GRANULARITY, NETWORK ASYMMETRY,
AND AGGREGATE VOLATILITY

JESSE TWEEDLE

ABSTRACT. I evaluate two competing theories for microfoundations of aggregate fluctuations. The network hypothesis suggests industry-level shocks propagate across the input-output (IO) network of the economy, resulting in aggregate fluctuations. The granular hypothesis suggests idiosyncratic shocks to very large firms result in aggregate fluctuations. My main contribution is to connect the two aggregate fluctuation hypotheses for the first time and theoretically and empirically quantify the contributions of each to volatility.

The network hypothesis depends crucially on certain plants being essential suppliers to the economy. However, they may be essential suppliers due to their productivity and not any underlying input-output requirements, which means productivity may be the source of both the granularity and network hypotheses. To disentangle these relationships, I document a plant-plant input-output network, then develop a model in which productivity and the exogenous IO network can vary independently and both combine to determine the observed IO network. Finally, I calibrate the model to uncover the underlying IO network and then investigate the empirical relationship between the uncovered IO network and aggregate volatility.

I find (i) the observed plant-plant IO network is very asymmetric, (ii) productivity doesn’t vary enough to explain the observed IO network, (iii) and therefore the true underlying IO network explains the majority of the plant size distribution and 34% of aggregate volatility.

1. Introduction

Why are microeconomic shocks sources of aggregate volatility and how do they propagate across the economy? Are shocks transmitted across input-output linkages or not? The answer would seem to depend on two competing theories of microfoundations of aggregate fluctuations: the granularity hypothesis of Gabaix [19] and the unbalanced network hypothesis of Acemoglu, et al. [2]. If granular plants are the sources of aggregate fluctuations, then plants should account for the majority of fluctuations, independent of the input-output (IO) network. However, at the plant level, the two theories are intertwined—plants may be essential suppliers in the IO network because of their granularity. I aim to document and explore the relationship between granularity and IO networks and how they contribute to aggregate volatility.

Date: June 11, 2016.
There are several dimensions of interaction between the two theories. First, the conceptual difference between them depends on the reason for the shape of the individual size distribution, be they plants or sectors. The granularity hypothesis is agnostic about the underlying cause of the shape of the distribution, and one typically assumes that a fat-tailed productivity distribution is responsible (e.g., in a standard Melitz model). The network hypothesis, on the other hand, claims the fat-tailed size distribution is caused by an exogenous asymmetry in the IO network, so that certain sectors are very large because they supply an inordinately large portion of the economy. The insight I add is to let productivity and the IO network vary independently at the plant level, and explain, with theory and data, how the size distribution is shaped by these two primary forces, and how that affects aggregate volatility.

Second, plants and sectors are typically treated very differently in economic models and data. How does that affect the argument for microfoundations of aggregate fluctuations? Both the granular and network hypotheses require the number of microeconomic units to be very large (otherwise there would be no micro to provide foundations for). However, in many models and data, sectors are the only ones with IO networks, while plants are assumed to have differences in productivity but no variation in IO characteristics. This presents a problem for the unbalanced network hypothesis: plant networks can’t be sources of aggregate fluctuations if there is no variation within sectors. On the other hand, sector-level models take expenditure shares as exogenous, implying no productivity or heterogeneity can affect the network itself.

In theory, there is no difference between sectors and plants—it is easy to include plant specific IO characteristics and reformulate a model with sectors into a model with plants with the same behaviour. This makes the distinction between plants and sectors an empirical one, in the sense that the deepest level we can study IO networks is the level at which data on inputs and outputs are recorded, which are typically industries. Put another way, the only way to explore the relationship between unbalanced IO networks and granularity at the plant level is to have data on inputs and outputs at the plant level. I combine theory and
data on granularity and an asymmetric IO network at the plant level, and show how this translates into both hypotheses coexisting at the most disaggregated level in the economy.

Putting the pieces together, I compare the properties of each plant, granularity and the IO network, and explore empirically how these factors affect aggregate volatility.

I use four crucial pieces of data and theory to explore these relationships. First, plant level data on commodity inputs and outputs to establish the unbalanced IO network at a much more disaggregated level than previous studies. Second, I use theory to identify conditions under which productivity and the underlying IO network determine the endogenous observed IO network. Third, I calibrate the model to match observed plant characteristics and uncover the underlying IO network. Fourth, I measure each plant’s importance to the IO network measure and link it to its contribution to aggregate volatility.

I use the Annual Survey of Manufactures (ASM), a long-term establishment-level survey in Canada, covering 99% of output and value-added. The ASM comes with detailed data on commodity inputs and outputs for each plant, crucial to exploring the disaggregated IO network. Using these data, I construct plant-to-plant direct-requirements tables, in the tradition of industry-level input-output accounting at statistical agencies. The ASM has the relevant data on other plant-level characteristics, including industry, location, sales, value added, and employment.

To disentangle the two forces shaping the observed unbalanced IO network—is a plant a central supplier because it supplies an essential product or because it is so productive that every plant substitutes toward it? The endogenous IO network depends on those two primary factors, the productivities of individual plants and the unobserved plant-to-plant supply linkages. I extend the standard Cobb-Douglas input-output model to accommodate productivity differences and substitutability across plants (both within and across sector boundaries), which induces productive plants to become more central suppliers. The key to differentiating between productivity and network asymmetry is the behaviour of the tails of each distribution, and how they affect the tail of the size distribution. Recall that the argument for microfoundations of aggregate volatility depends on the fat tail of the size
distribution as the number of plants in the economy gets very large. If productivity and network centrality are both distributed with power laws, I show that, as the number of plants gets large enough to apply the microfoundation argument, the fatter of the two tails will determine the tail of the size distribution. However, as in many applications of power laws, the empirics are more complex and both factors will matter.

Research on idiosyncratic shocks and aggregate volatility restarted in earnest when Gabaix [19] and Acemoglu et al. [2] revived the debate between Horvath [23, 24] and Dupor [14] on whether idiosyncratic shocks average out in aggregate. Gabaix [19] proposes that the largest, granular firms are so big that their idiosyncratic shocks do not average out at the aggregate level. Acemoglu et al. [2] suggest the reason for non-diversification of idiosyncratic shocks is an asymmetric input-output network, in which a shock to a sector that supplies a large number of other sectors propagates through the economy and generates aggregate fluctuations. I add an understanding of the connections between the two theories at an empirical level, specifically showing the complementarity between granularity and production networks and how idiosyncratic plant-level shocks rely on plant-level IO variation within industries. What really differentiates this work is that I explore the determinants of the observed IO network, whereas previous research assumes the network is exogenous.

The most direct predecessors of this paper are empirical studies of aggregate fluctuations. Starting with Shea [34], and continuing most recently with Di Giovanni, Levchenko and Méjean [12], Foerster, Sarte and Watson [16], Acemoglu et al. [1]. Foerster, Sarte and Watson [16] combined factor analysis with structural model of industrial production in the US, finding common shocks are the source of the majority of volatility, with idiosyncratic shocks becoming more important after the great moderation. Di Giovanni, Levchenko and Méjean [12] study fluctuations of French firm sales to individual countries and find idiosyncratic fluctuations account for the majority of aggregate volatility, and that much of it comes from covariances between firms. They suggest the firm covariances are due to firm-to-firm linkages, although they only observe industry-level IO data. In contrast to both papers, I use plant-level IO data to establish the determinants of plant covariances, using deeper levels of
disaggregation to examine both covariances (firm level to plant level) and IO (industry level to plant level). As well, I study the determinants of the network itself, something taken as exogenous in previous empirical work.

Any study of granularity builds on a body of work on the determinants of firm size and the characteristics of its distribution, from specific applications in international trade [13, 10, 11], or studies on general characteristics and theories of the size distribution itself [31]. I add an endogenous network perspective to this research and use it to further explore the determinants of the plant size distribution and the sources of granularity. My work also fits naturally with Hottman et al. [25], who use detailed price and sales data on consumer non-durables to suggest ‘firm appeal’ is the dominant source of firm heterogeneity, accounting for 50 – 70% of firm size. Holmes and Stevens [22] also provide evidence that demand characteristics are the main source of plant heterogeneity, in contrast to standard Melitz applications. In my case, the IO requirements of downstream plants translate into a dominant source of plant appeal, and therefore are a large determinant of plant size.

My argument is also related to recent work on customer-supplier relationships, especially Barrot and Sauvagnat [6], who study the disruption of production networks after natural disasters. In addition, research on customer-supplier relationships in Japan [7, 8] and the US [4] suggests larger plants have different input-output characteristics than smaller plants. Typically, customer-supplier relationship data only includes an indicator for whether a firm supplies another firm, not the strength of the relationship or the commodities made and used. In my case, I have measures of the strength of the interaction between plants. To this research, I add a characterization of the manufacturing IO network in Canada, focusing on differences across plants within industries.

These papers are also part of a recent wave of interest in the formation and effects of social and economic networks. Carvalho and Voigtlander [9], Oberfield [33] and Jones [29] each apply these ideas specifically to production and growth, whereas other works focus on volatility and contagion in financial markets, such as Acemoglu et al. [3], Golub, Elliot,
Jackson [15]. Other applications and background on several network measures used in this paper can be found in Jackson [27].

In Section 2, I present the plant-level volatility and IO data. I document an unbalanced IO network at a disaggregated level, with a few plants acting as central suppliers to the network. In Section 3, I present a simple, but necessary, extension to the IO model used in Acemoglu et al. [2] to allow plant IO characteristics to vary independently of productivity. The asymmetry of the network and the productivity distribution combine to determine plant sizes, which is the key to evaluating the granularity of the economy and its effect on aggregate volatility. In Section 4, I outline the asymptotic theory that gives a knife-edge prediction of the cause of granularity: the thicker tail of the distributions of productivity and network asymmetry are the sole cause of skewed firm size distribution.

