The Microfoundations of Aggregate Volatility: Productivity, Geography, or Network Asymmetry?

Job Market Paper

Jesse Tweedle*

November 20, 2016

Abstract

Theories of microfoundations for aggregate volatility rely on a skewed individual size distribution, termed *granularity* (in the sense that there are big firms that can't be subdivided). If so, what causes granularity? The firm size distribution may be skewed because of a skewed productivity distribution, geographic agglomeration, or skewed demand characteristics that result in an asymmetric production network. What matters more? I use detailed data on firm-firm trade in Canada to estimate a model in which productivity, geography, and demand characteristics vary independently. This allows me to recover unobserved demand characteristics from the observed production network, which conflates productivity, demand and geography. I find that the demand network accounts for 60% of the firm size distribution, productivity and geography explain little, and that approximately half of the demand network effect is due to higher order network interconnections. Microeconomic shocks can account for approximately 32% of aggregate volatility, and removing variation in the demand network would reduce aggregate volatility by 25%.

Keywords: Microfoundations, Firm production, Firm size, Input output, Networks, Herfindahl Index, Economic Geography

JEL classifications: D2; D57; E1; L1; L2; R1

*Email: jesse.tweedle@gmail.com. Statistics Canada, Economic Analysis Division and the University of Calgary, Department of Economics. Find the newest version at http://jesse.tw.

1 Introduction

Are idiosyncratic shocks sources of aggregate volatility? How do they propagate across the economy? The idea that aggregate demand and supply shocks are the only source of volatility in the economy leaves important mechanisms in the shadows. Previously, the possibility that idiosyncratic shocks to firms can cause aggregate fluctuations had been debunked by the law of large numbers; how can tens of thousands, or millions, of uncorrelated shocks average out to anything but zero? However, if the economy is structured in such a way that certain firms are disproportionately large, the law of large numbers argument may fail. Idiosyncratic shocks to these firms may propagate through the economy and make up a substantial portion of aggregate volatility.

Theories of microfoundations of aggregate volatility all rely on this skewness of the firm size distribution, called *granularity*. If granularity allows idiosyncratic shocks to cause aggregate fluctuations, what causes granularity? In this paper, I study three sources of skewness in the firm size distribution— productivity, the firm-firm production network, and geography—and how they affect aggregate volatility.

A skewed productivity distribution is the usual culprit in models with skewed size distributions: the standard source of firm heterogeneity is a Pareto productivity distribution in a simple Melitz (2003) model. However, the size distribution is also skewed if the firm-firm production network is skewed. A firm may produce goods that are required inputs in production for a substantial fraction other firms, which causes that firm to be very large. Finally, large firms tend to cluster in urban areas due to high trade costs and other benefits of agglomeration, so geography skews the firm size distribution as well. The first and second order effects of the interactions between these factors turn out to be very important. For instance, a firm may have low productivity and few customers in the production network, but if those customers are themselves large and this firm is the only close supplier, it will be large in turn.

The key to differentiating these features is to use a model in which they vary independently, and, more importantly, data that allows me to calibrate and estimate the model. I extend a model of firm-firm trade with firm heterogeneity in not only productivity, but pair-specific demand characteristics that define the production network and geography as well. The most important thing to note is that production networks are endogenous—using expenditures shares as measures of input-output requirements, such as those used in industry level input-output tables, conflates the three factors I study here. Recovering the true source of granularity requires data and a model that differentiate these things. After doing so, I perform counterfactuals on the parameters to see how changing the underlying productivity, demand and geography would change the size distribution and aggregate volatility.

The model extends a simple firm-firm Cobb-Douglas production network model to incorporate productivity differences, trade costs and substitutability across firms. Each firm is in a region, and total regional income is the sum of all value-added in that region; regional income can be spent on goods from any firm, subject to trade costs. The market structure determines the skewness of the size distribution, which in turn affect the way idiosyncratic shocks propagate across the economy. Almost all of the shocks are transmitted through input-output links, though the reasons the production links exist in the first place are determined by productivity, demand characteristics and geography.

One must note that it isn't enough to use industry input-output characteristics to define the economy, in the model or in the data. First, in the model, using industry input-output shares as demand characteristics implies all within-industry firm heterogeneity cannot be due to demand characteristics, which is refuted by the data. Using industry-level IO data also implies that within a pair of industries with an input-output linkage (i.e., a positive direct requirement coefficient, which is the term for the expenditure share in the industry-by-industry input-output tables produced by national statistical agencies), all firms trade with each other. And in any industry with an input-output linkage with itself, all firms within the industry trade with each other, including itself. This is again refuted by the data, which I turn to next.

The microdata[1] are from several sources: the Annual Survey of Manufacturing (ASM), the Surface Transportation File (STF), the detailed-confidential Input-Output and Supply-Use tables (IOT), the Inter-Provincial Trade Flow file (IPTF), and the Import-Export Register (IER). For more details of each database and on

[1]Here I make the first distinction between firms and establishments. I use the term 'firm' to be consistent with previous work on firm-to-firm production networks, firm size distributions and aggregate volatility, and because it's shorter and easier to say and write than 'establishment.' This is convenient for the writer and reader when describing establishment-establishment trade. Nevertheless, the data are at the establishment level. Firm-level microdata are difficult to study geographically, because they typically do not have 'locations' in the physical sense used in models of economic geography. When using administrative data, the firm unit is defined by tax accounting standards, not economic or physical standards, and so firms are not guaranteed to have actual physical locations. Instead, they have corporate headquarters that may have complex legal and operational heirarchies and no geographic information on economic activity.

data construction and benchmarking, see Appendix 8. Here, I describe the essential elements of the data.

The ASM is an establishment survey covering 99% of industrial output in Canada. The survey includes data on shipments by destination province (and exports), and inputs and outputs by commodity. The location of an establishment is defined by a postal code, which can be as detailed as a single building in urban areas. The STF is a transaction-level database of goods shipments in Canada, including trade to and from the United States. Each shipment includes value, tonnage, commodity classification, mode, and postal code geographic detail for origin and destination. The IOT, IPTF and IER serve to supplement the identification of the firm-firm production network and provide official statistics with which to benchmark the data. In other words, identification of the firm-firm production network is only valid if the data are internally consistent (the data are consistent with each other, given the model) and externally consistent (the data are consistent with the Canadian national accounts).

