all Input-Output (I-O) relationships between products or economic activities of an economy. In that sense, we understand the economy as a set of economic activities connected by relationships of demand or supply. In such system, an economic shock that affects the dynamics of one economic activity, can have consequences on all the other activities connected to it. For example, a sudden downturn in the demand for t-shirts, might hurt the sales of woven fabrics of cotton, and in consequence, reduce the demand for raw cotton in general, which might affect the aggregate level of output in the country, causing aggregate output volatility. However, the effect that a microeconomic shock can have on the aggregate output volatility is mediated by the structure of the production network. For example, in the case of an economy that produces a large variety of products, and the importance of those products is relatively homogeneous (a highly diversified economy), the effect of a negative demand shock in one sector might not have large consequences on the aggregate level of output because a related positive effect of the same magnitude in another sector can compensate it. Coming back to our previous example, think of a negative demand shock for t-shirts, affecting the demand for woven fabrics of cotton, therefore reducing its price, increasing the sales of woven fabrics of cotton to produce seats at a cheaper price, increasing the sales of seats, therefore, offsetting the initial negative shock. In that case, the microeconomic shock fades out fast, having a marginal or none effect on the aggregate level of product, and by consequence no effect on the aggregate output volatility.

It was only a few years ago that this argument was re-evaluated by showing that the “diversification argument” applies only under specific characteristics of the production network. Acemoglu et al. (2012) finds that for non-symmetrical production networks the rate at which microeconomic shocks fade out is smaller than $\sqrt{n}$. Therefore, the effect on aggregate volatility is no longer negligible, and the extent of its effect depends on the first and second order characteristics of the production network (Acemoglu et al., 2012). In other words, shocks at the sector level can propagate through the production network by two possible channels: direct and indirect connections. In the case of direct connections, if they are characterized as asymmetric, (meaning that some sectors are more important as suppliers than others) a shock to a sector that provides to an exceptionally large number of other sectors, will propagate directly to the many other sectors related to it. On the other hand, a shock to one sector can have effects not only over the sectors immediately related, but over other sectors connected to them, generating cascade effects. The authors formalize the relationship between the first and second order characteristics of the production network and the rate of decay of economic shocks, providing a base to calculate the influence of microeconomic shocks on aggregate volatility.

These findings are closely related to the granularity literature (Gabaix et al., 2011) that calls for attention towards the heterogeneity at the granular level of economic activity (firms or sectors) as a determinant of the behavior of macroeconomic aggregates. This literature argues that large heterogeneity at the firm level impedes micro shocks to cancel each other out, which in the end impacts macroeconomic variables. In a similar vein, Acemoglu, Ozdaglar, & Tahbaz-Salehi (2017) extend the analysis of the consequences of heterogeneity at the sector level, to the likelihood of large economic downturns. In a nutshell, the paper finds that the likelihood of large contractions on output depends on both the distribution of shocks and the structure of the production network. Deep recessions are as likely in an economy that is subject to fat-tailed shocks, as in an economy subject to thin-tailed shocks but with a production network with a non-homogeneous structure.

On top of the theoretical formulations, empirical research has uncovered the relationship between the structure of production networks and aggregate volatility. A series of papers using I-O data for the United States (US) have shown that sectoral heterogeneity is strong enough to allow aggregate volatility to be produced via microeconomic shocks (Acemoglu et al., 2012; Carvalho & Tahbaz-Salehi, 2019). Atalay (2017) calculates that 83% of aggregate output volatility can be attributable to industry-level shocks.

Evidence using industry-level I-O data for a group of countries shows that in a wide range of countries, the industrial production network exhibits similar characteristics to the US in terms of sector heterogeneity (Blöchl et al., 2011; McNerney et al., 2013). In the case of Ecuador, Romero et.al (2018) finds that the production network is non-homogeneous as in the US case, but in comparison with the US, it is a simpler economy with a large sparse network which can diminish the power of the I-O structure to transmit economic shocks into macroeconomic effects. In this work, they use a disaggregated version of the I-O table with 245 sectors, a smaller set of data than the one used for the US with 474 sectors. Similarly, Gonçalves et al. (2020) study the structure of the I-O table for Brazil, containing 67 sectors in 2010 and 2015. The authors find evidence of an asymmetric structure in which some sectors are prominent in their role as suppliers for the rest of the economy.

Most of the empirical work is based on aggregated data at the sector level, due to the lack of more disaggregated data on I-O linkages. However, the level of disaggregation is fundamental to test the diversification hypothesis with confidence. As we aggregate data into sectors, it is plausible to create a false structure that looks concentrated according to the distribution of large and small sectors in each aggregation group. Therefore, the higher the level of disaggregation of data, the larger is the confidence we can have in the distribution fit. However, to our knowledge, no studies used product-level data. This research fills that gap, by using industrial data at the product level for Colombia, to reveal the asymmetric character of the I-O network of industrial products, and to calculate to what extent this kind of economic structure is determinant for the propagation of microeconomic shocks into aggregate volatility.

Methodology

2.1 Model

The role of production networks in the propagation of microeconomic shocks into aggregate volatility can be understood via a baseline model of production networks. This section presents the basic model originally proposed by Long & Plosser (1983), in the static version analyzed by Acemoglu et al.(2012) and Carvalho & Tahbaz-Salehi (2019). The model formalizes the relationship between microeconomic shocks, and aggregate volatility on an economy with n distinct goods, each of them produced by a competitive sector denoted by $\{1,2,\ldots,n\}$. In this economy a good can be either consumed by households or used as intermediate input. The sectors produce with a Cobb – Douglas type of technology with constant returns to scale. The output of sector i is given by:

(1) $y_i = z_i l_i^{\propto_i} \prod_{j=1}^{n} x_{ij}^{a_{ij}}$,

where l_i is the amount of labor hired by sector i, x_{ij} is the quantity of good j used to produce good i, $\propto_i \in (0,1)$ is the share of labor, and z_i is an idiosyncratic productivity shock. Productivity shocks are independent across sectors. The exponent $a_{ij} \geq 0$ denotes the share of good j in the total intermediate demand of sector i. In that sense, a large a_{ij} means that good j is an important input to produce i, $a_{ij} = 0$ means that sector i does not use product j as an input. It is possible that $a_{ii} > 0$ if some proportion of the good is used for its production, and most likely $a_{ij} \neq a_{ji}$ as the reliance of one industry on the other as a supplier might not be the same. Under the assumption of constant returns to scale the input shares of all sectors add up to one; that is $\propto_i + \sum_{j=1}^{n} a_{ij} = 1$ for all $i = 1,2,\ldots,n$.