In Section 5, I calibrate the model to uncover the underlying IO network from the endogenous, observed IO network and evaluate the competing theories of aggregate fluctuations. Previewing the main calibration results, the productivity distribution is not heterogenous enough to account for the asymmetry in the observed IO network. The majority of the observed IO network is due to the underlying IO network, consistent with results in Holmes and Stevens [22] that challenge the reliance of the plant size distribution on productivity alone.

In Section 6, I provide direct empirical support for the importance of both productivity and network asymmetry for determining granularity and aggregate volatility. An 10% increase in network centrality is associated with a 2.66% increase in plant size, and a 10% increase in productivity is associated with an 8% increase in plant size. Eliminating any asymmetry in the plant-plant IO network reduces aggregate volatility by 34%.

Section 7 concludes, and several Appendices follow, giving details on theory, measurement and development of the plant-plant IO network, and other sundry details.
2. Data

2.1. Overview. The data is from the Annual Survey of Manufactures (ASM), which covers 99% of industrial output in Canada. It is a long-running annual panel of manufacturing establishments, including information on all relevant industrial characteristics, including sales, value added, total intermediate inputs, location, employment, industry, and parent firm. I analyze the period from 1973 to 1999, covering several volatile periods in Canadian manufacturing, including recessions and recoveries in the 1980s and the early 1990s, as well as oil shocks in the 1970s. The average value-added of manufacturing over this period was approximately 50%, declining from 60% in 1973 to 40% in 1999. Aggregate volatility, measured by the standard deviation of the aggregate growth rate of total output, over this period was approximately 7% in manufacturing, slightly higher than the overall for Canada during the same period, around 5%.

To get a sense of how much of aggregate volatility is due to plant-specific variance versus plant-pair specific covariance, use the fact that the aggregate growth rate is the weighted sum of individual growth rates, then

\[
\sigma_A^2 = \sum_i \text{Var}(w_{it-1}g_{it}) + \sum_i \sum_{j \neq i} \text{Cov}(w_{it-1}g_{it}, w_{jt-1}g_{jt})
\]

The variance component only accounts for 4% of total aggregate volatility. Firm or plant level studies of volatility typically focus on variances, but since 96% of aggregate volatility is due to covariance terms, that is the primary focus of this paper. Keep in mind that in many economic models, especially the input-output network outlined later in this paper, idiosyncratic productivity shocks will result in positive comovement between measured plants’ output growth rates—these covariances and connections are the basis of the asymmetric network theory.

The long-form survey, intended to provide additional detail for the biggest plants, covers approximately 18,000 plants and 92% of manufacturing output. It provides detailed data
on commodity level inputs and outputs for each plant at the 9-digit Standard Classification of Goods (SCG) level. This commodity survey, which serves as part of the basis for Canada’s input-output tables, provides the essential plant-level input-output data that allows me to investigate the plant-specific IO network and its effect on aggregate volatility. Just like industry-commodity level make-and-use tables in the national accounts, each plant can consume multiple inputs and produce multiple outputs, and each commodity may be produced or consumed by multiple plants, even across industries. Not surprisingly, there is considerable heterogeneity in input-output statistics at the plant-level, both within and across industries, and this heterogeneity plays a big part in the mechanisms outlined later in the paper.

2.2. Sales and growth and volatility statistics. In the Canadian manufacturing sector, a few industries play outsized roles in output, employment and value-added. Transportation equipment production alone accounted for 21.5% of total manufacturing output in Canada in 1997, suggesting that a shock to that industry will have significant effects on the economy as a whole. However, the top ten plants in that industry account for the vast majority of its output. Although industries look granular, plants within industries appear to have the same asymmetry and may themselves contribute to aggregate volatility.

Across the economy, the industry-level herfindahl is .159, showing the intense concentration of output in a few industries. Within industries, output concentration at the plant level is even greater, with a mean herfindahl of .264 and a standard deviation of .144. Heterogeneity at both the plant and industry level is clear, and they contribute to an overall

---

1The plant level herfindahl is year $t$ is $h_t = \sqrt{\sum_{I \in I} \sum_{i \in I} w_{it}^2}$. The industry level herfindahl in year $t$ is $h_{It} = \sqrt{\sum_{I \in I} w_{It}^2}$. I report the averages $\bar{h} = (1/T) \sum_{t=1}^T h_t$ and $\bar{h}_{It} = (1/T) \sum_{t=1}^T h_{It}$. Within-industry herfindahls are defined similarly using the weights $w_{it}/It$. There is a possibility that survey design generates a mechanical relationship between herfindahl and weights, in which small industries (in terms of total output) have fewer plants in the survey, possibly because they are concentrated geographically and thus not many plants are required to estimate the total industry output for a province. In this case, there will be a negative relationship between the within-industry herfindahls and industry weights, because there are less firms to reduce the weight of sampled plants. To account for this, I also calculated the herfindahls using only the top 20 plants in each industry. The results are very similar, with $\bar{h} = .056$, $\bar{h}_{It} = .187$, and $(1/N_{It}) \sum_{I \in I} \bar{h}_I = .270$. 


plant-level herfindahl of .0566. The potential for granularity is strong at both the industry and within-industry levels.

Table 3 also displays summary statistics of growth rates by plant, year and industry. The mean growth rate across all plants and years is .074 with a standard deviation of .329. The average aggregate growth rate is .075 with a standard deviation of .066. The mean industry growth rate is .067 with a standard deviation of .107. At more disaggregated levels, individual volatility is increasing. However, it is decreasing much slower than $1/\sqrt{N}$, which is a sign that shocks are not averaging out at either the industry or plant level. However, observed plant growth rates may be measuring aggregate shocks themselves and not idiosyncratic shocks.

In an effort to separate industry-specific and plant-specific shocks, I define the residual plant growth rate $g_{i/t}$ to be the difference between the plant growth rate and its industry average growth rate $g_{i/It} = \frac{1}{N_I} \sum_{i' \in I} g_{i't}$,

\begin{equation}
(2) \quad g_{i/It} = g_{i/t} - g_{It}.
\end{equation}

Residual growth rates have similar properties as the observed growth rates, with means and standard deviations of -.069 and .383. Further decomposing growth rates into industry-province bins produces similar results. Summary statistics can be found in Appendix ??.

2.3. **Input-output statistics.** Given two randomly chosen plants in Canadian manufacturing, how are they connected through their production processes? I attack the question in two ways: first, the industry-by-industry or plant-by-plant direct-requirements table, the empirical version of a matrix of input shares. This is one of the main IO measures developed by statistical agencies to measure direct and indirect input-output connections between industries, and is the basis of the empirical analysis in Acemoglu et al. [2]. Second, I create a new measure of input-output connections that measures the correlation of inputs, outputs, and direct linkages between two plants. The two measures complement each other and each have different advantages in different facets of the analysis.
The main IO measure is the share of expenditure on inputs of plant $j$, $g_{ij}$. The goal of the empirical analysis is to construct the plant-level version of $g_{ij}$ using the same method as the IO tables in statistical agencies. The ultimate result is a matrix $G$ of plant-plant input shares, where a typical element is $g_{ij}$ and satisfies $\sum_{j=1}^{N} g_{ij} = 1 - \beta_i$, where $\beta_i$ is the value added share of output of plant $i$ (see Appendix 9 for the full derivation). If you arrange the plants in $G$ by industry, it can be decomposed into industry-by-industry blocks. In addition, the weighted outdegree, $d_i$ is a measure of the importance of plant $i$ to the entire economy, measured by the sum of input shares across all other plants:

$$d_i \equiv \sum_{j \in J} \sum_{j} g_{ji} \quad (3)$$

Next, the novel measure of commodity correlation between plants is based on commodity-level similarity in inputs, outputs and input-output between two plants. Suppose $I_i$ is the input vector of plant $i$, then the input correlation $\rho^I_{ij}$ is the uncentered correlation between the two vectors,

$$\rho^I_{ij} \equiv \frac{I_i \cdot I_j}{||I_i||_2 \times ||I_j||_2} \quad (4)$$

The direct linkage correlation $\rho^L_{ij}$ and output correlation $\rho^O_{ij}$ are defined in the same way.

What is $\rho^I_{ij}$? Suppose $i$ and $j$ input totally different commodities, so they do not share any of the same intermediate inputs. Then the dot product $I_i \cdot I_j$ is 0, meaning the two input vectors are orthogonal and have zero correlation. If the two plants input the exact same commodities in the same proportions, then the input vectors overlap and are perfectly correlated. In terms of commodity inputs, the two plants are pointed in the exact same direction.

Another way of seeing the relationship between two plants is by the angle between their commodity vectors, defined by $\rho^I_{ij} \equiv \cos \theta^I_{ij}$, which means two uncorrelated input vectors are perpendicular in commodity space ($\theta^I_{ij} = \pi/2$), and two perfectly correlated input vectors overlap ($\theta^I_{ij} = 0$).
The benefit to using these input-output measures, in addition to the direct-requirements tables, is that it may offer a way to uncover commodity-level common shocks that are dispersed across industries that industry-level fixed effects may not pick up. In some cases, two plants in different industries with common inputs may be correlated in ways that are not related to their industries (and would be mistakenly measured as plant-specific covariance), but are really common commodity shocks that should not be considered correlated idiosyncratic plant shocks. A direct-requirements measure will not pick this up, but the input correlation $\rho_{ij}$ will.

The summary statistics of the IO measures are displayed in Appendix 9.4. The sparseness of the plant-level IO matrix is clear: of the 324 million possible connections between plants, less than 1% share any kind of link (in the sense of having strictly positive values of $G$), compared to 11% of industry pairs having connections. More disaggregated microeconomic levels become more and more sparse, suggesting there is significant heterogeneity in IO within sectors, because if every plant within an industry pair had IO characteristics that matched the aggregate level, the plant level connections would match the industry level. Instead, the IO network becomes more sparse as it becomes more disaggregated, so plants are only connected to certain other plants in another industry, and not all of them.