The data show clear skewness in the productivity and size distributions, a very asymmetric firm-firm production network, and very dense geographic agglomeration. The empirical strategy works in two parts. First, I start with the observed data and use the model to uncover unobserved demand characteristics and the implied demand network. Next, after uncovering the parameters that govern the firm size distribution, I turn to calculating aggregate volatility. It is difficult to infer the parameters that determine idiosyncratic volatility due to the general equilibrium nature of shock propagation: if the granularity hypothesis is true, a large aggregate shock is the result of idiosyncratic shocks, not evidence against them. This is called the "reflection problem." I attempt to circumvent this problem by estimating uncorrelated productivity shocks and using those as a lower bound for the contribution of idiosyncratic shocks to aggregate volatility.

After calibrating the model, I can investigate the effect of each feature's skewness on the economy. For instance, what happens to aggregate volatility if we remove the variation in productivity across firms? What if we remove variation in geography or the demand network instead? The relative changes in aggregate fluctuations after changing the distributions of each feature give important insights into the economy. Perhaps surprisingly, removing skewness in productivity and geography actually increases skewness in the size distribution, which would increase aggregate volatility by 11% and highlighting the importance of the complexity of the network.

Research on idiosyncratic shocks and aggregate volatility restarted in earnest when Gabaix (2011) and Acemoglu et al. (2012) revived the debate between Horvath (1998, 2000) and Dupor (1999) on whether idiosyncratic shocks average out in aggregate. Gabaix (2011) proposes that the largest, granular firms are so big that their idiosyncratic shocks do not average out at the aggregate level. Acemoglu et al. (2012) 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 firm-level shocks rely on firm-level input-output variation within industries. Furthermore, I incorporate geography to seriously study agglomeration, productivity, and network asymmetry and the interactions between them that determine the observed production network, which previous work assumes is exogenous.

The most direct predecessors of this paper are empirical studies of aggregate fluctuations. Starting with Shea (2002), and continuing most recently with Foerster et al. (2011), Di Giovanni et al. (2014), Acemoglu et al. (2015). Foerster et al. (2011) 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 et al. (2014) 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 firm-level network data to establish the determinants of firm covariances, using deeper levels of disaggregation to examine both covariances (firm level to establishment level) and input-output networks (industry level to establishment level). As well, I study the determinants of the network itself, adding geography and productivity as well as demand characteristics, 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 (Di Giovanni et al., 2011; Di Giovanni and Levchenko, 2012, 2013), or studies on general characteristics and theories of the size distribution itself (Luttmer, 2007). I add an endogenous network perspective to this research and use it to further explore the determinants of the firm size distribution and the sources of granularity. My work also fits naturally with Hottman et al. (2016), who use detailed retail scanner 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 (2014) also provide evidence that demand characteristics are the main source of firm heterogeneity, in contrast to standard Melitz applications. In my case, the input-output requirements of downstream firms translate into a dominant source of firm appeal, and therefore are a large determinant of firm size.

My argument is also related to recent work on customer-supplier relationships, especially Barrot and Sauvagnat (2015), who study the disruption of production networks after natural disasters. In addition, research on customer-supplier relationships in Japan (Bernard et al., 2015; Carvalho et al., 2014) and the US (Atalay et al., 2011) suggests larger firms have different input-output characteristics than smaller firms. Most recently, Lim (2016) studies creation and destruction of firm-firm relationships, although he notes the difficulty of matching geographic characteristics. 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 firms. To this research, I add a characterization of the geography and complexity of the production network in Canada.

These papers are also part of a recent wave of interest in the formation and effects of social and economic networks. Carvalho and Voigtländer (2014), Oberfield (2011) and Jones (2011) 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. (2013), Elliott et al. (2014). Other applications and background on network measures used in this paper can be found in Jackson (2010).

In Section 2, I present a simple, but necessary, extension to the Cobb-Douglas input-output model used in Acemoglu et al. (2012) to allow three features crucial to reconcile the empirical regularities in the economy: I incorporate productivity variation and substitutability across firms, geographic characteristics, and unobserved demand network characteristics. The asymmetry of the production network, the productivity distribution and geography combine to determine firm sizes, which is the key to evaluating the granularity of the economy and its effect on aggregate volatility. In Section 3, I present the firm-level volatility and production network data. I document an unbalanced production network at a disaggregated level, with a few firms acting as central suppliers to the network, as well as skewness in productivity and geographic agglomeration.

In Section 4, I calibrate the model to uncover the underlying demand characteristics network from the endogenous, observed input-output network and evaluate the competing theories of the microfoundations of aggregate fluctuations. In Section 5, I present results. Previewing the main calibration results, the productivity distribution is not heterogenous enough to account for the asymmetry in the observed production network network. The majority of the firm size distribution is due to the underlying demand network, consistent with results in Holmes and Stevens (2014) that challenge the reliance of the firm size distribution on productivity alone. In addition, higher order interconnections are economically significant determinants of the firm size distribution. Turning to the macroeconomy, I find idiosyncratic shocks can account for approximately 32% of aggregate volatility, and that removing variation in productivity and geography would actually increase firm size skewness and aggregate volatility by 11%.

Section 6 concludes, and several Appendices follow, giving details on theory, measurement and development of the firm-to-firm production network, and other necessary but tedious details.

2 Model

To study the relationships between volatility, endogenous asymmetric production networks and the factors that determine them, I adapt the sectoral model of Acemoglu et al. (2012), which is itself based on Long and Plosser (1983). There are three key additions.

First, I study individual firms and not sectors. Although technically easy (e.g., relabeling sectors as firms), it puts the focus on the determinants of granularity—is it the production network, or productivity, or geography? 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 incorporate trade costs into the model by dividing the economy into multiple regions, each with their own individual demand and geographic characteristics.

Third, and most importantly, I relax the assumption that the production network is exogenous. In my model, a firm may be a central supplier of the network because it is a required input in many other products (it has many high unobserved demand characteristics) or because it is so productive that many other firms substitute toward it, or because it is located close to important transportation arteries or other industrial areas.

To introduce these features, I need a model in which productivity, geography and unobserved demand characteristics can vary independently to create an observed firm-firm production network that I can take to the data. I give a table of important notation in Table 6 in Appendix 7. In general, I use capital letters to refer to matrices, lowercase to refer to vectors and elements of vectors and matrices, latin characters for observed variables and greek characters for the equivalent unobserved variables. For example, $G = [g_{ij}]$ is the observed expenditure share matrix, $\Gamma = [\gamma_{ij}]$ is the unobserved demand matrix.

2.1 Model Basics

To start, there are R regions. A representative household in a specific region r inelastically supplies a labour L_r, and has Cobb-Douglas preferences over N different goods (I relax this assumption later, but it is useful to focus first on firm-firm demand characteristics),

$$u_r(c_r) = \prod_{i \in N} c_{ri}^{\lambda^{ri}} \tag{1}$$

where c_{ri} is region r's consumption of good i. There is free migration between regions, so that the wage w in equilibrium is constant across regions. Later, I normalize $w = 1$.