Notice that under competitive factor markets, a_{ij} corresponds to an entry of the input-output table, measuring the value of j required as input, per dollar of production of good i. The input-output linkages between sectors can be summarized with an *input-output matrix* $\boldsymbol{A}$ with entries a_{ij}; or represented on a directed weighted graph on n vertices called *production network*. Each vertex in the graph corresponds to a sector and a directed edge (j,i) with weight $a_{ij} > 0$, if j is an input supplier to sector i. Two definitions are useful to the model:

the *weighted outdegree* of sector i, is defined as the share of sector i's output that is used as intermediate input by the entire economy, that is:

(2) $d_i \equiv \sum_{j=1}^{n} a_{ij}$,

and the Domar weight of sector i is defined as the sector's sales as a fraction of Gross Domestic Product (GDP), that is:

(3) $\lambda_{i=} \frac{p_i y_i}{GDP}$,

where, p_i is the price of sector i, and y_i is the output of sector i.
The representative household is endowed with one unit of labor, supplied inelastically, and has logarithmic preferences over the n distinct goods given by

(4) $u(c_1, \dots, c_n) = \sum_{i=1}^{n} \beta_i \log\left(\frac{c_i}{\beta_i}\right)$,

where c_i is the amount of good i consumed, and $\beta_i \geq 0$ are constants that measure the good's shares in the utility function, normalized such that $\sum_{i=1}^{n} \beta_i$ =1.

2.1.1 Competitive equilibrium

To determine prices and quantities in competitive equilibrium, firms in sector i maximize profits $\pi_{i=} p_i y_i - w_i l_{i-} \sum_{j=1}^{n} x_{ij}$, taking prices and wage w as given. The first-order conditions for firms are given by $x_{ij} = a_{ij} p_i y_i / p_j$ and $l_i = \alpha_i p_i y_i / w$. Substituting these values on the firm's production function ((1) and taking logarithms imply that: $\log(p_i/w)$= $\sum_{j=1}^{n} a_{ij} \log(p_j/w) - \epsilon_i$, where $\epsilon_i = \log(z_i)$. As this relationship must hold for all sectors, it represents an equation system to solve for all relative prices in terms of productivity shocks. It is possible to represent such equation system in matrix form as: $\hat{p} = \boldsymbol{A}\hat{p} - \epsilon$, where $\boldsymbol{A}$ is the economy's input-output matrix, $\hat{p} = (\log(p_1/w), \dots, \log(p_n/w))'$ is the vector of log relative prices and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is the vector of productivity shocks. In that sense, the equilibrium vector of relative prices is given by:

(5) $\hat{p} = -(1 - \boldsymbol{A})^{-1}\epsilon,$

where an equation is:

(6) $\log\left(\frac{p_i}{w}\right) = \log\left(\frac{p_i}{GDP}\right) = -\sum_{j=1}^{n} l_{ij}\epsilon_j$

The matrix $L = (1 - \boldsymbol{A})^{-1}$=$\sum_{k=0}^{\infty} \boldsymbol{A}^k$ is the *Leontief inverse* matrix, each element l_{ij} of the matrix describes the amount of output from sector j directly or indirectly required for production in sector i. This expression takes into account that sector i demands output from sector j indirectly via the consumption of goods j by other direct suppliers of i. In terms of the production network, an element of the *Leontief inverse* accounts for all possible directed walks[1] that connect sector j to sector i over the network.

The representative household demands a quantity $c_j = \beta_j w / p_j$ in equilibrium (Equation (4)), and the market-clearing condition for good is j: $y_j = c_j + \sum_{j=1}^{n} x_{ij}$. Replacing the first order conditions for firms and households we get: $p_j y_j = \beta_j w + \sum_{j=1}^{n} a_{ij} p_i y_i$. If we divide both sides by w which represents the household labor income and is equivalent to the economy's value-added, we get:

(7) $\lambda_j = \beta_j + \sum_{j=1}^{n} a_{ij}\lambda_i,$

where λ_i is the Domar weight defined in Equation (3). Writing in matrix form the same equation, and solving for the vector of Domar weights $\boldsymbol{\lambda}$ gives us: $\boldsymbol{\lambda} = (1 - \boldsymbol{A}')^{-1}\boldsymbol{\beta}$ or, equivalently, $\boldsymbol{\lambda} = \sum_{j=1}^{n} \beta_j l_{ji}$. Combinig this equation with Equation (6) we get:

[1] "*A walk in a network is any sequence of nodes such that every consecutive pair of nodes in the sequence is connected by an edge*" (Newman, 2018, p.131)

(8) $\log(y_i) = \sum_{j=1}^{n} l_{ij}\epsilon_j + \delta_i$,

where δ_i is a constant independent of the shocks ϵ_j Equation (8) express the output of sector i as function of shocks to industries $j \neq i$, and the Leontief coefficients l_{ij}. Therefore it shows that: a) shocks propagate from one industry to another via the production network, b) the propagation of shocks happens via direct and indirect production linkages and c) productivity shocks propagate downstream, meaning from a sector to their direct and indirect clients.

2.1.2 Aggregate volatility and the structural properties of the network

Furthermore, Acemoglu, et. al (2012) prove that it is possible to relate aggregate volatility to the structural properties of the economy's input-output matrix, such as the first- and second-order degree sequence. Starting with first-order interconnections, given an economy ε_i with a degree sequence $\{d_{1,}d_2, \ldots, d_n\}$, that describes the *weigthed outdegree* values d_i (as defined in Equation (2)) for all n sectors in the economy. The coefficient of variation of the first-order degree sequence is:

(9) $\mathrm{CV_n} \equiv \frac{1}{\bar{\mathrm{d}}_\mathrm{n}}\left[\frac{1}{(\mathrm{n}-1)}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{d_i} - \bar{\mathrm{d}}_\mathrm{n}\right)^2\right]^{1/2}$,

where $\bar{\mathrm{d}}_\mathrm{n} = {(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{d_i})}/{\mathrm{n}}$ is the average degree.

Then, output volatility, measured by the standard deviation of output can be written as:

(10) $(var\ y_{n)})^{1/2} = \Omega\left(\frac{1+CV_n}{\sqrt{n}}\right)$,

where Ω is a parameter that represents the standard deviation of productivity shocks under the assumption that they are independent and identically distributed with mean zero (more details can be found in Annex A).

Equation (10) is useful to illustrate that if the variability of the degree sequence of the input-output network is high, then the variability of aggregate output is high. The intuition tells us that the coefficient of variation is large when only some sectors are responsible for most input supplies in the economy. Shocks to these sectors propagate downstream, resulting in aggregate volatility that decays at a rate slower than $\sqrt{n}$. As Acemoglu, et. al (2012) have shown, this is true when the degree sequence exhibits relatively heavy tails (more details can be found in Annex A).

Such is the case when the economy ε_i exhibits a heavy-tailed power-law degree sequence with a scaling parameter $\beta > 1$. For such kind of distribution, the lower the value of β, the heavier the distribution tails, reflecting larger variations in the degree sequence. If $\beta \in (1{,}2)$, we have,

(11) $(var\ y_{n)}{}^{1/2} = \Omega\left(n^{-(\beta-1)}/_{\beta-\delta}\right)$, where, $\delta > 0$.