As shown by Acemoglu et al. [2], the increasing sparseness of the network does not necessarily mean idiosyncratic shocks are not important, as long as the network retains the asymmetric properties that represent the importance of a single plant or industry to the whole economy. In this case, the within-industry heterogeneity in IO suggests that there are important plants within important industries that are the source of aggregate fluctuations in the economy. The remaining question is whether the asymmetry in the observed IO network is truly due to asymmetry in the exogenous underlying IO network.

To illustrate the importance of asymmetry, Figure 1a plots the rank of $d_i$ versus $d_i$ itself, on a log-log scale (for the plants with a strictly positive $d_i$). The asymmetry and heavy-tailed distribution of outdegree is apparent—there are a few plants are very important to the network, and are significant suppliers of a large number of other plants. Figures 1b and 1c...
show similar relationships for labour productivity and plant size, respectively, suggesting all three have power law tails and that productivity and outdegree both determine the plant size distribution.

**Figure 1.** Hexbin rank plots for plant characteristics, all with linear (power law) right tails. In panel (D), we can see that degree is positively correlated with plant size.
Viewing plants as sectors, this confirms Acemoglu et al.’s [2] conjecture that the network asymmetry is preserved as the economy becomes more and more disaggregated. In addition, this result suggests that there is also significant plant-level variation of characteristics we normally only associate with industries. This is mainly because of lack of data at the plant or firm level, and so this paper provides the first evidence of IO network asymmetry at the plant level.

To confirm the power law behaviour of the sequence of plant outdegrees, I estimate the share parameter of the tail of the distribution. Following Gabaix and Ibragimov [20], I trim the distribution to the top 20\textsuperscript{th} percentile of outdegree and estimate

\begin{equation}
\log(\text{rank}_i - 1/2) = \alpha - \beta \log d_i
\end{equation}

The estimated shape parameter $\hat{\beta}$ is a measure of the strength of the asymmetry in the distribution—a shape parameter of 1 is Zipf’s law.

The estimated parameter is $\hat{\beta} = 1.21$ (s.e. = 0.011), slightly lower (heavier tailed) than the sector level results from Acemoglu et al. [2], suggesting the plant level outdegree distribution is just as asymmetric as the industry level outdegree distribution in the US. Furthermore, plant size is positively correlated with outdegree in the hexbin plot in Figure 1(D). The elasticity of weight with respect to outdegree has an elasticity of .31 (s.e. .0049), suggesting a strong relationship between being an important plant to the economy (having a high outdegree) and being large (having a large influence vector).

There are two main takeaways from the IO data. First, there is incredibly asymmetry in plant-level IO network, suggesting certain plants are very important suppliers in the economy, especially within industries. Second, that asymmetry is positively associated with plant weight, confirming the relationship between the theoretical measure of asymmetry and the

\footnote{Hexbin plots have two advantages here. First, for confidentiality reasons I cannot display scatterplots because you can identify characteristics of each individual establishment. Second, scatterplots with thousands of points can be impossible to decipher. The hexbin plot is like a two-dimensional histogram, with the opacity of each hexagon representing the number of plants in that bin (representing the same thing as the length of a histogram bar).}
reduced form weight vector used for applying the granularity theory against diversification of idiosyncratic shocks.

However, keep in mind that these are observed IO statistics, not a true underlying IO network. Given the relationships between productivity and IO at the plant level, it is crucial to understand the data through the lens of an appropriate model.

3. Model

To study the relationships between volatility, endogenous unbalanced IO networks and the factors that determine them, I adapt the sectoral model of Acemoglu et al. [2], which is itself based on Long and Plosser [30]. There are two key additions. First, I study individual plants and not sectors. Although technically easy (e.g., relabeling sectors as plants), it puts the focus on the determinants of granularity—the IO network or something else? This becomes crucial as we turn to the study of a very disaggregated economy, which is the primary reason for studying microfoundations of aggregate volatility. Second, I relax the assumption that the IO network is exogenous. In my model, a plant may be a central supplier of the network because it is a required input in many other products (it has many high exogenous direct input coefficients) or because it is so productive that many other plants substitute toward it. To introduce these features, I need a model in which productivity and the exogenous IO network can vary independently to create an observed plant-level IO network that I can take to the data.

To start, a representative household inelastically supplies a single unit of labour, and has Cobb-Douglas preferences over $N$ different goods,

$(6)\quad u(c) = \prod_{i=1}^{N} c_i^{1/N}$

where $c_i$ is consumption of good $i$. Each good is produced by a single plant using Cobb-Douglas combination of labour and a plant-specific intermediate input which is itself a CES
aggregate of other products,

\[ q_i = z_i l_i^\beta \left( \sum_{j=1}^{N} \gamma_{ij}^{1/\sigma} q_{ij}^{\sigma-1} \right)^{(1-\beta)\sigma} \frac{(1-\beta)\sigma}{\sigma-1} \]

where \( z_i \) is productivity, \( \beta \) is the labour share in production, \( q_{ij} \) is the quantity of plant \( j \)'s product demanded by plant \( i \), and \( \sigma \) is the elasticity of substitution between intermediates.

The crucial part of production is \( \gamma_{ij} \), which is the exogenous direct input coefficient. If \( \gamma_{ij} \) is high, then independent of plant \( j \)'s productivity, plant \( i \) requires a lot of plant \( j \)'s input to produce. If \( \gamma_{ij} \) is low but positive, then plant \( i \) may still demand a lot of \( q_{ij} \) if plant \( j \) is very productive. In this way, the endogenous IO network is determined jointly by productivity, substitutability and the exogenous IO network.

With perfect competition, prices equal marginal costs for plant \( i \),

\[ p_i = C z_i^{-1} \left( \sum_{j=1}^{N} \gamma_{ij} p_j^{1-\sigma} \right)^{1-\beta-\sigma} \]

where \( C \equiv \beta^{-\beta}(1-\beta)\beta^{-1}w^{\beta} \) is independent of \( i \).

Remark 1. Observed expenditure shares depend on productivity and exogenous IO characteristics.

The IO tables provided by statistical agencies gives an expenditure share of industry \( i \) on goods from industry \( j \). The plant IO table I detail in the previous section is constructed the same way, an expenditure share of plant \( i \) on plant \( j \). If we assume production is Cobb-Douglas, then the expenditure share parameter in production exactly determines the observed expenditure share. This is no longer true if the elasticity of substitution is not equal to 1. Define the observed expenditure share \( g_{ij} \),

\[ g_{ij} = \frac{p_j q_{ij}}{p_i q_i} \]
In equilibrium, this simplifies to

$$g_{ij} = (1 - \beta) \left( \frac{\gamma_{ij} p_j^{1-\sigma}}{\sum_{k=1}^{N} \gamma_{ik} p_k^{1-\sigma}} \right)$$

If $\sigma = 1$, the observed expenditure share is exactly determined by the relative exogenous coefficient $\gamma_{ij}$ (that is, if you rederive the solution starting with $\sigma = 1$ in the production function). However, it is clear that the observed expenditure shares are jointly determined by the vector of direct input coefficients $\gamma_{ij}$ and the vector of prices, which are themselves determined by the vector of plant productivities (and more complex interconnections). Again, the observed IO network is endogenously determined by the vector of plant productivities and the exogenous IO network.

**Remark 2.** Expenditure shares still “determine” size, but they say nothing about the underlying determinants of the size distribution.

In an important result, Acemoglu et al. [2] shows that the vector of industry sizes, normalized by total sales in the economy, which he calls the influence vector $v$, is the crucial link between the IO network and volatility. The influence vector determines the extent to which microeconomic shocks contribute to aggregate volatility, and the influence vector is determined by the characteristics of the exogenous IO network. Hence their claim that the IO network is the main determinant of aggregate volatility. Here I show that the same holds for the observed IO network. That is, an empirical association between the influence vector and the true IO network does not tell you the effect of the IO network on volatility, because the observed network may be entirely determined by productivity. Write the system of market clearing equations,

$$p_i c_i + \sum_{j=1}^{N} p_i q_{ji} = p_i q_i, \text{ for } i = 1, \ldots, N$$
And rewrite in terms of $g_{ij}$ using (9),

\begin{equation}
\label{eq:12}
p_i c_i + \sum_{j=1}^{N} g_{ji} p_j q_j = p_i q_i, \text{ for } i = 1, \ldots, N
\end{equation}

Then a similar derivation to Acemoglu et al. [2] (see Appendix 8) gives you the influence vector as a function of the matrix of observed expenditure shares $G = [g_{ij}]$,

\begin{equation}
\label{eq:13}
v' = \frac{\beta}{N} 1'(I - G)^{-1}
\end{equation}

The influence vector, $v$, is always related to the observed IO network, but the observed IO network is endogenous. So observing the association between the influence vector and the IO network does not give you any information on the importance of the underlying IO network, $\Gamma = [\gamma_{ij}]$.

**Example 1.** Suppose $\gamma_{ij} = 1$ for all $i, j = 1, \ldots, N$. Then there is no exogenous IO network variation, and all of the observed IO characteristics are due to productivity.

If $\gamma_{ij} = 1$, then all plants use the same intermediate bundle and face the same intermediate input price. This means the expenditure share equation (9) reduces to

\begin{equation}
\label{eq:14}
g_{ij} = (1 - \beta) \left( \frac{z_j^{\sigma-1}}{N \sum_{k=1}^{N} z_k^{\sigma-1}} \right)
\end{equation}

Which is determined solely by relative productivities. In this case, if productivities are distributed with a power law, we will still observe an influence vector consistent with the unbalanced IO network, even though the underlying IO network is as balanced as possible.

**Example 2.** Suppose $z_i = 1$ for all $i = 1, \ldots, N$. Then there is no productivity variation, and all of the observed IO characteristics are due to the exogenous IO network.