Each good is produced by a single firm using Cobb-Douglas combination of labour and a firm-specific intermediate input which is itself a CES aggregate of other products,

$$q_i = z_i l_i^{\beta} \left(\sum_{j \in N} \gamma_{ij}^{\frac{1}{\eta}} q_{ij}^{\frac{\eta-1}{\eta}} \right)^{\frac{(1-\beta)\eta}{\eta-1}} \tag{2}$$

where z_i is productivity, β is the labour share in production, q_{ij} is the quantity of firm j's product demanded by firm i, and η is the elasticity of substitution between intermediates. The crucial part of production is $\gamma_{ij} \geq 0$, which is the exogenous direct input coefficient. If γ_{ij} is high, then independent of firm j's productivity, firm i requires a lot of firm j's input to produce. If γ_{ij} is low but positive, then firm i may still demand a lot of q_{ij} if firm j is very productive. In this way, the endogenous production network is determined jointly by productivity, substitutability and unobserved demand characteristics. Firm i can only draw labour from its region r. There can be multiple firms in any given region.

Trade between regions and firms comes at a cost. For instance, firm i pays for quantity q^s_{ij} to be shipped, but only receives $q^d_{ij} = q^s_{ij}/\tau_{ij}$. In other words, the buyer pays a unit price $\tau_{ij}p_{ij}$, and the total expenditure that i spends on j is $\tau_{ij}p_{ij}q^d_{ij}$.

With perfect competition, prices equal marginal costs for firm i, incorporating the iceberg trade costs,

$$p_i = C z_i^{-1} \left(\sum_{j \in N} \gamma_{ij} (\tau_{ij} p_j)^{1-\eta} \right)^{\frac{1-\beta}{1-\eta}} \tag{3}$$

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

$$p_{ri} = \prod_{i \in N} \left(\frac{\tau_{ri} p_i}{\lambda_{ri}} \right)^{\lambda_{ri}} \tag{4}$$

The full derivation of the model, along with any extra notation needed, can be seen in Appendix 7.1.

2.2 Important model features

The model is simple, but it delivers several important results that are typically ignored when looking at models of production networks.

Remark 1 *Observed expenditure shares depend on productivity, geography and unobserved demand characteristics.*

The input-output tables provided by statistical agencies give an expenditure share of industry i on goods from industry j. The firm production network I detail in the previous section is constructed in a similar way, an expenditure share of firm i on firm 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},

$$g_{ij} = \frac{\tau_{ij} p_j q_{ij}}{p_i q_i} \tag{5}$$

In equilibrium, this simplifies to

$$g_{ij} = (1-\beta)\left[\frac{\gamma_{ij}(\tau_{ij} p_j)^{1-\eta}}{\sum_{k\in N} \gamma_{ik}(\tau_{ik} p_k)^{1-\eta}}\right] \tag{6}$$

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

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. (2012) 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 production network 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 production network. Hence their claim that the production network is the main determinant of aggregate volatility. Here I show that the same holds for the observed production network. That is, an empirical association between the influence vector and observed production network does not tell you the effect of the production network on volatility, because the observed network may be entirely determined by productivity. Write the system of market clearing equations,

$$\sum_{r\in R} p_{ri} c_{ri} + \sum_{j\in N} \tau_{ij} p_i q_{ji} = p_i q_i, \text{ for } i \in N \tag{7}$$

And rewrite in terms of g_{ij} using (5),

$$\sum_{r \in R} p_{ri} c_{ri} + \sum_{j \in N} g_{ji} p_j q_j = p_i q_i, \text{ for } i \in N \tag{8}$$

Then a similar derivation to Acemoglu et al. (2012) (see Appendix 7.2) gives you the influence vector as a function of the matrix of observed expenditure shares $G = [g_{ij}]$, observed demand shares $A = [a_{ri}]$, and regional labour $L = (L_1, \ldots, L_R)$,

$$v' = \beta \left(\frac{L'}{\mathbf{1}'L} \right) A(I - G)^{-1} \tag{9}$$

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

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

If $\gamma_{ij} = \tau_{ij} = 1$, then all firms use the same intermediate bundle and face the same intermediate input price. This means the expenditure share equation (5) reduces to

$$g_{ij} = (1 - \beta) \left[\frac{z_j^{\eta - 1}}{\sum_{k \in N} z_k^{\eta - 1}} \right] \tag{10}$$

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 production network, even though the underlying demand characteristics are homogenous.

Example 2 *Suppose $z_i = 1$ for all $i \in N$ and $\tau_{ij} = 1$ for all $i, j \in N$. Then there is no productivity or geography variation, and all of the observed production network characteristics are due to the exogenous demand characteristics.*

When productivities and trade costs are identical across all firms, the expenditure share terms reduce to

$$g_{ij} = (1-\beta)\left[\frac{\gamma_{ij} p_j^{1-\eta}}{\sum_{k\in N}\gamma_{ik} p_k^{1-\eta}}\right] \tag{11}$$

where the prices can be written as a recursive function of prices and demand parameters, which implies the expenditure shares are determined only by demand parameters.

2.3 Outdegree and unbalanced production networks

An unbalanced production network is one in which individual firms are central suppliers to the entire economy. The easiest way to ask how central a firm is by adding up the demand parameters of a firm's customers (unobserved outdegree, δ_i), or the observed expenditure shares of a firm's customers (observed outdegree, d_i),

$$\delta_i = \sum_{j\in N}\gamma_{ji};\ d_i = \sum_{j\in N} g_{ji}, \tag{12}$$

Example 3 *Suppose $\gamma_{ij} = \delta_j/N$, for $j \in N$, and $\tau_{ij} = 1$ for $i, j \in N$.*

Expenditure shares are

$$g_{ij} = (1-\beta)\left[\frac{\delta_j z_j^{\eta-1}}{\sum_{k\in N}\delta_k z_k^{\eta-1}}\right] \tag{13}$$

Observed outdegree is

$$d_i = (1-\beta)\left[\frac{\delta_i z_i^{\eta-1}}{(1/N)\sum_{k\in N}\delta_k z_k^{\eta-1}}\right] \tag{14}$$

And one element of the influence vector is

$$v_i = \frac{\beta}{N} + (1-\beta)\left[\frac{\delta_i z_i^{\eta-1}}{\sum_{k\in N}\delta_k z_k^{\eta-1}}\right] \tag{15}$$

This examples highlights the dependence of the influence vector on productivity and the unbalanced production network—the distribution of v_i is determined by the distribution of $\delta_i z_i^{\eta-1}$. 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 firms grows large. However, as the number of firms 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.