Equation (11) relates aggregate volatility to the scaling parameter of the degree distribution. A small β is an indicator of a heavy-tailed degree distribution, meaning that few sectors are disproportionately important as input suppliers, meanwhile many others contribute only a bit as suppliers for the whole economy. In that case, the aggregate volatility of the economy is smaller than with a large β. However, the degree sequence only gives us information about the structure of direct connections between industries. To understand the whole network structure, we need to look at second-order interconnections of the input-output network, because shocks to an industry not only affect its direct clients, but also the downstream customers of those industries.

The effect of second-order interconnections can be summarized by looking at the tail of the second-order degree sequence of the input-output network. In here we define the second-order degree of sector i as the sum of degrees of the industries that uses i as an input weighted by the input shares, that is,

(12) $q_i \equiv \sum_{j=1}^{n} d_j a_{ji}$

If the economy ε_i has a power-law second-order degree sequence with shape parameter $\zeta \in (1,2)$, then aggregate volatility can be written as:

(13) $(var\, y_{n)}{}^{1/2} = \Omega\left(n^{-(\zeta-1)/_{\zeta-\delta}}\right)$ for all $\delta > 0$.

Equation (13) shows that if the distribution of second-order degrees is heavy-tailed, aggregate volatility decreases at a slower rate than predicted by the diversification argument. Acemoglu et al. (2012) show that second-order effects can dominate first-order effects in determining the decay rate of aggregate volatility. For an economy in which the distribution of first and second-order degrees sequences have a power-law tail, the lower bound for the decay rate of aggregate volatility is given by $min\{\beta, \zeta\}$..

2.2 Methods

2.2.1 Estimation procedure

To study empirically if the input-output structure of an economy favors the propagation of microeconomic shocks into aggregate volatility, we need to determine if the first and second-order degree sequence of the intersectoral network is heavy-tailed and can be characterized by a power-law distribution. In practice, it implies fitting a power-law form into the empirical distribution of first and second-order degrees.

There are different ways to estimate the scaling parameters β and ζ of equations ((11)(13) and the lower bound on the scaling region x_{min}. However, the most reliable method is proposed by Clauset, Shalizi, & Newman (2007), because it provides a method to test if the distribution fits with the characteristics of a power-law form[2]. To perform the test the first step is to calculate the node in-degree and out-degree sequences of the production network, following equation (2). In terms of the input-output network, the node in-degree corresponds to the column sum of the input coefficient matrix, and the out-degree to the row-sum of the output coefficient matrix.

Then, we can calculate the second-order out-degree sequence of the input-output network, following equation (12). The focus here is on the role of industries as input suppliers to the economy. Next, we can evaluate the goodness of fit to a power-law distribution of the first and second-order degree sequence, following the methodology proposed by Clauset et al. (2007) that can be summarized in three steps:

1. Estimate the scaling parameter and the scaling range of the power-law distribution.
2. Calculate the goodness-of-fit between the data and the power-law using the Kolmogorov-Smirnov statistic to compare between the empirical data and the model data to obtain a p-value. If the resulting p-value is greater than 0.1 the power-law distribution hypothesis is not rejected

[22] For a detailed discussion see Appendix A of Clauset et al. (2007).

3. Compare the power-law distribution hypothesis with the alternative hypothesis via a likelihood ratio test.

The estimation of the scaling range of the power-law distribution is equivalent to find a lower bound x_{min}, above which the empirical distribution follows a power-law. This step is necessary because very few phenomena exhibit a power-law behavior for all values. The adequate choice of x_{min} guarantees that the estimate of the scaling parameter is not biased due to the inclusion of non-power law data, and that all legitimate data points are included in the estimation. In that way is possible to minimize the statistical error on the scaling parameter and the bias from finite-size effects (Clauset, Shalizi, et al., 2007). To find an optimal x_{min} value the method proposed by Clauset, Young, et al. (2007) is to fit a power law for each unique value in the dataset and select the one that minimizes the Kolmogorov – Smirnov statistic:

(14) $D = \max_{x > x_{min}} |S(x) - P(x)|,$

where $S(x)$ is the Cumulative Distribution Function (CDF) of the observed data with values larger than x_{min}, and $P(x)$ is the CDF for the power-law model that fits the data better in the same region $x \geq x_{min}$. Then we proceed to the estimation of the scaling parameter. Assuming that the data is drawn from a power-law distribution for $x \geq x_{min}$, is possible to calculate a maximum likelihood estimator (MLE) of the scaling parameter:

(15) $\hat{\beta} = 1 = n \left[\sum_{i=1}^{n} \frac{x_i}{x_{min}}\right]^{-1}$ where x_i, $i = 1, \ldots, n$ are observed values such that $x_i \geq x_{min}$.

To quantify the uncertainty on the estimation of x_{min} and the scaling parameter, we use a non-parametric *"bootstrap"* procedure. Given *n* observations, we generate a synthetic data set with a distribution like the original data, by sampling with replacement from the original data set. Then we estimate x_{min} and β for the synthetic data. Many repetitions of the process allow us to derive estimates of the uncertainty in the original estimation, by calculating the standard deviation of the estimates. In this paper, we perform 1500 repetitions of the procedure.

Once we obtain the estimation of a power-law distribution for the data, we need to know if such a model is plausible for the data. Such assessment is done via a goodness-of-fit test, which generates a p-value to quantify the plausibility of the hypotheses:

$$H_0: data\ is\ generated\ from\ a\ power-law\ distribution;$$
$$H_1: data\ is\ not\ generated\ from\ a\ power-law\ distribution$$

To calculate the p-value the first step is to fit the data to a power-law model and calculate the KS statistic. Then we generate a large number (1500 in the exercise below) of synthetic data sets distributed power-law with x_{min} and β equal to those obtained in the first step. Next, we fit a power-law model and calculate the KS statistic for each synthetic data set. Then we compare the KS statistic from the estimation for the synthetic data set, vs the KS obtained with the original data, and count the fraction of times that the KS of the synthetic data is larger than the KS for the original data to obtain the p-value. If the p-value ≤ 0.1 then the power-law distribution is ruled out.

The final step is to compare the power-law distribution versus other alternatives to describe the data. To this end, the likelihood ratio test provides a statistic and a p-value to signal which alternative fits better the data. In this paper, we compare the power-law distribution with a log-normal distribution and an exponential distribution. Consider two competing distributions for the data set with probability density functions $p(x)$ y $q(x)$, we need to compute the likelihood of the data for both distributions and choose the one with the greater likelihood. Then we calculate the ratio of the likelihood:

(16) $R = \prod_{i=1}^{n} \frac{p(x_i)}{q(x_i)}$

And taking logs, the log-likelihood ratio:

(17) $LR = \sum_{i=1}^{n} [ln\, p(x_i) - ln\, q(x_i)]$

And the p-value is an estimate of the probability that we obtain a value of LR ratio, when the true value of LR is close to zero, in other words, the probability of obtaining a useless LR ratio to determine which distribution is better. If the p-value is small ($p \leq 0.1$) the value of LR most likely is not the result of chance, therefore is reasonable to trust it as an indicator of which model is a better fit for the data.