When productivities are identical across all plants, the expenditure share terms reduce to

\begin{equation}
\label{eq:15}
g_{ij} = (1 - \beta) \left( \frac{\gamma_{ij}}{\sum_{k=1}^{N} \gamma_{ik} (p_k/p_j)^{1-\sigma}} \right)
\end{equation}
where \((p_k/p_j)^{1-\sigma}\) terms can be written as a recursive function of relative prices and IO parameters, which implies the expenditure shares are determined only by IO parameters.

3.1. **Outdegree and unbalanced IO networks.** An unbalanced IO network is one in which individual plants are central suppliers to the entire economy. The easiest way to ask how central a plant is by adding up the expenditure shares of a plant’s customers,

\[
\hat{d}_i = \sum_{j=1}^{N} g_{ji}
\]

**Example 3.** Suppose \(\gamma_{ij} = d_j/N\).

Expenditure shares are

\[
g_{ij} = (1 - \beta) \left( \frac{d_j z_j^{-\sigma-1}}{\sum_{k=1}^{N} d_k z_k^{-\sigma-1}} \right)
\]

Observed outdegree is

\[
\hat{d}_i = (1 - \beta) \left( \frac{d_i z_i^{-\sigma-1}}{(1/N) \sum_{k=1}^{N} d_k z_k^{-\sigma-1}} \right)
\]

And one element of the influence vector is

\[
v_i = \frac{\beta}{N} + (1 - \beta) \left( \frac{d_i z_i^{-\sigma-1}}{\sum_{k=1}^{N} d_k z_k^{-\sigma-1}} \right)
\]

This examples highlights the dependence of the influence vector on productivity and the unbalanced IO network—the distribution of \(v_i\) is determined by the distribution of \(d_i z_i^{-\sigma-1}\). Recall that the argument for microfoundations of aggregate shocks requires the distribution of \(v_i\) to have a thick tail even as the number of plants grows large. However, as the number of plants grows large, the thick tail of \(v_i\) will tend to be dominated by the thickest tail of the two distributions of outdegree and productivity.

The next section pins down the theoretical basis for these concepts, and the following sections explore the empirical support for them.
4. Asymptotic Theory

Asymptotic results are key to the arguments for and against the microfoundations of aggregate shocks. The granular hypothesis relies on a thick tail of the size distribution. The unbalanced network hypothesis claims the reason why the size distribution has a thick tail is because of a thick tail of outdegree, a telling characteristic of an unbalanced IO network. Only by combining the two approaches can we understand the forces that shape the observed centrality and size distributions.

In what follows, I rely especially on the following property of power law distributions:

Remark 3. Suppose the random variables $X$ and $Y$ follow power law distributions with parameters $\zeta_X$ and $\zeta_Y$. Then the distribution of $X + Y$ and the distribution of $XY$ both follow power laws with parameter $\min\{\zeta_X, \zeta_Y\}$.

The same result follows for many similar combinations of power law random variables (see [18] or [28]). Using Remark 3, we are interested in explaining the tail parameter of the size distribution, $\beta_v$, given the tail parameters of the distributions of observed outdegree ($\zeta_d$) and productivity ($\zeta_z$).

Therefore, if the asymptotic results hold for this economy, network asymmetry cannot be the fundamental cause of the skewed plant size distribution because of the relative values of each tail parameter. But like so many other applications of power laws, the reality is not so black and white. In any case, we must understand the asymptotic argument first, and then ask if and when is it reasonable to apply it.

The network hypothesis relies on two sequential arguments. First, the tail of the distribution of the plant-level exogenous IO network characteristics must determine the tail of the distribution of the observed plant-level IO network characteristics. Second, the tail of the distribution of the observed IO network determines the tail of the plant size distribution. If

---

3In Appendix 8, I use Hulten’s Theorem to show aggregate volatility depends on the herfindahl of the economy, and the herfindahl of the economy depends on the distribution of outdegree and productivity. These results are standard when applying the granular and network theories of aggregate fluctuations, so I omit them and focus on the new idea provided in this paper.
either of these arguments fail, it is unlikely the underlying IO network is the cause of the skewed plant size distribution.

I approach the second part of the argument first. For the observed network to matter asymptotically, the outdegree distribution must have a thick tail. If not, outdegree cannot be the ultimate source of the thick tail of the size distribution. If the outdegree distribution does have a thick tail, the parameter must match, or be “close” to matching (in a statistical sense) the tail of the size distribution. However, the measured tail parameter for the network is 1.21, about 20% higher than the plant size distribution’s parameter of 1.04, which is consistent with a Zipf’s law distribution of plant size. Therefore \( \zeta_z < \zeta_d \) implies the degree distribution is dominated by some other plant characteristic, and thus does not determine plant size asymptotically or turn idiosyncratic shocks into aggregate fluctuations.

We can see this conclusion supported by prior research in different settings. A plethora of research on the firm size distribution conclude it is approximately described by Zipf’s law in the upper tail (see [31] or [18], while Acemoglu et al. [2] measure the tail of the sector outdegree distribution at 1.38, much larger than the typical Zipf’s law size distribution parameter of 1.

The first part of the argument, the required relationship between the observed and unobserved network characteristics is more problematic. The IO data are necessarily the observed shares, and so depend on both the underlying IO network and other plant characteristics, especially productivity. However, in absence of direct evidence for or against the underlying network, I suggest that, asymptotically, productivity is more likely to be the cause of the observed IO network, and possibly the final size distribution.

To establish this formally, I show that, under the assumptions of the model in the previous section, the tail of the size distribution is dominated by the thickest tail between productivity (adjusted for substitutability) and outdegree.
Proposition 4.1. Suppose the distributions of outdegree and productivity both follow power laws with parameters $\zeta_d$ and $\zeta_z$,

\[ P(d > x) = C_d x^{-\zeta_d} L_d(x), \]
\[ P(z > x) = C_z x^{-\zeta_z} L_z(x) \]

Here, $L_d(x)$ and $L_z(x)$ are slowly varying functions, $C_d$ and $C_z$ are constants, and $\zeta_d$ and $\zeta_z$ are positive. Then the size distribution also follows a power law with parameter

\[ \min\{\zeta_d, \zeta_z/(\sigma - 1)\}, \]

\[ P(v > x) = C_v x^{-\min\{\zeta_d, \zeta_z/(\sigma - 1)\}} L_v(x) \]

Proof. See Appendix 8. □

The distribution of labour productivity has a tail parameter of approximately 1.97, so for a suitable choice of $\sigma$, it is easy to match the empirical tail parameter of the plant size distribution. In particular, if $\sigma \approx 2.89$, the size distribution will approximately satisfy Zipf’s law. It also could satisfy both, if substitutability for final goods is higher than for intermediates. Note that similar studies on productivity and size, especially ones focusing on international trade models, (e.g., see Appendix 10 for an extension of the model with monopolistic competition and plant entry and exit) gives the same result—firm size is determined by a combination of productivity and substitutability, with the size tail parameter being very close to 1 (see, e.g., a series of papers by di Giovanni and Levchenko and their co-authors [13, 10, 11]). The difference here is that they observe the size distribution and assume it must be because of productivity. For more on power laws and the determination of firm size, see [31] or [18].

Although the asymptotic theory gives clear cut answers as to which factor is responsible for the shape of the size distribution, the empirical results suggest the truth is somewhere between the two extremes.
5. Calibration

In this section, I calibrate the model to match features of the data to further explore the relationships between the unbalanced IO network and volatility. Instead of applying asymptotic results directly to infer which tail, productivity or outdegree, dominates the size distribution, using the model described in Section 3 I use data on plant productivity $z_i$, the observed input share matrix $G$ to solve for the unobserved technical requirement matrix $\Gamma$.

Although final demand didn’t add to the model and asymptotic theory, it is important empirically. Therefore, to match the data better, I change the consumer’s utility function to a CES combination of each product,

\[
    u(c) = \left( \sum_{i=1}^{N} \alpha_i c_i^{\sigma - 1} \right)^{\frac{\sigma - 1}{\sigma}}
\]

(23)

Where $c_i$ is consumption of plant $i$’s output. Now the unobserved final demand characteristic $\alpha_i$ is similar to a $\gamma_{ji}$ in firm $j$’s production function, and the observed final demand share $a_i$ is similar to the observed expenditure share $g_{ji}$.

5.1. Parameters. First, I need a measure of productivity, $z_i$, for each establishment. There are several methods to estimate productivity, although they give similar results. First, I use the method of Gandhi, Navarro and Rivers [21] to estimate total-factor productivity (TFP). Second, labour productivity, either calculated by value-added per worker or value-added per dollar spent on wages (which accounts for labour and capital quality better than raw value-added per worker or hours worked). Lastly, in an effort to distinguish revenue productivity from physical productivity and accounting for the number of products produced by each plant, I focus on a few industries for which I have commodity quantity and value data, I solve the model separately for those industries. The results for all are similar. Since revenue TFP varies less than labour productivity, and revenue TFP varies considerably more than physical TFP (consistent with Foster and Haltiwanger [17] and Huttman et al. [25]), revenue and physical productivity can explain even less of the observed IO network than labour productivity.
Next, the matrix of input-shares has been described already, $G$. In addition, the share of value-added in production, $\beta_i$, is calculated as exactly that. Next, the final demand parameter $a_i$ (similar to the final demand category in the industry input-output tables), is calculated as that left over after all within-manufacturing production has been taken into account. Finally, I set the elasticity of substitution $\sigma = 2$. Later, I test the robustness of the model against changing the elasticity of substitution, since that directly affects the tail of the plant-size distribution as we saw in Section 4.