2.4 Geography

Firms have a tendency to cluster in specific geographic areas, leaving groups of firms in close proximity (and low trade costs), while other firms are spread out and isolated over the rest of the regions in the economy. A firm that is closer to other firms, especially large ones, tend to be large themselves and also tend to purchase inputs from their neighbours.

Example 4 *Suppose $\tau_{ij} = \tau_j$, for $j \in N$, and that productivity and the exogenous demand parameters are constant, $z_i = 1$, for $i \in N$, and $\gamma_{ij} = 1$ for $i, j \in N$.*

The expenditure share depends solely on trade costs, and firm i buys the most from the firm with the lowest τ_j:

$$g_{ij} = (1 - \beta) \left[\frac{\tau_j^{1-\eta}}{\sum_{k \in N} \tau_k^{1-\eta}} \right] \tag{16}$$

In addition, trade costs between firms are typically clustered so that groups of firms are closer to each other and all have low τ_{ij} values relative to firms in remote areas. This has a special effect on the size distribution of firms. While productivity and demand characteristics usually have distributions with very skewed right tails, trade costs can have the opposite effect—there are a high number of firms concentrated in center areas of the economy, with other small firms in remote areas. Since trade costs are bounded below by 1 and are inversely related to g_{ij}, trade costs have a much more significant effect at the middle and lower-end of the size distribution, not the right tail.

3 Data

The microdata are from several sources: the Annual Survey of Manufacturing (ASM), the Surface Transporation File (STF), the detailed-confidential Input-Output and

Supply-Use tables (IOT), the Inter-Provincial Trade Flow file (IPTF), and the Import-Export Register (IER). For more details of each database and on data construction and benchmarking, see Appendix 8.

The establishment data is from the ASM, which covers 99% of industrial output in Canada. It is a long-running annual panel of manufacturing establishments, including data on shipments by destination province (and exports), and inputs and outputs by commodity. The location of an establishment is defined by a postal code, which can be as small as a single building in urban areas, and larger in rural areas.

I analyze volatility over the period from 1990 to 2010, covering several volatile periods in Canadian manufacturing, including in the early 1990s, as well as 2001 and the Great Recession. Aggregate volatility, measured by the standard deviation of the aggregate growth rate of total output, over this period was approximately 6% in manufacturing, slightly higher than the overall for Canada during the same period, around 4%.

The trade data is from the STF, a transaction-level database of goods shipments in Canada, including trade to and from the United States. Each shipment includes value, tonnage, commodity classification, mode, and postal code geographic detail for origin and destination. This allows the identification of origin and destination establishments from the ASM, as well as establishment origins and final demand destinations.

In many empirical studies, the observations need only be consistent in one or two dimensions—e.g., the distribution of value-added should make sense, or there should be consistent classifications of firm and industry identifiers over time in a panel. However, in a model with firm-firm trade, it is also important to make sure the interactions between firms are also consistent, meaning every firm must have at least one supplier and at least one customer (which may be final demand only). To achieve this empirical goal, the IOT and IPTF serve to supplement the identification of the firm-firm production network and provide official statistics with which to benchmark the data. In other words, identification of the firm-firm production network is only valid if the data are internally consistent (the data are consistent with the model) and externally consistent (the data are consistent with the Canadian national accounts).

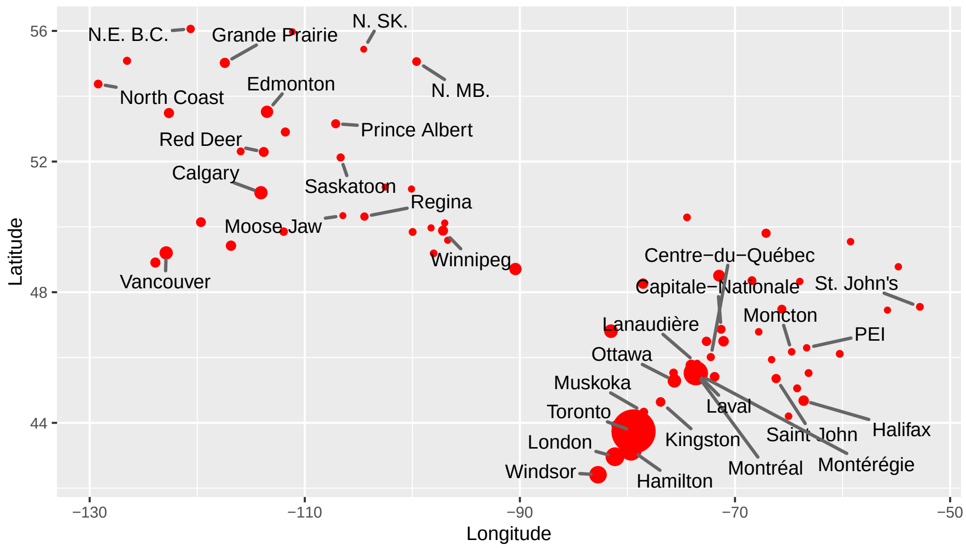

Figure 1: Economic Regions and their volatilities

Notes: The x- and y-axes are the longitude and latitude of the population weighted centroid of each region. Larger circles represent Economic Regions with higher volatilities, with Toronto as the highest. Volatility is calculated as the standard deviation of weighted growth rates of manufacturing value-added within each Economic Region, 1990-2010.

3.1 Motivating empirical features

Three empirical features of the data motivate the study of productivity, geography and the production network. First, there is considerable skewness in each measure. Next, there are important interactions between them: large firms are more productive, tend to cluster together, and are central suppliers to the economy. Third, these factors are empirically related to aggregate volatility.

I'll start with describing the geographic features of aggregate volatility. To get a sense of the geographic distribution of economic activity and its relationship to aggregate volatility, Figure 1 displays the standard deviation of value-added growth in each Economic Region in Canada and their geographic positions (I focus on Economic Regions in the motivating empirical work for two reasons: first, to reduce the dimension of the visual display, and second, to comply with confidentiality restrictions on the data). The largest source of volatility is Toronto, along with the surrounding regions. This area has the highest concentration of overall activity, trade, input-output links, and volatility. Of course, variance is only one aspect of aggregate volatility. Covariance between regions can matter just as much, especially for understanding