Finally, to calculate the effect of the structure of the I-O network on the volatility of output, we use the estimation of $\hat{\beta}$, and $\hat{\zeta}$ and equations (11) and (13) to calculate the rate of decay of microeconomic shocks in the manufacturing sector in Colombia, and relate them to the projected rate of decay of $1/\sqrt{n}$ under the assumption of an homogeneous distribution in the first and second-order degree sequence.

2.2.2 Data

Data comes from the Colombian Industrial Survey (EAM by the name in Spanish). The EAM is a census of manufacturing plants with 10 or more workers gathering all the variables commonly found in industrial surveys to measure economic activity (production values, sales, employment, intermediate consumption, etc.). A distinctive feature of the EAM is the availability of detailed information describing the use of inputs (consumption in quantities and values, plus transfers among plants) and the kind of products obtained (quantities and values of production and sales) in each plant. The granularity of the data makes it ideal to study the properties of the production network.

In this paper, we use data from 1982 to 2012 of the EAM and variables on sections eight and nine of the survey to build use tables. Products and materials are originally recorded at 7 digits of the ISIC Rv. 2 until 2000, and from 2001 onwards at 7 digits of the Central Products Classification (CPC) v. 1.0. The selection of the period of the analysis is related to the availability of information at the finest level of disaggregation. Before 1982 data on products and materials cannot be used due to problems with the data recording. In 2013 there was a major change in the classification system for products and materials in the survey towards a CPC classification adapted to Colombia version 2.0 which might artificially change the network structural characteristics.

The data describes the list of goods produced in a plant with values of production and quantities for each of them. The list of inputs demanded, and the corresponding value of consumption is included. However, it is not possible to differentiate the origin of inputs, nor the destination of products from the data. In other words, it is not possible to know the proportion of imported inputs that each firm uses for production, and in the same token is not possible to know the proportion of production of a particular good that is exported by each firm. However, this information is not crucial for the analysis in this paper, as we focus on the diversity of connections and the general network structural properties as a source of aggregate volatility. In a broad sense, the origin and destination of inputs and products might be a variable of interest if product-level shocks have geographically bounded effects. In that case for example, local product-level shocks might affect only to locally sourced inputs, generating aggregate volatility only through their connections to other products excluding imported inputs. However, such analysis is out of the scope of this paper that focus on the whole interconnectedness of the I-O network.

The data does not tell us about the proportions in which inputs are used to produce products inside the plant. To overcome the issue, the well-known principle of proportionality can be used to distribute input consumption at the firm level. Each input consumption value is distributed among all products produced at the establishment according to their importance, measured by the share of the product on the production value. Then we aggregate data at the product level to build I-O matrices.

The use matrix must be transformed on product networks, linkages between products are built using input and output coefficients. In an I-O table for each pair of products p an q it is possible to build two output coefficients:

(18) $A_{pq}{}^{O} = \frac{z_{pq}}{z_p}$ Output coefficient pq,

(19) $A_{qp}{}^{O} = \frac{z_{qp}}{z_q}$ Output coefficient qp,

where, Z_{pq}=q's Intermediate consumption of p; Z_{qp}= p's intermediate consumption of q; Z_p= intermediate demand of p.

To calculate the *weighted out- and in-degrees*, we use the output coefficients described above. To understand changes in the structural properties along the period chosen we perform calculations on data every 5-years (1982, 1987, 1992, 1997, 2002, 2007, 2012). The choice of five-years windows is a standard practice in the network literature. The five-year-window allows the researcher to capture structural changes that usually take longer than one or two years to manifest in the data. A longer time-window might be too long to describe the change dynamics on the network over time by capturing only the longer and more permanent changes. To implement all the steps described above to our data, we use the R package *poweRlaw* (Gillespie, 2015).

Results

The analysis of I-O interconnections between products can be performed in two directions: demand and supply. The first direction refers to the sector's power to pull inputs from the economy as intermediates. The indegree metric is useful to measure such demand power for each sector relative to its size. The second direction is focused on input supply, we characterize products according to their power to push inputs into the economy to be used as intermediates in the production of other goods. The outdegree metric is useful to understand the relative importance of each product as an input supplier to the economy, given the production scale in each sector.

3.1 Input demand

We start with the demand direction by looking at the distribution of weighted in-degree values in each year in our data. The number of goods produced by the manufacturing industry in Colombia[3] goes from 2382 in 1982 to 3117 in 2012, reflecting both a natural expansion in the number of categories and types of products produced in the country (due to increasing market specialization and sophistication of consumption) and a small but progressive industrial diversification of the Colombian industry[4]. On average about half of the demand for intermediates is satisfied with locally produced goods. The mean value of indegree averaging over all years is 0.47 and registers only small variations between years. Basic statistics of in-degree values do not suggest a heavy-tailed distribution. The coefficient of variation is below 0.55 and the median is close to the mean in all years (Table 1).

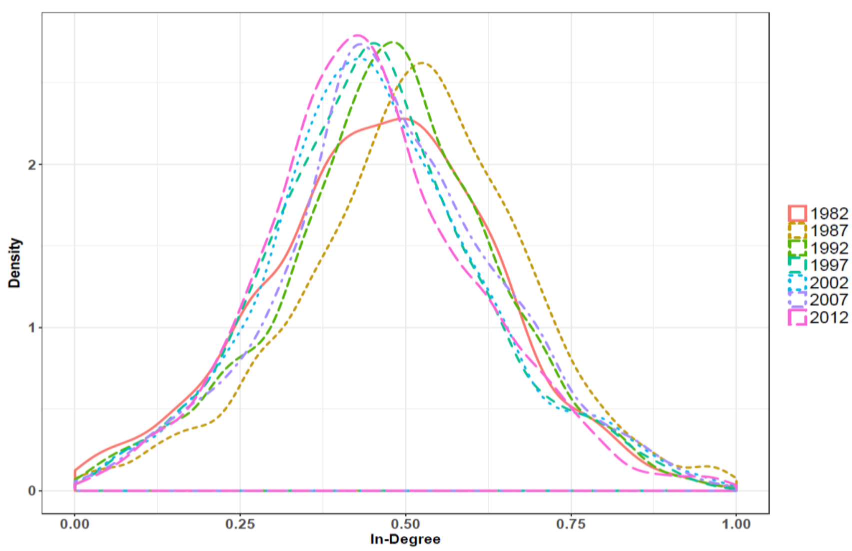
[3] The number of goods corresponds to the number of product codes in the dataset given the ISIC version 2 (adapted for Colombia) classification between 1982 – 2000, and the CPC version 1.0 (adapted for Colombia) classification between 2001 and 2012.
[4] Between 1997 and 2002 there is a classification change, which has an impact on the universe of product codes. This means that some product codes are not comparable between classifications but the structure of intersectoral relationships is stable over time. After 2013 classification changes might translate into artificial changes in the network structure, we took the advice of the technical team at DANE and limited our analysis to the period before 2013.

Table 1. Summary statistics of product weighted in-degree by year.