The model is simple enough to be solved directly, using the $N \times N$ observations in $G$, $N$ final demand observations for $a_i$, and $N$ productivity observations to solve for the $N \times N$ unknowns in $\Gamma$ and $N$ unknowns in $\alpha_i$ (along with the $N$ prices and the final demand price $P$ normalized to 1).

5.2. Results. The main result is that productivity cannot explain much of the observed IO network; there is not enough heterogeneity in productivity to explain either the size distribution, the IO network or the final demand parameters. This is consistent with other recent work challenging the dependence of the plant size distribution on productivity.

(To be completed pending confidentiality review.)

5.3. Robustness. (To be completed pending confidentiality review.)

6. Regressions

Circling back to the differentiation between industries and plants, the ultimate goal is to understand the relationships between plant-specific input-output linkages, plant covariances and aggregate volatility. I boil this down to several underlying questions, all relying on within-industry variation—in other words, the main question is: does within-industry variation in input-output linkages affect aggregate volatility, and how? First, does plant outdegree affect aggregate volatility? Second, since aggregate volatility is a weighted sum of covariances, does outdegree affect the weights or the covariances or both at the same time, and what is the relative importance of each? How much does productivity matter for
determining weights? Then I dig deeper into the plant covariance measures to try to uncover the determinants of plant-level covariance. Do direct linkages matter more than producing common inputs or common outputs?

To simplify the moving parts, I adopt the method of Di Giovanni, Levchenko and Méjean [12] to focus on a single year $\tau$, and view $\sigma^2_{A\tau}$ as an estimate of aggregate volatility in year $\tau$.

\[
\sigma^2_{A\tau} = \sum_i \sum_j w_{i\tau-1} w_{j\tau-1} \text{Cov}(g_{i/It}, g_{j/Jt})
\]

In this way, I can use the commodity data from a single year, $\tau = 1992$, and estimate the effect of the IO network on the volatility estimates for 1992. From here on, I suppress the $\tau$ term, with the understanding that the following measures $V_i$, $d_i$ and $w_i$ are measured in year $\tau = 1992$. The analysis is robust to the choice of year.

6.1. Does within-industry variation in the IO network affect aggregate volatility?

Plant $i$’s contribution to aggregate volatility is

\[
V_i = w_i \sum_{j \neq i} w_j \text{Cov}(g_{i/It}, g_{j/Jt})
\]

There are three elements: (i) holding the sum constant, a bigger plant matters more for the economy, and for some plant $j$, $V_i$ is higher if (ii) that plant is bigger and (iii) it covaries more with plant $i$. To test whether outdegree affects aggregate volatility, regress $V_i$ on $d_i$, including industry fixed effects:

\[
V_i = \beta + \beta_d d_i + \mu_I + \epsilon_i
\]

If $\beta_d > 0$, then plants with higher outdegrees and thus higher importance to the IO network, have larger contributions to aggregate volatility. This leads to the next question.

6.2. How does within-industry variation in outdegree and productivity affect plant size? We expect an industry $I$ with a high outdegree $d_I$ to be larger than otherwise.
For the same reason, a plant with a higher outdegree relative to the rest of the industry should be larger than the rest of the industry. Is outdegree correlated with size? How does productivity factor in?

\[
\log w_i = \beta + \beta_d \log d_i + \beta_z \log z_i + \mu_I + \epsilon 
\]  

(27)

If \( \beta_d > 0 \) here, then plants with higher importance in the network (relative to other plants in the industry) are larger, which would confirm that they are more influential in the economy and have higher contributions to aggregate volatility. Similarly, we expect more productive plants to be larger, \( \beta_z > 0 \).

6.3. Does within-industry variation in outdegree affect a plant’s unweighted contribution to aggregate volatility? Rewrite plant \( i \)’s unweighted contribution to aggregate volatility as \( V_i/w_i \). Is this, the weighted sum of plant \( i \)’s covariances with every other plant, correlated with outdegree?

\[
V_i/w_i = \beta + \beta_d d_i + \mu_I + \epsilon 
\]  

(28)

If \( \beta_d > 0 \), then plants with more influence have higher contributions to aggregate volatility, independent of whether outdegree has an effect on the plant’s weight.

6.4. How much does productivity matter for aggregate volatility? Having established theoretically that productivity is an important determinant of plant weights, and plant weights are an important determinant of aggregate volatility, I ask how the empirical results change when I add productivity into the mix. For instance, the asymptotic results suggest that productivity may be the only determinant of plant weights.

To that end, I regress plant \( i \)’s contribution to aggregate volatility on productivity and outdegree (limiting the sample to those with strictly positive measures of both, but similar
results obtain for the full sample with non-log measure of outdegree).

$$V_i = \beta + \beta_z \log z_i + \beta_d d_i + \mu_I + \epsilon_i$$

![Formula](image)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(5.1)</th>
<th>(5.2)</th>
<th>(5.3)</th>
<th>(5.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>outdegree, $d_i$</td>
<td>1.62</td>
<td>1.27</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>log $d_i$</td>
<td>(.256)*</td>
<td>(.631)*</td>
<td>(.258)*</td>
<td></td>
</tr>
<tr>
<td>log $z_i$</td>
<td>.796</td>
<td>4.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>3.38</td>
<td>−18.8</td>
<td>49.9</td>
<td>−48.7</td>
</tr>
<tr>
<td>Ind. FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Scale</td>
<td>$(\times 10^{-8})$</td>
<td>$(\times 1)$</td>
<td>$(\times 10^{-5})$</td>
<td>$(\times 10^{-8})$</td>
</tr>
<tr>
<td>No. obs</td>
<td>17280</td>
<td>10887</td>
<td>17280</td>
<td>16994</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.107</td>
<td>.613</td>
<td>.015</td>
<td>.160</td>
</tr>
</tbody>
</table>

**Table 1.** Regression results (standard errors are in parentheses).

The results, shown in Table 1, confirm that the network and granular theories of aggregate volatility operate at the plant-level. Controlling for industry effects, outdegree is positively associated with the plant’s overall contribution to aggregate volatility $V_i$ (column (5.1)).

Furthermore, outdegree is positively associated with within-industry weights, suggesting plants with higher network importance have a higher contribution to aggregate volatility partly through size relative to their industry (column (5.2)). Outdegree is also positively associated with the unweighted contribution ($V_i/w_i$, column (5.3)), increasing its effect on aggregate volatility. Although outdegree would still have an effect on aggregate volatility if this coefficient were insignificant, as long as the mean unweighted covariance is positive; however, it is significant and thus suggests outdegree exerts influence on this term as well. Overall, outdegree is positively associated with both parts of a plant’s contribution to aggregate volatility, its own weight $w_i$ and the sum of the weighted covariances with other firms $V_i/w_i$. 
Column (5.4) reports that productivity and outdegree both have positive and significant effects on aggregate volatility, and the coefficient on outdegree is largely unchanged from the regression in (5.1) that does not control for productivity.

To calculate the total effect of within-industry network asymmetry on aggregate volatility, I use a regression similar to 29 and column (5.4) and a decomposition similar to the volatility decomposition in Di Giovanni, Levchenko and Méjean [12]. I decompose the predicted value of aggregate volatility $\hat{\sigma}_A^2$ into the part explained by degree $d$ and that leftover. First, I eliminate the network asymmetry by setting $d_i = 1/N$ and calculate the predicted volatility, $\bar{\sigma}^2 = \sum_i \bar{V}_i$. Then the network asymmetry accounts for the rest of the predicted volatility, $\sigma_d^2 = \hat{\sigma}_A^2 - \bar{\sigma}^2$. This means the relative standard deviation is 0.34, which means network asymmetry accounts for 34% of aggregate volatility. Put another way, eliminating network asymmetry reduces the standard deviation of aggregate growth by 34%.

In all, the empirical results support two conclusions. First, within-industry variation in input-output networks exists and contributes significantly to aggregate volatility. This supports and extends results from Acemoglu et al. [2] on industry-level IO networks at a much more disaggregated level, and challenges the common assumption that all plants within an industry have the same IO characteristics. Second, it offers insight into the factors that determine the key cog in the argument for microfoundations of aggregate fluctuations—the thick tail of plant size. Productivity and network asymmetry both contribute to plant granularity, despite the common asymptotic argument that suggest only one will dominate, which implies there is room for Acemoglu et al.’s network asymmetry theory to co-exist with productivity under the umbrella of Gabaix’s granular theory of aggregate fluctuations. Network asymmetry contributes to granularity, but is not the only factor.

6.5. Do input-output linkages affect plant-plant covariances? Although plant size is a large determinant of aggregate volatility, if the covariances did not vary across plants, then each plant’s unweighted contribution would be exactly the same. Moreover, if covariances

$\hat{\sigma}_A^2 = \sum_i \tilde{V}_i$, where $\tilde{V}_i$ is the predicted value for plant $i$. [27]
were negative they would actually reduce aggregate volatility, partially offsetting the strictly positive plant variances. Or, if the covariance terms are zero, or random, then covariances would not have any relationship with plant weights and would not have any effect on aggregate volatility. On top of that, if common trends in growth rates can be eliminated, then covariances are direct measures of the transmission of idiosyncratic shocks, making the covariances an important subject to study in their own right.

To test the effect of input-output connections on plant covariances, regress plant specific covariance terms on input, output and linkage correlations:

\[
\text{Cov}(g_{it}, g_{jt}) = X_{ij}\beta + \beta_I\rho_{Iij} + \beta_L\rho_{Lij} + \beta_O\rho_{Oij} + \mu_{IJ} + \epsilon_{ij}
\]

(30)

Recall \(g_{it}\) is the observed sales growth rate of plant \(i\). The \(X_{ij}\) are a vector of plant-pair specific characteristics which could include common characteristics, such as a dummy for whether both plants are owned by the same firm, whether they are both in the same province, or in the same industry. In another specification, I replace \(g_{it}\) with the residual plant growth rate \(g_{i/I_t}\) in an attempt to reduce common sectoral shocks between plants \(i\) and \(j\). In addition, I add weights to the covariances to test how the IO measures matter for the contributions to aggregate volatility.