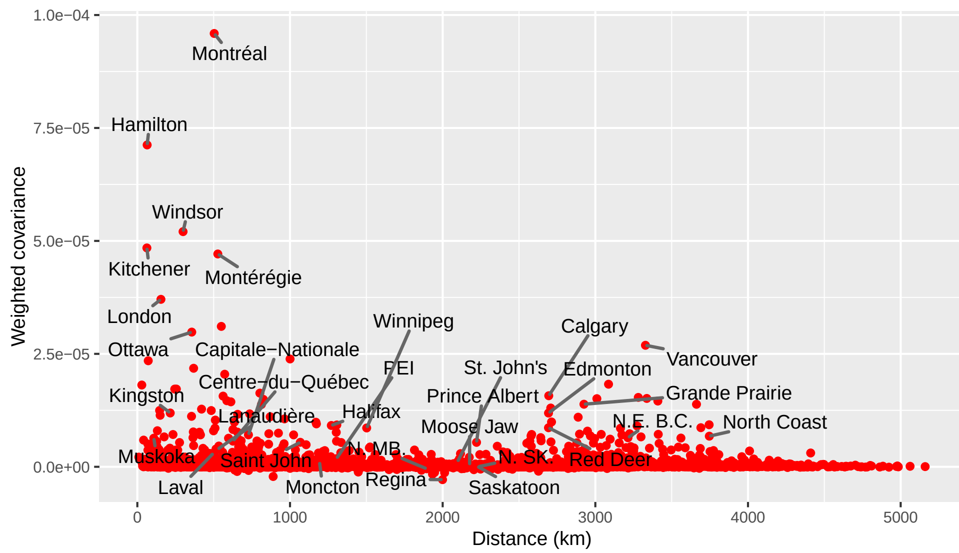

Figure 2: Economic Region pairs, covariance vs. distance

Notes: Labeled pairs are Toronto-y observations, for label y. Covariance is calculated as the covariance of weighted growth rates of manufacturing value-added between each Economic Region pair, 1990-2010. Distance is the mean distance of shipments in the STF between the regions, weighted by value.

the interactions between geography and the rest of the economy. Figure 2 shows the relationship between all components of aggregate volatility and geography, plotting covariance between regions vs. distance between regions them. The geographic concentration is clear, with Toronto's covariance with nearby regions the obvious outliers again. However, other important features emerge: there is a bump in covariances around the 3000km mark, roughly the distance between Toronto and cities in Western Canada. These covariances show that although closer regions are larger and comove more than distant regions, this effect can be outweighed by other factors.

I've shown considerable geographic patterns in weighted volatility, but does it have anything to do with firm-firm trade? In Figure 3, I show the geographic density of output, along with the density of trade within and across Canada. Here, we dispense with Economic Regions and work with the microdata directly. In red, I show the density of value-added in by longitude. Although the Toronto area is clearly the largest, there are also peaks in farther West and East. However, the density of trade to and from Toronto is much, much more concentrated. In blue, I show the value-weighted density of imports and exports of any goods shipment coming in and

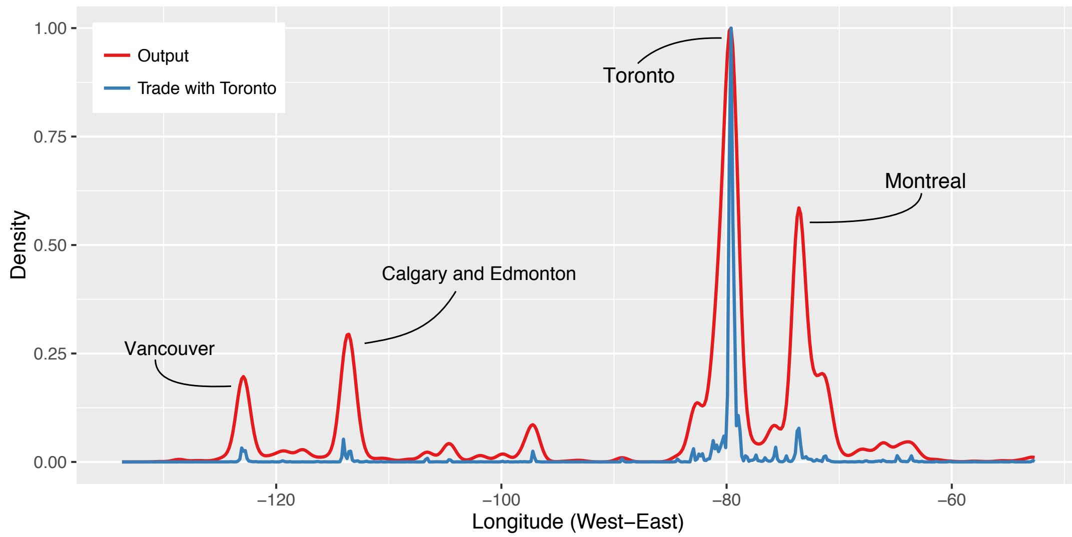

Figure 3: Geographic density of output across Canada compared to the geographic density of trade with Toronto.

Notes: The red line is the longitudinal geographic density of manufacturing output in Canada in the ASM, 2010. The blue line is the longitudinal geographic density of the value of trade transactions in the STF with an origin *or* destination in Toronto, 2010. The peak of each density is normalized to one to make it easy to compare the two.

out of the Toronto Economic Region. Although there are smaller peaks in trade at long distances with the East and West, the far, far greater share of Toronto trade is with itself and nearby regions. These empirical features show that contributions to aggregate volatility are very concentrated geographically, that economic activity concentrated in the same way, and that firm-firm trade is even more concentrated.

3.1.1 Skewed distributions: output, productivity, geography and demand

In this section, I document the skewness in each feature of the economy. For one-dimensional firm measures (i.e., firm size and firm productivity), I measure skewness with the herfindahl, the 90/10 percentile ratio and the slope of the right tail of the distribution on a log-log rank-size plot. Herfindahls are directly related to the granular theory (see Gabaix, 2011), with more concentrated distributions supporting more aggregate volatility. To estimate the shape parameter of the tail of the distribution, following Gabaix and Ibragimov (2011), I trim the distribution to the top 20$^{\text{th}}$

percentile of variable x and estimate

$$\log(\text{rank}(x)_i - 1/2) = \alpha - \beta \log x_i \tag{17}$$

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.

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, and the top *ten* firms in that industry account for the vast majority of its output. The herfindahl of firm sales is 0.048, and the tail parameter of the log-log rank-size plot is 0.99. The firm size distribution is clearly skewed.

I measure firm productivity in several ways. First, labour productivity, defined as total value-added divided by employment. Next, labour productivity, defined as total value-added divided by total payroll. Next, naïve total-factor-productivity, measured as the residual of a log-linear regression of output on employment, capital and total input cost. Finally, the estimation procedure developed by Gandhi et al. (2013), which I refer to as TFP (GNR).