Year	# Products	Mean	Standard deviation	Median	Coefficient Variation
1982	2382	0.4608	0.1748	0.4656	0.3792
1987	2732	0.5127	0.1728	0.5196	0.3370
1992	2946	0.4762	0.1696	0.4779	0.3561
1997	2943	0.4542	0.1685	0.4487	0.3710
2002	2928	0.4572	0.1702	0.4460	0.3724
2007	3041	0.4890	0.2630	0.4669	0.5378
2012	3117	0.4551	0.2501	0.4364	0.5496

The non-parametrical estimates of the empirical density of in-degree (Figure 1) show that product in-degree values are concentrated around the mean. Between 1982 and 2012 we observe an increased concentration of values around the mean and a shift to the left of the distribution, due to a general decrease of in-degree values. In other words, industrial goods demand relatively less locally produced inputs.

Figure 1. Empirical densities of product weighted in-degree by year.



3.2 Input supply

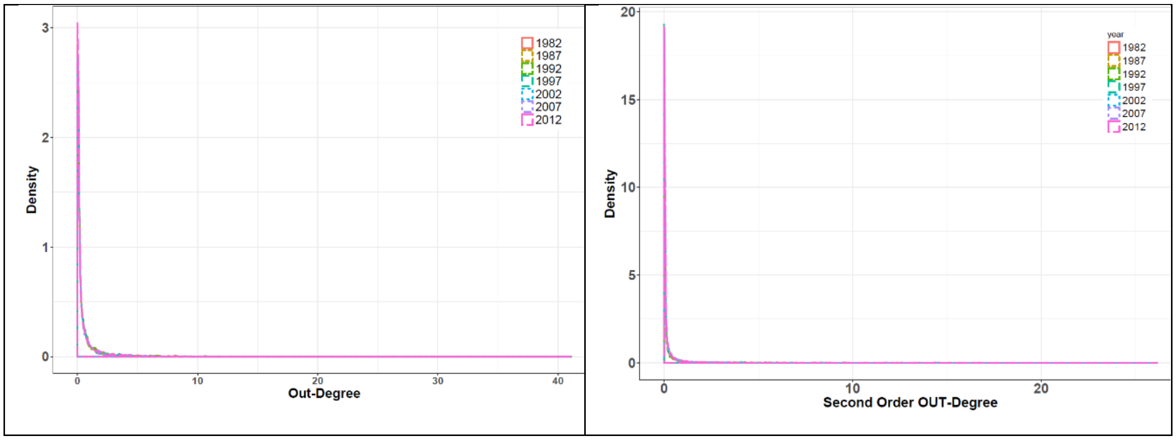
Now we turn to the supply side of product interrelationships. The number of inputs varieties used by the manufacturing industry in Colombia increased from 1987 in 1982 to 2962 in 2012. The increase is related to a more sophisticated choice of intermediates by local producers and a natural increase in the number of available input varieties. On average about half of the total product is used as an intermediate input in the production of other goods with a mean value of 0.53 of out-degree averaging across years. A lower value of the median compared to the average indicates already an overrepresentation of less connected products in the network (Table 2). The out-degree sequence of the I-O network exhibits a high variability measured by the coefficient of variation, which is larger than 2.8 in all years.

Table 2. Summary statistics of product weighted out-degree by year.

Year	# Products	Mean	Standard deviation	Median	Coefficient Variation
1982	1987	0.552	1.564	0.087	2.831
1987	2227	0.629	1.959	0.085	3.114
1992	2503	0.560	1.665	0.080	2.970
1997	2600	0.514	1.468	0.078	2.856
2002	2737	0.489	1.419	0.069	2.902
2007	2854	0.521	1.533	0.067	2.942
2012	2962	0.479	1.355	0.067	2.829

An examination of the empirical density of the first- and second-order out-degree sequence, confirms that both distributions are heavy-tailed and skewed (Figure 2). A very skewed out-degree distribution means that some inputs are used by many other sectors, also known as “general-purpose inputs”, accompanied by many other inputs that only provide to a few other sectors. This feature of the first- and second-order out-degree sequence is very stable over time.

Figure 2. Empirical densities of first- and second-order out-degrees.



Summary statistics of the second-order out-degree distribution show that the second-order out-degree sequence is more skewed and has a heavier right tail than the first-order out-degree distribution (Table 3). Meaning that there is a large variability in the strength and number of product-to-product connections, featuring some inputs with multiple and strong connections to other sectors and many other inputs with few and/or weak connections to other products. Additionally, well-connected inputs tend to be connected to other well-connected inputs meanwhile less connected inputs are connected to other poorly connected goods.

Table 3 Summary statistics of second-order out-degrees.

Year	# Products	Mean	Standard deviation	Median	Coefficient Variation
1982	1802	0.199	0.864	0.007	4.352
1987	2040	0.262	1.086	0.010	4.151
1992	2275	0.234	0.976	0.007	4.167
1997	2417	0.205	0.921	0.007	4.498
2002	2583	0.165	0.735	0.007	4.449
2007	2649	0.180	0.833	0.007	4.627
2012	2780	0.162	0.683	0.007	4.229

A visual examination of the distribution of first- and second-order out-degree sequences suggests that the power-law might be a good fit for the tail distribution. Figure 3 and Figure 4 plot the counter-cumulative distribution functions (CCDF) of the first- and second-order out-degrees on a log-log scale. In both cases, we observe a rather linear relationship at the right tail of the distribution. To further characterize the heavy tail behavior of the out-degree distribution, we proceed to perform a power-law estimation for the data following the method described in section 2.2.

Figure 3. Empirical counter-cumulative distribution function of first-order out-degrees.

Figure 4. Empirical counter-cumulative distribution function of first-order out-degrees