\textit{A priori}, the effect of the commodity correlation between plants on their covariance could be positive or negative. Consider the input correlation \(\rho_{Iij}\): on one hand, two plants may negatively covary if they are competing for the same inputs. On the other hand, if they both heavily depend on a single commodity input, a supply shock to that commodity will affect sales of both of those plants in the same way, inducing positive covariance. The output correlation term \(\rho_{Oij}\) may work in a similar way.

The linkage correlation term \(\rho_{Lij}\) is a measure of the strength of the IO relationship between two plants, being higher when one plant inputs a commodity the other plant produces. In
this case, we expect the effect to be positive, and plants with overlapping inputs and outputs will covary more.

Table 2 presents the results. All commodity correlation measures are positively correlated with covariances between plants. The effect of linkage and input correlations are approximately the same, with correlation between output commodities having a slightly lower effect. However, two plants are much more likely to have similar inputs than they are to be linked in the supply chain, so the overall effect is higher for inputs.

The same patterns result in residual growth rate covariances, suggesting that accounting for both industry growth effects and industry-pair covariances, two plants still significantly covary more if they share input and output commodities.

Similar properties emerge when testing the weighted plant covariances, although the coefficients on linkages have approximately twice the magnitude as input correlations, and output correlations are no longer significant. This suggests that pairs of plants that are in a supply chain are bigger as a pair.

All results control for industry-pair fixed effects. Two more observations: notice the mean of each covariance term is positive, and the $R^2$ is very low even after controlling
for industry pair effects, less than 1% in the first two specifications and less than 10% in the weighted specifications. This suggests the correlation measures tell us very little about plant covariances. One possible culprit is measurement error in covariances—26 periods is a relatively short period to measure covariances, and the lives of many plants do not overlap for even that length of time. This will lead to much unexplained variation in measured covariances.

7. Conclusion

I investigate the relationship between idiosyncratic shocks, unbalanced input-output (IO) networks and aggregate volatility. Using detailed data on commodity inputs and outputs in Canadian manufacturing, I study a plant-level IO network and its effect on aggregate volatility. My main contribution is to account for the endogenous observed IO network and quantify the separate effects of productivity and the underlying IO network on plant size and aggregate volatility.

To differentiate between the granular and network hypotheses of aggregate fluctuations, I use a model in which productivity and the underlying IO network vary independently and use the plant-plant IO network data to uncover the model parameters. I find that productivity cannot explain the asymmetry in the observed IO network and that the majority of the variation in plant size, and therefore aggregate volatility, is caused by the underlying IO network.

I compare the properties of the IO network to each plant’s contribution to aggregate volatility, and confirm that more central plants matter more for aggregate volatility. Specifically, a 10% increase in a plant’s outdegree is associated with a 2.66% increase in size, while a 10% increase in labour productivity is associated with an 8% larger plant. The asymmetry of the IO network contributes 34% to aggregate volatility in Canadian manufacturing.

In conclusion, to investigate the propagation of idiosyncratic shocks, I acknowledge and investigate the endogeneity of the observed IO network and find the underlying IO network does account for a sizable proportion of aggregate volatility. Future research can extend this
work in several ways: using the plant-plant IO network to directly investigate the propagation mechanism of idiosyncratic shocks, adding in financial linkages between establishments within firms, or identifying supply chains across the economy, instead of just manufacturing. Doing so will increase our knowledge of the complex linkages that underpin our economy.

REFERENCES


8. Appendix: Theory

8.1. Derivation of influence vector. Using the definition of observed expenditure shares,\n\n\[(31)\] \[g_{ji} = \frac{p_i q_{ji}}{p_j q_j}\]

Rewrite the system of market clearing equations\n\n\[(32)\] \[p_i c_i + \sum_{j=1}^{N} p_i q_{ji} = p_i q_i, \text{ for } i = 1, \ldots, N\]

as\n\n\[(33)\] \[p_i c_i + \sum_{j=1}^{N} g_{ji} p_j q_j = p_i q_i, \text{ for } i = 1, \ldots, N\]

Then replace \(p_i c_i = wL/N\) and define total sales as \(s_i = p_i q_i\),\n\n\[(34)\] \[\frac{wL}{N} + \sum_{j=1}^{N} g_{ji} s_j = s_i, \text{ for } i = 1, \ldots, N\]

Rewrite in vector form, using \(g_i\) as the \(i\)-th column of \(G\),\n\n\[(35)\] \[\frac{wL}{N} + g_i^t s = s_i, \text{ for } i = 1, \ldots, N\]
Now stack those $N$ equations on top of each other, which stacks the vectors $g'_i$ (now the row vectors of $G'$), which gives

\[ \frac{wL}{N} \mathbf{1} + G's = s \]  

Rearrange and factor out $s$,

\[ s - G's = \frac{wL}{N} \mathbf{1} \]  
\[ (I - G')s = \frac{wL}{N} \mathbf{1} \]  

Then pre-multiply by the Leontief matrix, the inverse of $(I - G')$,

\[ s = \frac{wL}{N} (I - G')^{-1} \mathbf{1} \]

To get the form in the text, use $wL = \beta \sum_{i=1}^{N} s_i$ and $v_i = s_i / \left( \sum_{j=1}^{N} s_j \right)$, and finally take the transpose of both sides:

\[ v' = \frac{\beta}{N} (\mathbf{1}')(I - G)^{-1} \]

8.2. Aggregate volatility depends on the product of the distributions of outdegree and productivity. Aggregate volatility scales according to $||v||_2$ (see Hulten’s Theorem \[26\] and Theorem 1 of Acemoglu et al. \[2\]). To add to those results, I characterize the behaviour of $||v||_2$ in terms of the distributions of outdegree and productivity.

Write an element of the influence vector $v_i$ as

\[ v_i = \frac{\beta}{N} + \left(1 - \beta\right) \left( \frac{d_i z_i^{\sigma-1}}{\sum_{k=1}^{N} d_k z_k^{\sigma-1}} \right) \]
Then the Euclidean norm of $v$ can be written

$$
(42) \quad ||v||_2 = \sqrt{\sum_{i=1}^{N} \left( \frac{\beta^2}{N^2} + (1 - \beta)^2 \left( \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right)^2 + 2(1 - \beta) \left( \frac{d_i z_i^\sigma - 1}{N} \right) \right)}
$$

$$
(43) \quad ||v||_2 = \sqrt{\frac{\beta^2}{N} + (1 - \beta)^2 \sum_{i=1}^{N} \left( \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right)^2 + 2(1 - \beta) \left( \frac{d_i z_i^\sigma - 1}{N} \right) \sum_{i=1}^{N} \left( \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right)}
$$

Rewrite slightly,

$$
(44) \quad ||v||_2^2 = \frac{\beta^2}{N} + (1 - \beta)^2 \sum_{i=1}^{N} \left( \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right)^2 + 2(1 - \beta) \left( \frac{d_i z_i^\sigma - 1}{N} \right)
$$

$$
(45) \quad ||v||_2^2 = \frac{\beta^2}{N} + 2(1 - \beta) \left( \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right) + (1 - \beta)^2 h_g^2
$$

$$
(46) \quad ||v||_2^2 = \frac{\beta(2 - \beta)}{N} + (1 - \beta)^2 h_g^2
$$

$$
(47) \quad ||v||_2^2 \geq (1 - \beta)^2 h_g^2
$$

Implying $||v||_2^2 = \Omega \left( h_g^2 \right)$.

In addition, $||v||_2^2 = O \left( h_g^2 \right)$. To see this, first note

$$
(48) \quad h_g^2 \geq \frac{1}{N} \left( \sum_{i=1}^{N} \frac{d_i z_i^\sigma - 1}{N \sum_{k=1}^{N} d_k z_k^\sigma - 1} \right)^2 = \frac{1}{N}
$$

which we can rearrange to get $1/(Nh_g^2) \leq 1$.

$$
(49) \quad ||v||_2^2/h_g^2 = \frac{\beta(2 - \beta)}{Nh_g^2} + (1 - \beta)^2
$$

Meaning

$$
\limsup_{N \to \infty} \frac{||v||_2^2}{h_g^2} = \limsup_{N \to \infty} \left[ \frac{\beta(2 - \beta)}{Nh_g^2} + (1 - \beta)^2 \right]
$$
Using the result that \((Nh_g^2)^{-1}\) is bounded above by 1,

\[
\limsup_{N \to \infty} \frac{||v||^2}{h_g^2} \leq \limsup_{N \to \infty} \left[ \beta(2 - \beta) + (1 - \beta)^2 \right]
\]

(51)

\[
\limsup_{N \to \infty} \frac{||v||^2}{h_g^2} \leq \beta(2 - \beta) + (1 - \beta)^2 < \infty
\]

(52)

So \(||v||^2 = O(h_g^2)\), which combined with the Big-\(\Omega\) result gives

(53)

\[ ||v||_2 = \Theta(h_g) \]

8.3. **Proof of Proposition 4.1**

*Proof.* One element of the influence vector, \(v_i\), is

\[
v_i = \frac{\beta}{N} + (1 - \beta) \left( \frac{d_i z_i^{\sigma-1}}{\sum_{k=1}^{N} d_k z_k^{\sigma-1}} \right)
\]

(54)

As \(N \to \infty\), the first term approaches zero, and the distribution of \(v\) is determined by the relative product term \(d_i z_i^{\sigma-1}\), which means

\[
v_i \to \chi d_i z_i^{\sigma-1}
\]

(55)

\[
F_v(x) = F_v\left(\chi d_i z_i^{\sigma-1}\right)
\]

(56)

\[ P(v > x) \to P(\chi dz^{\sigma-1} > x) \]

(57)

\[ = P(dz^{\sigma-1} > \chi^{-1}x) \]