Each method has benefits and drawbacks. The goal is to rely on the robustness of the results to a variety of different productivity estimates, rather than stick to a single productivity estimation procedure. First among the drawbacks, all estimates are of revenue productivity, not physical productivity (see Foster et al., 2008, for a discussion of the relevant differences). The drawback here isn't as stark as it would be in reduced form studies that rely on the difference between revenue TFP and physical TFP, since I can recover the unobserved demand characteristics in the model, conditional on the assumptions. The productivity measures and results are consistent with previous work on firm heterogeneity, specifically that demand characteristics matter more for firm heterogeneity than physical productivity itself. Therefore, although I have no *a priori* justification for only using revenue productivity, the results suggest revenue productivity is a decent measure of productivity, as long as I account for the unobserved demand characteristics in the model.

In addition, both labour productivity measures have the obvious drawback of being partially determined by capital. Using payroll instead of employment tends to reduce this bias (since firms with higher capital-per-worker tend to pay higher wages,

Table 1: Skewness

	Mean	Median	S.D.	90/10	Tail, $\hat{\beta}$
Output ($\$ \times 10^6$)	16.46	2.1	147.03	42.92	0.99
Value added ($\$ \times 10^6$)	6.42	1.1	36.69	49.80	1.05
Value added share	0.55	0.6	0.18		
TFP (Naïve)	1.10	1.0	1.46	2.12	1.98
TFP (GNR)	1.06	1.0	1.17	1.71	1.99
Labour prod. (Emp.)	1.23	1.1	0.85	4.89	3.99
Labour prod. (Pay.)	1.11	1.0	0.60	2.78	3.81
Outdegree	0.45	0.1	2.09	439.11	1.61

Notes: Output and value added are measured in millions of Canadian dollars. The 90/10 ratio is the ratio of the 90th percentile to the 10th percentile of the distribution of the variable. Output, value added, TFP and labour productivity are from the ASM, 2010. The tail parameter is estimated using the method of Gabaix and Ibragimov (2011). Outdegree d_i is calculated with the observed production networks A and G.

which reduces the variation in the payroll-based measure due to capital). Again, the real strategy is to show robustness across each measure.

The observed production network is defined by expenditure shares between firms, $G = [g_{ij}]$, and the firm-region expenditure shares $A = [a_{ij}]$. A directed link exists from firm j to firm i if i buys some positive amount of firm j's output. The intensity of the link is determined by the value of $g_{ij} \in [0, 1]$. In this setting, observed (d_i) and unobserved (δ_i) outdegrees are

$$g_i = \sum_{r \in R} a_{ri} + \sum_{j \in N} g_{ji}; \quad \delta_i = \sum_{r \in R} \lambda_{ri} + \sum_{j \in N} \gamma_{ji} \tag{18}$$

The observed shares and show considerable asymmetry. As we saw in the model in Section 2, the asymmetry of the influence vector and the asymmetry of the observed production network do not necessarily let us infer anything about the underlying economic relationships between firms. We only know that a firm buys a lot of input from another firm, not why.

Finally, I use two measures of geography, shown in Table 2. First, I use the STF to calculate the *ad valorem* rates of shipments between each pair of firms:

$$\tau_{ij} = 1 + \frac{\text{mean shipment cost}_{ij}}{\text{mean shipment value}_{ij}} \tag{19}$$

Table 2: Geography

	Mean	Median	S.D.
Distance	2456	2203	1704
Distance, output weighted	1745		2022
Distance, trade weighted	873		1761
Estimated τ_{ij}	73.06	71.05	44.45
Estimated τ_{ij}, output weighted	53.55		57.83
Estimated τ_{ij}, trade weighted	27.71		48.30
Ad valorem rate	1.03		0.08

Notes: Distance (in km) and *ad valorem* rates (1.0x means a rate of x%) are from the STF, 2004-2012. Estimated τ_{ij} are the predicted trade costs from a gravity regression by region, normalized so the minimum τ_{ij} is equal to 1.

This gives a direct measure of the direct cost of shipping goods at a distance. Lastly, I estimate a gravity model at the Economic Region level and use predicted trade costs as a measure of τ_{ij}. *Ad valorem* rates are much smaller, with average rates of 3% per shipment, while the predicted trade costs are much, much higher, with a mean that's 73 times higher than the minimum trade cost between regions. This is more representative of Canada's geography, but may confuse bilateral weakness in demand between regions with higher trade costs. On the other hand, *ad valorem* rates do not include costs associated with geography that aren't included in the cost of shipping, such as time, the cost of goods damaged in transit, and so on. Despite the low rates, firms engaging in just-in-time inventory management may desire close proximity to customers and suppliers independent of the actual cost of shipping. This suggests the true geography of trade costs in Canada lie somewhere between the two extremes. However, as we will see, the results are consistent across trade cost specifications.

3.2 The importance of higher-order interconnections

Can we simplify the study of the complex firm-firm network to a one-dimensional measure? Acemoglu et al. (2012) cite outdegree as the main measure of network importance; can we focus on that one-dimensional firm measure and leave the complex network alone? Here, I show that one-dimensional measures do not explain much of the firm size distribution, and therefore higher-order interconnections are significant factors in explaining the economy—we cannot rely on one-dimensional firm measures alone.

Suppose the input-output connection is constant across firms, and equal to δ_j/N for firm j, as in Example 3. Then first-order outdegree and productivity alone explain the firm-size distribution,

$$\log v_i = \chi + (\eta - 1)\log z_i + \log \delta_i, \tag{20}$$

If this equation defines the firm-size distribution, estimating this equation with OLS should give an R^2 close to 1, subject to measurement and numerical error. However, the estimated R^2 is only 26.4% (about 5% when including productivity alone, and 21% when including outdegree alone). This leaves 73.6% of the firm-size distribution unexplained, which means the higher-order interconnections matter—it matters which firms you supply, and which firms they supply, and so on, and the complex effects of the network cannot be reduced to one-dimensional firm measures. Note that using a skewed distribution δ_j as demand parameters implies (skewed across suppliers j, constant across customers i within a given supplier) implies skewed distributions of second-order and higher-order outdegrees as well (see Acemoglu et al., 2012). This suggests it is not only the higher order demand connections, but how they interact with productivity as well.[2]

4 Calibration

In this section, I calibrate the model to match features of the data to further explore the relationships between productivity, geography, the unbalanced production network and volatility. Instead of relying on asymptotic results to infer which factor dominates the size distribution (see Appendix 7.3), using the model described in Section 2, I use data on firm productivity z, trade costs T, the observed region demand A, and the observed input share matrix G to solve for the unobserved region demand characteristics Λ and the unobserved technical requirement matrix Γ.