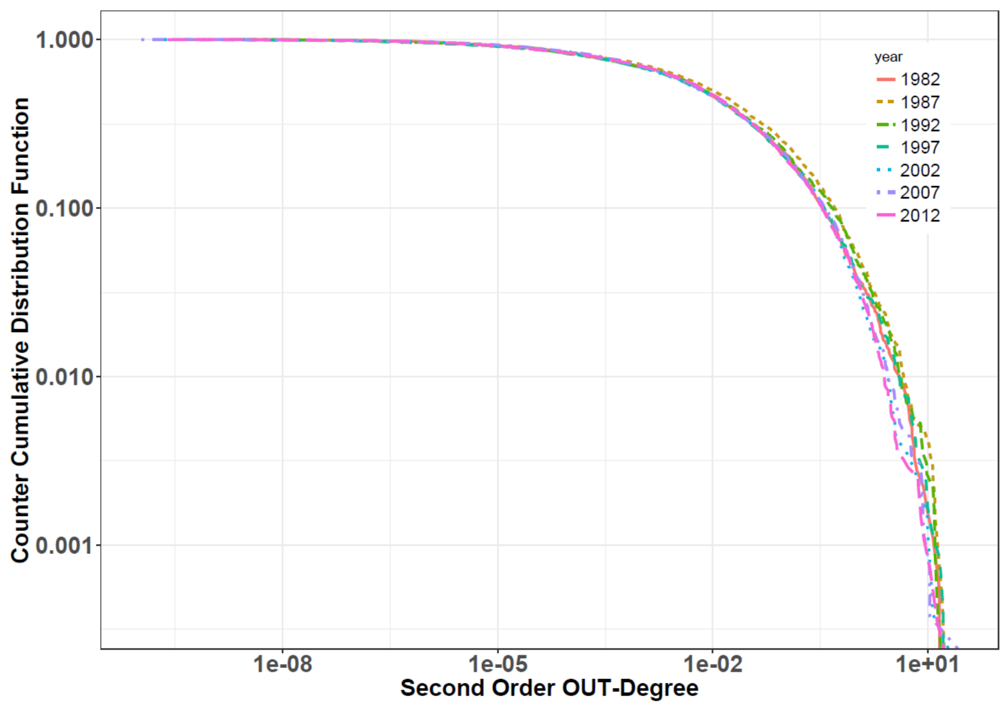


Table 4 shows the results of the power-law test including a) estimates of the power-law distribution: the lower bound ($\hat{x}_{min}$), the tail size (n-tail), the scaling parameter ($\hat{\beta}$), the uncertainty of the scaling parameter (standard deviation of $\hat{\beta}$), and the p-value from the goodness-of-fit test for the power-law; b) The log-likelihood ratio (LR) and p-value from the likelihood test to compare the likelihood of the data under power law vs the log-normal distribution, and c) the LR and p-value from the likelihood test to compare the likelihood of the data under power law vs the exponential distribution.

A first look at the results suggests that there are large differences between products concerning the role they play as input suppliers for other industrial sectors. The size of $\hat{\beta}$ confirms the existence of a heavy tail in the outdegree distribution in all years. However, the existence of a heavy tail does not imply that the power law distribution is the best fit for the data. In fact, according to the p-value from the goodness-of-fit test, there is evidence of a good fit from 1997 onwards. In 1982, 1987, and 1992 the results of the likelihood test point towards the log-normal distribution as a better fit for the data. It is worth noticing that the lower bound of the fat-tail is larger from 1997 to 2012, compared with previous years. In our case, that means that the number of products behaving as large input providers is smaller in the last four years in our data than in previous years. In other words, after 1997, some inputs have become very central as direct providers for the rest of the manufacturing sector, increasing their importance for industrial production.

Table 4. Power law test for first-order out-degree by year.

	Power Law					Log-normal		Exponential	
Year	***n-tail***	$\hat{x}_{min}$	$\hat{\beta}$	***SD***	***p-value***	***LR***	***p-value***	***LR***	***p-value***
1982	235	1.13	2.20	0.340	0.014	-2.240	**0.025**	0.958	0.338
1987	264	1.28	2.22	0.307	0.096	-1.830	**0.068**	2.030	**0.042**
1992	328	1.01	2.14	0.398	0.010	-2.820	**0.005**	1.320	0.187
1997	100	3.31	2.96	0.532	**0.591**	-0.571	0.568	1.190	0.232
2002	69	3.99	2.99	0.396	**0.624**	-0.586	0.558	0.634	0.526
2007	101	3.48	2.95	0.486	**0.680**	-0.573	0.567	1.260	0.202
2012	106	3.10	2.91	0.551	**0.394**	-0.878	0.380	0.788	0.431

Now we turn to the second-order out-degree distribution fit of a power law. Results in Table 5, show that the second-order degree sequence is right-skewed. The average value of $\hat{\zeta}$ is 2.17, which is lower than the average of $\hat{\beta}$, 2.62. Uncertainty over the parameter estimation is lower as well for the second-order degree power-law estimation than for the first-order degree. The p-value of the goodness-of-fit test indicates a good fit of the functional form to the data for 2002, 2007, and 2012. There is moderate evidence in favor of the power law in 1987 and 1992 because the log-normal appears as a possible candidate in the likelihood test. The likelihood test comparing the power-law vs the exponential distribution provides support in favor of the power-law in 1982 and 1997 and confirms the result of the goodness-of-fit test for all other years. In summary, is possible to affirm that the second-order degree distribution of the data can be described as a power law.

Table 5. Power law test for second-order out-degree by year.

	Power Law					Log-normal		Exponential	
Year	***n-tail***	$\hat{x}_{min}$	$\hat{\zeta}$	***SD***	***p-value***	***LR***	***p-value***	***LR***	***p-value***

1982	152	0.45	1.98	0.178	0.082	-1.590	0.112	2.300	**0.022**
1987	210	0.49	1.99	0.176	**0.123**	-1.740	**0.083**	3.100	**0.002**
1992	220	0.50	2.02	0.234	**0.177**	-1.830	**0.067**	2.960	**0.003**
1997	212	0.44	1.99	0.220	0.053	-1.580	0.114	3.180	**0.001**
2002	227	0.42	2.17	0.194	**0.111**	-1.070	0.283	3.040	**0.002**
2007	112	0.96	2.35	0.245	**0.455**	-0.754	0.451	1.760	**0.079**
2012	67	1.47	2.71	0.431	**0.709**	-0.098	0.922	2.180	**0.030**

As Acemoglu et al. (2012) shows, for an economy in which the distribution of first and second-order degrees sequences have a power-law tail, the lower bound for the decay rate of aggregate volatility is given by $min\{\beta, \zeta\}$. Therefore, we can expect that the second-order degree dominates the network effect of the first-order degrees in the case of the manufacturing sector in Colombia. To illustrate the effect on aggregate volatility of the heavy-tailed first- and second-order degree sequences, Table 6 displays estimations of the rate of decay of product level shocks (see equations (11) and (13)), versus the rate of decay predicted by classical macroeconomics ($1/\sqrt{n}$), and the ratio between the two measures.

Table 6. Estimated effect on output volatility of the first- and second-order outdegree distribution.

	First-Order Out-Degree			**Second-Order Out-Degree**		
Year	$\|\|\boldsymbol{\lambda}\|\|_2$	$\mathbf{1/\sqrt{n}}$	$\|\|\boldsymbol{\lambda}\|\|_2 / \mathbf{1/\sqrt{n}}$	$\|\|\boldsymbol{\lambda}\|\|_2$	$\mathbf{1/\sqrt{n}}$	$\|\|\boldsymbol{\lambda}\|\|_2 / \mathbf{1/\sqrt{n}}$
1982	0.051	0.022	2.269	0.083	0.024	3.532
1987	0.047	0.021	2.203	0.070	0.022	3.159
1992	0.046	0.020	2.286	0.066	0.021	3.131
1997	0.047	0.020	2.416	0.070	0.020	3.422
2002	0.060	0.019	3.125	0.054	0.020	2.727
2007	0.047	0.019	2.528	0.066	0.019	3.422
2012	0.047	0.018	2.550	0.070	0.019	3.713

In both cases is evident that intersectoral linkages increase the impact of microeconomic shocks. If we take only direct linkages into account, there is an increase of roughly 2.5 times in the expected impact of a product level shock into aggregate volatility. However, taking the second-order degree into account, the impact increases to 3.3 times on average. Therefore, the role of the I-O network structure for the transmission of product-level shocks on industrial output volatility is not negligible. According to this result, we can expect that shocks at the product level will propagate in the economy, thanks to the heterogeneous structure of the industrial I-O network, affecting the manufacturing output volatility.