(58)

\[
P(v > x) = P(dz^{\sigma-1} > \chi^{-1}x)
\]

(59)

\[
= \int_{d}^{\infty} P\left( z > \left[ \frac{x}{\chi d} \right]^{1/(\sigma-1)} \right) dF_d(d)
\]

(60)

\[ = \int_{d}^{\infty} C_z \left[ \frac{x}{\chi d} \right]^{-\xi_z/(\sigma-1)} dF_d(d) \]

(61)

\[ = \chi^{\xi_z/(\sigma-1)} C_z x^{-\xi_z/(\sigma-1)} \int_{d}^{\infty} d^{\xi_z/(\sigma-1)} dF_d(d) \]

(62)
For the integral to exist, we need \( \frac{\zeta_z}{\sigma-1} < \zeta_d \). If so, it is a constant (independent of \( x \)), so combine the other constants into \( C_v = \chi^{\zeta_z/\sigma-1} \int_d^\infty d^{\zeta_z/\sigma-1} dF(d) \), and write

\[
P(v > x) = C_v x^{-\frac{\zeta_z}{\sigma-1}}
\]

So \( v \) has a power law distribution with parameter \( \frac{\zeta_z}{\sigma-1} \). If \( \frac{\zeta_z}{\sigma-1} > \zeta_d \), we need to derive it the other way, and end up with a power law distribution with parameter \( \zeta_d \). Therefore the distribution can be expressed by

\[
P(v > x) = C_v x^{-\min\{\zeta_d, \zeta_z/\sigma-1\}}
\]

Or,

\[
\log P(v > x) = \log C_v - \min\{\zeta_d, \zeta_z/\sigma-1\} \log x
\]

\[
\square
\]

9. Appendix: Data and Empirics

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate</td>
<td>26</td>
<td>.075</td>
<td>.066</td>
</tr>
<tr>
<td>Herfindahl</td>
<td>26</td>
<td>.0566</td>
<td>.0071</td>
</tr>
<tr>
<td>Industries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate</td>
<td>232 \times 26</td>
<td>.067</td>
<td>.107</td>
</tr>
<tr>
<td>Herfindahl</td>
<td>26</td>
<td>.159</td>
<td>.0129</td>
</tr>
<tr>
<td>(Within-I) herfindahl</td>
<td>232 \times 26</td>
<td>.264</td>
<td>.144</td>
</tr>
<tr>
<td>Plants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth rate</td>
<td>306146</td>
<td>.0739</td>
<td>.329</td>
</tr>
<tr>
<td>Residual growth rate</td>
<td>306146</td>
<td>-.0697</td>
<td>.383</td>
</tr>
<tr>
<td>Labour productivity</td>
<td>32710</td>
<td>51771</td>
<td>87407</td>
</tr>
</tbody>
</table>


9.2. Direct requirements table. To construct the plant-level direct requirements table, a matrix \( G = [G(i,j)] \), start with some notation: \( D = V \hat{q}^{-1} \), where \( V \) is the \( N \times C \) matrix
of outputs, where rows are plants and columns are commodities. So a typical entry is \( O_{ic} \), where plant \( i \) outputs value \( O \) of commodity \( c \). \( V \) is the make table. \( \tilde{q}^{-1} \) is a diagonal matrix, with \( 1/O_c \) on the diagonal, where \( O_c \) is the total output of commodity \( c \) in the economy. It is \( C \times C \). This makes a typical entry in \( D(i, c) \)

\[
D(i, c) = \frac{O_{pc}}{\sum_{i'} O_{i'c}}
\]

(66)

Now \( B = U \tilde{q}^{-1} \), where \( B \) is the \( C \times N \) matrix of input values (the Use matrix), and \( \tilde{q}^{-1} \) is a diagonal \( N \times N \) matrix with typical element \( 1/O_i \), where \( O_i \) is the total output of plant \( i \).

A typical element \( B(c, j) \) is

\[
B(c, j) = \frac{I_{jc}}{O_j}
\]

(67)

The direct requirements matrix I’m interested in is \( G = (DB)' \), a plant-plant \( N \times N \) matrix of shares. Instead of writing it out, it’s easier to calculate a specific element of the matrix \( G(i, j) \), which is the empirical version of \( \gamma_{ij} \), the share of \( j \)’s output in \( i \)’s production (aka \( \gamma_{ij} \) in \( i \)’s Cobb-Douglas production function).

So \( G(i, j) = B(:, i)' \cdot D(j, :)' \);

\[
G(i, j) = \sum_c \left[ \left( \frac{O_{jc}}{\sum_{j'} O_{j'c}} \right) \left( \frac{I_{ic}}{\sum_{i'} O_{i'c}} \right) \right]
\]

(68)

Some rearrangements may be useful later:

\[
G(i, j) = \frac{1}{\sum_c O_{ic}} \left( \sum_c \left( \frac{O_{jc} I_{ic}}{\sum_{j'} O_{j'c}} \right) \right)
\]

(69)

\[
G(i, j) = \frac{1}{S_i} \left( \sum_c \left( \frac{O_{jc} I_{ic}}{\sum_{j'} O_{j'c}} \right) \right)
\]

(70)
It’s easier to write Now, if I were using industry level instead of plant level,

\[ O_{Jc} = \sum_{j \in J} O_{jc} \]  
\[ I_{ic} = \sum_{i \in I} I_{ic} \]  
\[ S_I = \sum_{i \in I} S_i = \sum_{i \in I} \sum_{c'} O_{ic'} \]  

The goal is to write the number \( G(I, J) \) as a function of the block matrix \( G(i, j) \) when \( i \in I \) and \( j \in J \):

\[ G(I, J) = \frac{1}{S_I} \left( \sum_c \left( \frac{O_{Jc}I_{ic}}{\sum_{j' \in J} O_{j'c}} \right) \right) \]  
\[ G(I, J) = \frac{1}{\sum_{i \in I} S_i} \left( \sum_c \left( \frac{O_{Jc}I_{ic}}{\sum_{j' \in J} O_{j'c}} \right) \right) \]  
\[ G(I, J) = \frac{1}{\sum_{i \in I} S_i} \left( \sum_c \left[ \frac{\sum_{j \in J} O_{jc}}{\sum_{j' \in J} O_{j'c}} \right] \right) \]  

Simplify by trying to work with the terms inside the inner square brackets first,

\[ X_{I,Jc} = \left( \sum_{j \in J} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \]  
\[ X_{I,Jc} = O_{j_1 c} \left( \sum_{i \in I} I_{ic} \right) + \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \]  
\[ X_{I,Jc} = O_{j_1 c} I_{i_1 c} + O_{j_1 c} \left( \sum_{i \neq i_1} I_{ic} \right) + \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \]  

Dividing by the total sum term \( O_c \) (which is the same for commodity \( c \) regardless of whether the decomposition is at the plant or industry level), and summing over all \( c \in C \), the first term is \( G(i_1, j_1)S_{i_1} \) [recall that \( G(i_1, j_1)S_{i_1} = \sum_c O_{j_1 c}I_{i_1 c}/O_c \),
\[
\sum_c X_{I,J,c}/O_c = G(i_1, j_1) S_{j_1} + \sum_c \left[ O_{i_1c} \left( \sum_{j \neq j_1} I_{je} \right) + \left( \sum_{i \neq i_1} O_{ie} \right) \left( \sum_{j \in J} I_{je} \right) \right]/O_c,
\]

then go back to calculating \( G(I, J) \), and sub that term back in,

\[
G(I, J) = \frac{1}{S_i} \left( \sum_c \left[ \frac{X_{I,J,c}}{O_c} \right] \right)
\]

\[
G(I, J) = \frac{1}{S_i} \left( G(i_1, j_1) S_i + \sum_c \left[ O_{j_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right) + \left( \sum_{j \neq j_1} \sum_{i \in I} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right) \right]/O_c \right)
\]

\[
G(I, J) = \frac{1}{S_i} G(i_1, j_1) S_i + \frac{1}{S_i} \sum_c \left[ O_{j_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right) + \left( \sum_{j \neq j_1} \sum_{i \in I} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right) \right]/O_c
\]

Keep going,

\[
G(I, J) = \frac{S_i}{S_i} G(i_1, j_1) + \frac{1}{S_i} \sum_c \left[ O_{j_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right) + \left( \sum_{j \neq j_1} \sum_{i \in I} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right) \right]/O_c
\]

\[
G(I, J) = w_{i_1} G(i_1, j_1) + \frac{1}{S_i} \sum_c \left[ O_{j_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right) + \left( \sum_{j \neq j_1} \sum_{i \in I} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right) \right]/O_c
\]

Switch the \( c \) and \( i, j \) sums around.

\[
G(I, J) = w_{i_1} G(i_1, j_1) + \frac{1}{S_i} \sum_c O_{i_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right)/O_c + \frac{1}{S_i} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right)/O_c \right]
\]

\[
G(I, J) = w_{i_1} G(i_1, j_1) + \frac{1}{S_i} \sum_c O_{j_1c} \left( \sum_{i \neq i_1} \sum_{i \neq i_1} I_{ic} \right)/O_c + \frac{1}{S_i} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \left( \sum_{i \in I} I_{ic} \right) \right)/O_c \right]
\]
(88) \[ G(I, J) = w_{i_1}G(i_1, j_1) + \frac{1}{S_I} \sum_c \left( \sum_{i \neq i_1} O_{j_1c} I_{ic} \right) + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(89) \[ G(I, J) = w_{i_1}G(i_1, j_1) + \frac{1}{S_I} \sum_{i \neq i_1} \sum_c O_{j_1c} I_{ic} / O_c + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(90) \[ G(I, J) = w_{i_1}G(i_1, j_1) + \frac{1}{S_I} \sum_{i \neq i_1} \sum_c O_{j_1c} I_{ic} / O_c + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(91) \[ G(I, J) = w_{i_1}G(i_1, j_1) + \sum_{i \neq i_1} S_i \left( \sum_c O_{j_1c} I_{ic} \right) + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(92) \[ G(I, J) = w_{i_1}G(i_1, j_1) + \sum_{i \neq i_1} w_{i_1} G(i_1, j) + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(93) \[ G(I, J) = \sum_{i \in I} w_{i_1} G(i_1, j) + \frac{1}{S_I} \sum_c \left[ \left( \sum_{j \neq j_1} O_{jc} \right) \left( \sum_{i \in I} I_{ic} \right) \right] / O_c \]