Although final demand didn't add to the explanation of the model and asymptotic theory, it is important empirically. Therefore, to match the data better, I change the

[2] δ_i is calculated as the column sums of Λ plus the column sums of Γ. Using observed outdegree (via A and G) gives similar results. Variation in β also matters quantitatively for the firm size distribution, but this again suggests it matters which firms you supply, and who they supply, and so on, not just that you have a high outdegree.

regional consumer's utility function to a CES combination of each product,

$$u_r(c_{ri}) = \left(\sum_{i \in N} \lambda_{ri}^{\frac{1}{\epsilon}} c_{ri}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{21}$$

Where c_{ri} is region r's consumption of firm i's output. Now the unobserved final demand characteristic λ_{ri} is similar to a γ_{ji} in firm j's production function, and the observed final demand share a_{ri} is similar to the observed expenditure share g_{ji}. In addition, variation in the value added share of output per firm matters for the distribution of output. After adding these features, the goal is to use the model to uncover the unobserved region-firm and firm-firm demand parameters from the data.

4.1 Parameters

There are several sets of parameters that determine the model. Most of the parameters I can select directly from data, a few I need to set, and the rest I use the model (and the given parameters) to solve. The observable set of parameters are: output s_i, the expenditure share matrices A and G, value added shares β_i, productivities z_i, regional income wL_r, and trade costs T. Next, I set the elasticities of substitution η and ϵ at 2. Finally, using the data and model, I solve for the unobserved demand parameters Λ and Γ. For a full description of the data sources, benchmarking, calibration and solutions to the model, see the Appendix.

4.2 Productivity vs. demand

Productivity and demand characteristics are tough to define. Productivity z_i is some technology specific to firm i that tells us how effective that firm is at turning inputs into outputs. However, with CES production technology, firm i is more productive (and is larger) if it uses more inputs (and $\gamma_{ij} = 1$ for all inputs j), even holding z_i constant. In this case, even though the demand characteristics are increasing its size, we'd like to associate that effect with productivity. In other words, if we normalize the demand characteristics γ_{ij} for each i, and associate that effect with productivity instead, we can more accurately describe the relative effects of demand

and productivity.

$$q_i = C_i w^{\beta_i} \underbrace{\left[z_i \left(\sum_j \gamma_{ij}^{1/\eta} \right)^{\frac{(1-\beta_i)\eta}{\eta-1}} \right]}_{\tilde{z}_i} \left(\sum_j \underbrace{\left[\frac{\gamma_{ij}^{1/\eta}}{\sum_k \gamma_{ik}^{1/\eta}} \right]}_{\tilde{\gamma}_{ij}} q_{ij}^{\frac{\eta-1}{\eta}} \right)^{\frac{(1-\beta_i)\eta}{\eta-1}}, \tag{22}$$

and I refer to $\tilde{z}_i$ as augmented productivity, and $\tilde{\Gamma}$ as augmented demand. In the following empirical results, I use these augmented measures instead. The final results using the augmented measures suggest demand accounts for a significant portion of firm size, and using the raw productivity and demand measures only reinforce that result. Using the augmented measures serves to adjust for a producer's demand characteristics that results in higher productivity. It may also adjust for bias in raw productivity measures, since the model uncovers demand parameters that justify the size distribution—if a firm with low raw productivity z_i ends up with large measured demand characteristics, then the raw productivity measure wasn't enough to justify the firm's size, and the demand characteristics provide an augmented productivity measure $\tilde{z}_i$ that is correct and consistent with the model and data.

4.3 Dynamic model

To adapt the static model in Section 2 to include volatility, I use a strategy similar to Acemoglu et al. (2012). In each period, firms receive idiosyncratic demand shocks γ'_{ijt} and λ'_{rit}, as well as productivity shocks z'_{it}. In each period, the equilibrium is equal to the static model with the new parameters $\gamma_{ijt} = \gamma_{ij}\gamma'_{ijt}$, $\lambda_{ijt} = \lambda_{ij}\lambda'_{ijt}$, and $z_{it} = z_i z'_{it}$.

There are several important factors in the dynamic model that help us study the microfoundations of aggregate fluctuations, and the relative contributions of granularity, geography and exogenous production characteristics to aggregate volatility. Similar to the rest of the paper, the difference between the unobserved and observed parameters matters. The data are observed sales growth rates, but we would like to know the unobserved idiosyncratic shocks that gave rise to them. Furthermore, uncorrelated idiosyncratic shocks naturally result in correlated sales growth rates, depending on the linkages between firms and firms, and firms and regions.

Next, demand and productivity shocks may contribute differently to aggregate volatility. In previous work (see, e.g., Acemoglu et al., 2015; Shea, 2002), productivity shocks only propagate downstream, and demand shocks only propagate upstream. However, using a CES function in productivity and demand, both types of shocks can propagate in both directions. For example, a positive productivity shock can propagate upstream because it affects downstream expenditure for the product (positively, if the elasticity of substitution is greater than one).

The distinction between demand and productivity is an important factor in the literature on the firm-size distribution (see Foster et al., 2008, and Section 2 above), so it's reasonable to expect the same pattern in volatility. Demand variation by firm contributes significantly more to the firm-size distribution than does variation in productivity. Similarly, idiosyncratic demand shocks may contribute significantly more to volatility than does idiosyncratic productivity shocks.

The last important note: idiosyncratic shocks may or may not be correlated. First, I attempt to match aggregate volatility by using uncorrelated shocks, but if the simulations can't match the data, I'll re-examine the assumptions, and see how far idiosyncratic shocks can go with reasonable parameter estimates.

4.4 Counterfactuals of the firm size distribution

To examine the effect of the demand network, productivity, geography and the interplay between these factors, I perform several counterfactuals on the data and model. The general idea is to remove variation in one or more of the parameters, solve the model, and (i) compare the true firm density with the counterfactual firm density, and (ii) regress the firm size from the data on the firm size implied by the counterfactual. The density comparison gives an effective visual comparison of the effect of each factor, but lacks sufficient detail to reject any hypotheses. Specifically, the firm density may be similar, but the rank of firm sizes may be scrambled, suggesting the distribution of parameters may give rise to similar aggregate effects, but the underlying parameters do not match the data well. In this case, it is better to compare the individual firm sizes with their corresponding counterfactuals. That is, compare firm i's actual size v_i with its implied size $\hat{v}_{xi}$ after performing some counterfactual x. That gives a better idea of what is truly determining the density by asking what determines the individual units that make up the density.

4.4.1 Demand network

In order to test the importance of the unobserved demand network to the firm size distribution, I eliminate variation in all other factors, recalculate the model and compare the resulting firm sizes with the firm sizes observed in the data. Specifically, I set $z_i = \bar{z}$ and $\tau_{ij} = \bar{\tau}$, and leave β_i, Λ and Γ at their original levels, and then recalculate the set of firm sizes v_i implied by the model.