It is hard to compare the results obtained in this exercise with similar attempts made for other countries. No other exercise has been made at the level of disaggregation reached in this paper, and we cover only the manufacturing industry, meanwhile, others include aggregated activities in all economic sectors. However, it is worth noticing that the results reached in this paper are consistent with findings in the related literature. For example, estimations of $\hat{\beta}$ and $\hat{\zeta}$ at two and three-digit levels of ISIC for the I-O tables of Ecuador between 1975 and 2012, obtain average values of 2.27 and 1.87, and calculate an increase in the impact of the I-O structure on aggregate volatility of 1.67 on average (Romero et al., 2018).

These numbers make sense in the context of a country like Colombia if we consider that at more disaggregated levels is possible to capture better the magnitude of the effect of the network structure on aggregate volatility (Acemoglu et al., 2012), in consequence, we can expect lower values of decay with more aggregated data. Moreover, the fit of the power-law needs a large amount of data to give unbiased estimates (Clauset, Shalizi, et al., 2007), so we can expect a better estimation for the parameters with our data.

Conclusions

In this paper, we characterize the network structure of the industrial I-O network in Colombia between 1982 and 2012. We find that from the supply point of view, the I-O industrial network is highly heterogeneous. On one hand, input suppliers diverge greatly in their importance as providers of intermediates to produce other industrial products. Moreover, the central role of some inputs as suppliers for the rest of the industry has been increasing after 1997. This is an industry where few products are important suppliers, meanwhile, many other products have a meager role as input suppliers. This feature is spotted in the data in the form of right-skewed distributions of first-order out-degrees.

On the other hand, there are clusters of big suppliers in the Colombian manufacturing industry. Central input providers sell to other big input suppliers, producing clusters of central products. In other words, there are few central value chains connected among them and some other disconnected chains on the periphery. This characteristic of the I-O network is revealed by the power-law behavior on the tail of the second-order out-degree distribution.

As predicted by the theory, this is the perfect structure to propagate and amplify microeconomic shocks that can affect the product supply or demand, to generate aggregate volatility. We estimate that the effect of the heterogeneous I-O product network structure on industrial product volatility is on average 3.3 times larger than what one could expect in a balanced structure.

This research reveals a structural weakness of the Colombian industry structure, related to the high level of concentration of economic activity in some products, that demand inputs preferentially from each other at the same time. In such a structure, idiosyncratic shocks that affect central products will have effects on a large proportion of sectors that demand such inputs, creating a cascade effect on industrial output.

This paper is just the first step in the understanding of sources of aggregate volatility related to the economic structure in Colombia. A natural step forward is to further analyze the effect of aggregation on the estimation of fat tails and the impact on the industrial output volatility. Once the amount of reduction on the estimates at higher levels of aggregations is understood, is possible to extend the analysis to the whole economy using the existent use tables. On a more disaggregated level, one alternative is to exploit data at the firm level to understand the effect of firm-level shocks in aggregate volatility. To further understand the role of different product-level shocks on the generation of output volatility, the use of trade data at firm level in combination with industrial survey data is an interest venue to understand the role of trade on the network structure and more broadly on the generation of output volatility (Joya & Rougier, 2019; Mundt, 2021).

A. Annex: Macroeconomic effects

To understand the macroeconomic implications of Equation (8), recall the equilibrium vector of relative prices (Equation (6)), multiply both sides by β_i and sum over the i sectors to obtain:

$$\log(GDP) = \sum_{i,j=1}^{n} \beta_i l_{ij} \epsilon_j + \sum_{i=1}^{n} \beta_i \ \log(p_i),$$

where $\sum_{i=1}^{n} \beta_i \ \log(p_i) = 0$ if we normalize the ideal price index $P_c = \prod_{i=1}^{n} P_i^{\beta_i}$ to 1.

Then, the economy's (log) real value added is:

$$\log(GDP) = \sum_{i,j=1}^{n} \beta_i l_{ij} \epsilon_j = \sum_{i=1}^{n} \lambda_i \epsilon_j$$

This equation shows that (log) aggregate output is a linear combination of sector-level productivity shocks with Domar weights as coefficients. Therefore, the sector Domar weight is a statistic of the impact of sector-level shocks on aggregate output. At the same time, the equation shows that Domar weights depend on the economy's production network, an increase in l_{ji}, increases λ_i, meaning a larger impact of a sector's i shock on aggregate output.

To illustrate the relationship between sector-level shocks, the production network, and aggregate volatility, which is the interest variable, two assumptions are useful: a) log productivity shocks $\epsilon_i = \log(z_i)$ are independent and identically distributed with mean zero and standard deviation σ, and b) the share of labor is the same across all sectors $\alpha_i = \alpha$

for all i. In consequence, output volatility measured by the standard deviation of aggregate output is given by:

$$(var\ y_{n)})^{1/2} = \sqrt{\sum_{i=1}^{n} \sigma_i^2 \lambda_i^2},$$

given the assumptions stated above, this equation is equivalent to:

$$(var\ y_{n)})^{1/2} = \sigma\sqrt{\sum_{i=1}^{n} \lambda_i^2} = \sigma\|\lambda\|_2$$

Since $\sum_{i=1}^{n} \lambda_i = \sum_{i,j=1}^{n} \beta_i l_{ij} = 1/\alpha$, the last equation can be written as:

$$(var\ y_{n)})^{1/2} = \frac{\sigma/\alpha}{\sqrt{n}}\sqrt{1 + n^2\sigma^2}var(\lambda_1, \dots, \lambda_n)$$

This equation illustrates the relationship between sector heterogeneity, measured by the variance of Domar weights and output volatility. The larger the variance of Domar weights, all else equal, the larger the output volatility of the economy. In particular, when all Domar weights are identical, output volatility is proportional to $1/\sqrt{n}$, consistent with the diversification argument (Lucas, 1977).

To understand the relationship with first-order interconnections, take an economy ε_i with a degree sequence $\{d_{1,}d_2, \dots, d_\mathrm{n}\}$ the coefficient of variation is:

$$\mathrm{CV_n} \equiv \frac{1}{\mathrm{\bar{d}_n}}\left[\frac{1}{(\mathrm{n}-1)}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{d_i} - \mathrm{\bar{d}_n}\right)^2\right]^{1/2}$$

Where $\mathrm{\bar{d}_n} = {}^{(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{d_i})}/_{\mathrm{n}}$ is the average degree. Then:

$$(var\ y_{n)})^{1/2} = \sigma\|\lambda\|_2 = \Omega\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} d_i^2}\right)$$

and

$$(var\ y_{n)}{}^{1/2} = \Omega\left(\frac{1 + CV_n}{\sqrt{n}}\right)$$

In particular, if the economy ε_i has a power-law degree sequence, meaning that there is a constant $\beta > 1$, a constant c, and $k < d_{max} = \Theta\left(n^{1/\beta}\right)$, such that

$$P(k) = ck^{-\beta}$$

Where $P(k) \equiv \frac{1}{n}|\{d_i > k\}|$is the empirical counter-cumulative distribution function and d_{max} is the maximum degree of ε_i. The parameter $\beta > 1$, is the scaling parameter. If $\beta \in (1{,}2)$, we have,

$$(var\ y_{n)}{}^{1/2} = \Omega\left(n^{-(\beta-1)/\beta-\delta}\right) \qquad \text{Where } \delta > 0.$$

Similarly if the economy ε_i has a power-law second-order degree sequence with shape parameter $\zeta \in (1{,}2)$, then aggregate volatility can be written as:

$$(var\ y_{n)}{}^{1/2} = \Omega\left(n^{-(\zeta-1)/\zeta-\delta}\right) \qquad \text{for all } \delta > 0.$$

For further details and proof refer to Acemoglu et al. (2012).