(94)

So, applying the same logic to each element \( j \neq j_1 \) (specifically, \( j \in J \backslash \{j_1\} \)), this gives

(95) \[ G(I, J) = \sum_{i \in I} \sum_{j \in J} w_{i_1} G(i, j) \]

Which is exactly like the theoretical counterpart,

(96) \[ \Gamma_{IJ} = \sum_{i \in I} w_{i_1} \left( \sum_j \gamma_{ij} \right) \]

(97) \[ G(I, J) = \sum_{i \in I} w_{i_1} \left( \sum_j G(i, j) \right) \]
9.3. **Input, linkage and output correlations.** I construct three measures that measure how connected any two plants are, in terms of using the same commodities (input correlation, $\rho_{ij}^I$), making the same commodities (similar outputs, $\rho_{ij}^O$), or one plant that uses a commodity that the other makes (direct linkage correlation, $\rho_{ij}^L$).

First, for plant $i$, write the vector of commodity outputs as $O_i$ and the vector of commodity inputs as $I_i$. Then plants $i$ and $j$ have similar outputs if their output vectors overlap, in the sense that the dot-product is positive:

\[
\rho_{ij}^O = \cos \theta_{ij}^O = \frac{\sum_c O_{ci}O_{cj}}{\sqrt{\sum_c O_{ci}^2} \sqrt{\sum_c O_{cj}^2}}
\]

Two plants have similar inputs if their input vectors overlap (adjusted for the shares of intermediates in total output),

\[
\rho_{ij}^I = \cos \theta_{ij}^I = \frac{\sum_c I_{ci}I_{cj}}{\sqrt{\sum_c O_{ci}^2} \sqrt{\sum_c O_{cj}^2}}
\]

Two plants have direct linkages if one plant makes a commodity the other plant uses, and vice versa,

\[
\rho_{ij}^L = \cos \theta_{ij}^L = \frac{\sum_c I_{ci}O_{cj}}{\sqrt{\sum_c T_{ci}^2} \sqrt{\sum_c O_{cj}^2}}
\]

In the vector notation used in the text, these are

\[
\rho_{ij}^O = \frac{O_i \cdot O_j}{||O_i||_2 \times ||O_j||_2}
\]

\[
\rho_{ij}^I = \frac{I_i \cdot I_j}{||I_i||_2 \times ||I_j||_2}
\]

\[
\rho_{ij}^L = \frac{I_i \cdot O_j}{||I_i||_2 \times ||O_j||_2}
\]
where $||X||_2$ is the Euclidean norm of the vector $X$.

<table>
<thead>
<tr>
<th>IO Measure</th>
<th>No. Obs</th>
<th>Mean</th>
<th>Median†</th>
<th>99th pct†</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct input (plant), $g_{ij}$</td>
<td>11597686</td>
<td>.0003</td>
<td>0</td>
<td>.0026</td>
<td>.0064</td>
</tr>
<tr>
<td>Direct input (industry), $g_{IJ}$</td>
<td>28738</td>
<td>.0025</td>
<td>0</td>
<td>.0784</td>
<td>.01822</td>
</tr>
<tr>
<td>Outdegree (plant), $d_i$</td>
<td>17211</td>
<td>.1984</td>
<td>.0003</td>
<td>3.566</td>
<td>1.787</td>
</tr>
<tr>
<td>Outdegree (industry), $d_I$</td>
<td>230</td>
<td>.3087</td>
<td>.1180</td>
<td>3.265</td>
<td>.6028</td>
</tr>
<tr>
<td>Input correlation (plant), $\rho_{ij}^I$</td>
<td>11419288</td>
<td>.0359</td>
<td>.0025</td>
<td>.3727</td>
<td>.07753</td>
</tr>
<tr>
<td>Linkage correlation (plant), $\rho_{ij}^L$</td>
<td>11183779</td>
<td>.001829</td>
<td>0</td>
<td>.1126</td>
<td>.02148</td>
</tr>
<tr>
<td>Output correlation (plant), $\rho_{ij}^O$</td>
<td>10960722</td>
<td>.06251</td>
<td>0</td>
<td>1</td>
<td>.1959</td>
</tr>
</tbody>
</table>

| No. of possible plant pairs 324 million |
| No. of possible industry pairs 53824   |

Table 4. Summary statistics: direct requirements $G = [g_{ij}]$ and commodity correlations. Industries are 1980 SIC. All data from 1992. Measures from other years are similar. †: for confidentiality reasons, I do not report exact percentiles, and instead report an average value over the 49-51 percentiles for the median, and 98-99.9 percentiles for the 99th percentile.

9.4. Summary statistics, input-output measures.

10. Appendix: Intensive and Extensive Margins of Volatility

In the main text, I assume there is no extensive margin of volatility. One may wonder how the results change if I allow for plant entry and exit. I address this two ways: first, empirically, using a similar decomposition to Di Giovanni, Levchenko and Méjean [12], and also theoretically, adapting a Melitz [32] model to incorporate plant-specific IO characteristics.

10.1. Intensive vs. extensive margins in the data. First, write sales of plant $i$ at year $t$ as $s_{i,t}$. Let $I_t$ be the set of plants operating in year $t$, and $I_{t/t-1}$ be the set of plants operating in both years $t$ and $t-1$. Then the log-difference aggregate growth rate of sales is

$$\tilde{g}_{At} \equiv \ln \left( \sum_{i \in I_t} x_{it} \right) - \ln \left( \sum_{i \in I_{t-1}} x_{it-1} \right)$$

$$= \ln \left( \frac{\sum_{i \in I_t} x_{it}}{\sum_{i \in I_{t-1}} x_{it-1}} \right) - \left[ \ln \left( \frac{\sum_{i \in I_{t/t-1}} x_{it}}{\sum_{i \in I_t} x_{it}} \right) - \ln \left( \frac{\sum_{i \in I_{t/t-1}} x_{it-1}}{\sum_{i \in I_{t-1}} x_{it-1}} \right) \right]$$

$$= g_{At} - \ln \left( \frac{\nu_{t,t}}{\nu_{t,t-1}} \right)$$
where $g_{At}$ is the intensive margin of growth and the other term is the extensive margin of growth. Now aggregate volatility is

\begin{equation}
\sigma^2_A = \sigma^2_A + \sigma^2_\nu - 2\text{Cov}(g_{At}, g_\nu)
\end{equation}

Calculating each of these in the data, we see that the extensive margin matters little (consistent with the results in Di Giovanni, Levchenko and Méjean [12].

<table>
<thead>
<tr>
<th>S.D.</th>
<th>Relative S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Volatility ($\sigma_A$)</td>
<td>.065</td>
</tr>
<tr>
<td>Intensive Volatility ($\sigma_A$)</td>
<td>.066</td>
</tr>
<tr>
<td>Extensive Volatility ($\sigma_\nu^2$)</td>
<td>.009</td>
</tr>
</tbody>
</table>

### 10.2. A Melitz model with IO

Next, we can adapt a Melitz model to include plant-specific IO characteristics. Profit for a plant $i$ with productivity $z_i$ and outdegree $\gamma_i$ is

\begin{equation}
\pi_i = p_i q_i - \eta \left( f + \frac{q_i}{z_i} \right)
\end{equation}

Demand for $q_i$ is

\begin{equation}
q_i = \gamma_i X \left( \frac{p_i}{P} \right)^{-\sigma}
\end{equation}

where $X$ is total demand for the CES composite (used for both final and intermediate goods), including all units used to pay fixed costs. Prices are constant markups over marginal cost

\begin{equation}
p_i = \left( \frac{\sigma}{\sigma - 1} \right) \frac{\eta}{z_i}
\end{equation}

And the overall price index is
\[ P = \chi_P^{1/\beta} \left( \sum_{i=1}^{N} \gamma_i z_i^{\sigma - 1} \right)^{\frac{1}{\beta(1-\sigma)}} \]

And \( \chi_P = \left( \frac{\sigma}{\sigma - 1} \right) \beta^{-\beta} (1 - \beta)^{\beta - 1} \).

The distribution of plant size is again determined by the distribution of \( \gamma_i z_i^{\sigma - 1} \). Write sales as \( s_i = p_i q_i \), which is

\[ s_i = \left( \frac{\sigma}{\sigma - 1} \right) \left( \frac{\eta}{z_i} \right) \gamma_i X \left( \frac{P_i}{P} \right)^{-\sigma} \]

\[ s_i = \left( \frac{\sigma}{\sigma - 1} \right) \left( \frac{\eta}{z_i} \right)^{1-\sigma} \gamma_i X P^\sigma \]

\[ s_i = \left( \frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \eta^{1-\sigma} X P^\sigma \gamma_i z_i^{\sigma - 1} \]

\[ v_i = \frac{\gamma_i z_i^{\sigma - 1}}{\sum_{j=1}^{N} \gamma_j z_j^{\sigma - 1}} \]

It is unlikely that the extensive margin of volatility matters because entry and exit happen at the lower bound of the distribution, so those small adjustments should not matter for aggregate volatility. This is consistent with the empirical evidence, which stands in contrast to mechanisms suggested by Baqee \[5\]. Further investigation is required.

**CDER at Statistics Canada and the University of Calgary**

_E-mail address: jesse.tweedle@gmail.com_