Bibliography

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, *80*(5), 1977–2016. https://doi.org/10.3982/ECTA9623

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2017). Microeconomic Origins of Macroeconomic Tail Risks. *American Economic Review*, *107*(1), 54–108. https://doi.org/10.1257/aer.20151086

Aizenman, Joshua, & Marion, N. (1999). Volatility and Investment: Interpreting Evidence from Developing Countries. *Economica*, *66*(262), 157–1179. https://doi.org/10.1111/1468-0335.00163

Aizenman, Joshua, & Pinto, B. (2005). *Managing Economic Volatility and Crises* (Joshua Aizenman & B. Pinto (eds.)). Cambridge University Press. https://doi.org/10.1017/CBO9780511510755

Atalay, E. (2017). How Important Are Sectoral Shocks? *American Economic Journal: Macroeconomics*, *9*(4), 254–280. https://doi.org/10.1257/mac.20160353

Badinger, H. (2010). Output volatility and economic growth. *Economics Letters*, *106*(1), 15–18.

Baldwin, R., & Weder di Mauro, B. (2020). *Economics in the Time of COVID-19* (VoxEU.org (ed.)). CEPR Press.

Blöchl, F., Theis, F. J., Vega-Redondo, F., & Fisher, E. O. (2011). Vertex centralities in input-output networks reveal the structure of modern economies. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *83*(4 Pt 2), 046127. http://www.ncbi.nlm.nih.gov/pubmed/21599260

Breen, R., & Garcia-Penalosa, C. (2005). Income Inequality and Macroeconomic Volatility: An Empirical Investigation. *Review of Development Economics*, *9*(3), 380–398. https://doi.org/10.1111/j.1467-9361.2005.00283.x

Calderon, C., & Yeyati, E. L. (2009). *Zooming in : from aggregate volatility to income distribution*. *April*. http://documents.worldbank.org/curated/en/2009/04/10421833/zooming-aggregate-volatility-income-distribution

Carvalho, V. M., & Tahbaz-Salehi, A. (2019). Production Networks: A Primer. In *Annual Review of Economics* (Vol. 11, Issue 1). https://doi.org/10.1146/annurev-economics-080218-030212

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2007). *Power-law distributions in empirical data*. https://doi.org/10.1137/070710111

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the Frequency of Severe Terrorist Events. *Journal of Conflict Resolution*, *51*(1), 58–87. https://doi.org/10.1177/0022002706296157

del Rio-Chanona, R. M., Mealy, P., Pichler, A., Lafond, F., & Farmer J., D. (2020). Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. *Covid Economics*, *1*(6), 65–103.

Gabaix, B. Y. X., Basu, S., Bénabou, R., Blanchard, O., Caballero, R., Canning, D., Caplin, A., Chaney, T., Chari, V. V, Christiano, L., Comin, D., Davis, D., Durlauf, S., Edmans, A., Eichenbaum, M., Engel, E., & Fernald, J. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, *79*(3), 733–772. https://doi.org/10.3982/ECTA8769

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The powerlaw package. *Journal of Statistical Software*, *64*(2), 1–16. https://doi.org/10.18637/jss.v064.i02

Gonçalves, J., Matsushita, R., & Da Silva, S. (2020). The asymmetric Brazilian input–output network. *Journal of Economic Studies*. https://doi.org/10.1108/JES-05-2020-0225

Hakura, D. (2007). Output Volatility and Large Output Drops in Emerging Market and Developing Countries. *IMF Working Papers*, *07*(114), 1. https://doi.org/10.5089/9781451866780.001

Jaimovich, N., Pruitt, S., & Siu, H. E. (2013). The Demand for Youth: Explaining Age Differences in the Volatility of Hours. *American Economic Review*, *103*(7), 3022–3044. https://doi.org/10.1257/aer.103.7.3022

Joya, O., & Rougier, E. (2019). Do (all) sectoral shocks lead to aggregate volatility? Empirics from a production network perspective. *European Economic Review*, *113*, 77–107. https://doi.org/10.1016/j.euroecorev.2019.01.004

Laursen, T., & Mahajan, S. (2005). Volatility, income distribution and poverty. In J. Aizenman & B. Pinto (Eds.), *Managing Economic Volatility and Crises: A Practitioner's Guide.* (pp. 101–136). Cambridge University Press.

Long, J. B., & Plosser, C. (1983). Real Business Cycles Author. *Journal of Political Economy*, *91*(1), 39–69.

Lucas, R. E. (1977). Understanding Business Cycles. In *Carnegie–Rochester Conference Series on Public Policy* (No. 5).

McNerney, J., Fath, B. D., & Silverberg, G. (2013). Network structure of inter-industry flows. *Physica A: Statistical Mechanics and Its Applications*, *392*(24), 6427–6441. https://doi.org/10.1016/j.physa.2013.07.063

Mundt, P. (2021). The formation of input–output architecture: Evidence from the European Union. *Journal of Economic Behavior and Organization*, *183*, 89–104.

https://doi.org/10.1016/j.jebo.2020.12.031

Newman, M. (2018). *Networks* (Second Edi). Oxford University Press. https://global.oup.com/academic/product/networks-9780198805090?cc=co&lang=en&#

OECD/World Trade Organization. (2019). Promoting economic diversification and structural transformation through industrialisation. In *Aid for Trade at a Glance 2019: Economic Diversification and Empowerment* (pp. 81–108). OECD Publishing. https://doi.org/10.1787/785f021c-en

Ramey, G., & Ramey, V. A. (1995). Cross-Country Evidence on the Link Between Volatility and Growth. *The American Economic Review*, *85*(5), 1138–1151. http://www.jstor.com/stable/2950979

Romero, P. P., López, R., & Jiménez, C. (2018). Sectoral networks and macroeconomic tail risks in an emerging economy. *PLoS ONE*, *13*(1), 1–17. https://doi.org/10.1371/journal.pone.0190076

Sahay, R., & Goyal, R. (2006). Volatility and Growth in Latin America: An Episodic Approach. *IMF Working Papers*, *06*(287), 1. https://doi.org/10.5089/9781451865479.001

Szirmai, A. (2012). Industrialisation as an engine of growth in developing countries, 1950–2005. *Structural Change and Economic Dynamics*, *23*(4), 406–420. https://doi.org/10.1016/j.strueco.2011.01.005