4.4.2 Geography and trade costs

Similar to the demand network example, I eliminate variation in productivity and demand, and solve the model using only geographical parameters. I set $z_i = \bar{z}$, $\lambda_{ri} = 1/N$ and $\gamma_{ij} = 1/N$, and then solve for the set of firm sizes, v_i, that come from variation in geography alone.

4.4.3 The importance of higher order interconnections, counterfactual version

In Section 3.2, I found that higher order interconnections were significant determinants of the firm size distribution (or more specifically, one dimensional firm attributes like productivity and outdegree cannot explain much of the observed firm size distribution, which leaves the rest to be explained by the interactions betweent the two). Here, I offer similar evidence from a different method. Suppose the counterfactual firm demand networks Λ' and Γ' were such that the outdegrees were the same as the original networks, but the variation across customers for a given supplier was eliminated. Instead of variation across γ_{ij} for a given j, they are all set at a constant value of δ_j/N. The biggest source of change here is the extensive margin—setting the demand network to a constant adds all the firm connections that originally did not exist, turning the network from incredibly sparse to as dense as possible. Then, keeping productivity and value added shares as they are, recalculate the firm sizes.

This strategy keeps the outdegree centrality of a firm constant across the data and counterfactual, but eliminates the true variability in the higher-order interconnections betweens firm demand and productivity. Specifically, a firm that had been a central supplier to some subset of the economy, the same firm is of equal importance to the economy, but spreads the importance of its demand over the entire set of firms in the economy. This eliminates variation in the set of customers each firm has (and the set of customers those customers have), while keeping its 'importance' measures (and ranking thereof) intact. The resulting equilibrium sizes tell us how important the higher-order interconnections are for the economy.

Again, it is important to note that the skewness in the distribution of outdegree shown by Acemoglu et al. (2012) is not enough to explain the firm size distribution, since a skewed distribution of outdegree that results from a γ_{ij} that is constant across i will result in a skewed distribution of higher order outdegrees, but there's no guarantee that the resulting second order outdegree distribution explains firm sizes. In other words, there may be higher order variation in the data that doesn't match the pattern implied by a constant γ_{ij} across i. So, the evidence here will show not only that the higher order interconnections matter for the shape of the firm size distribution, but that the higher order interconnections matter for explaining the individual firm sizes themselves.

4.4.4 Productivity

Here, I ask whether productivity alone can match the firm size distribution. To remove the demand and geography from the model, I eliminate all variation in demand and geography, and calculate the implied firm sizes. To be specific, I set $\gamma_{ij} = 1/N$ for all $i, j \in N$ and $\lambda_{ri} = 1/N$ for all $r \in R$ and $i \in N$, and $\tau_{ij} = \tau_{ri} = 1$ for all $r \in R$ and $i, j \in N$.

5 Results and discussion

There are several main results. The counterfactual firm densities are shown in Figure 4. The herfindahl, ratio of $90^{\text{th}}/10^{\text{th}}$ percentiles, and regression results are in Table 3. First, productivity accounts for very little, between 5-10%, of the existing firm size distribution. Second, the demand network accounts for much more, around 60% of the firm size distribution, and much of that comes from higher-order interconnections. Third, geographic characteristics do not seem to be very helpful in explaining firm sizes. Finally, a reasonable calibration of idiosyncratic shocks can explain approximately one-third of aggregate volatility.

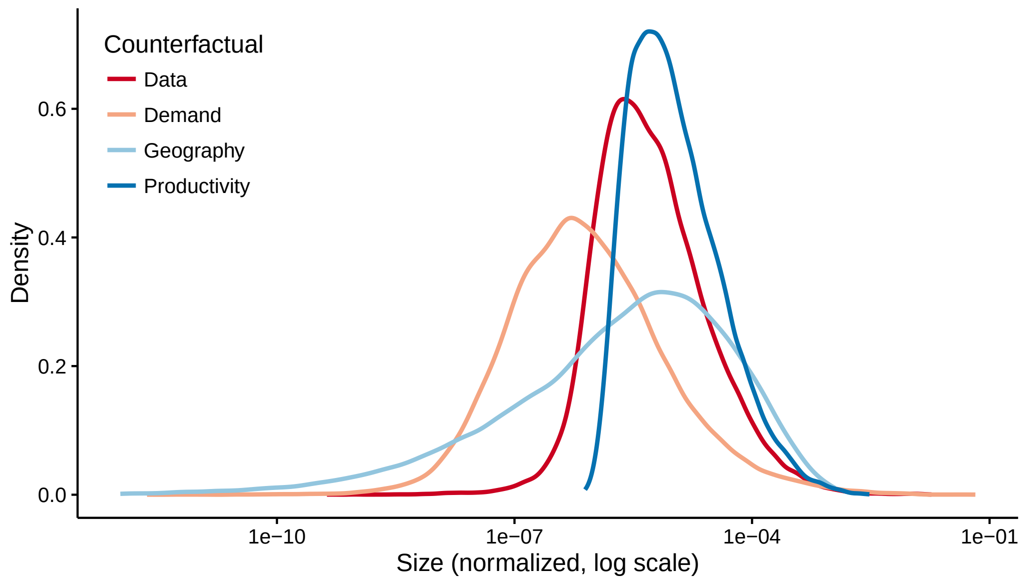

Figure 4: Counterfactual firm densities

Notes: 'x' is the resulting firm size density after removing all variation in the model except value added shares and demand parameters, where x is 'Demand,' 'Geography' or 'Productivity.'

5.1 Counterfactual firm size densities

The firm density defined by productivity alone bears some resemblance to the empirical density but lacks the long right and left tails, suggesting there are demand and geographic characteristics that make some firms very small and very big relative to their productivity levels. This is reflected in the herfindahl, which is about 38% of the data, which would make aggregate volatility that much lower if productivity were the only source of variation in the data (see Section 5.2 for additional volatility results). Furthermore, the R^2 of a regression of $\log v_i$ on $\log v_{xi}$ is 0.092. This shows that productivity, although bearing visual similarities to the empirical distribution, cannot match the individual firm sizes themselves.

This result is robust to different measures of productivity, including different methods of estimating TFP and labour productivity. The fact that productivity doesn't vary enough or in the right directions to explain the firm size distribution accords well with other firm-level studies, including Holmes and Stevens (2014) and Hottman et al. (2016). Both show that demand characteristics explain much more of the firm size distribution than productivity, but in much different settings; Holmes and Stevens (2014) focuses on product differentiation and Hottman et al. (2016) focuses on scanner data for retail goods. Here, I show this same idea applies if you consider the input-output production network as defining demand characteristics.

That brings me to my main result: demand parameters explain much more of the firm size distribution. Visually, the shape of the Demand counterfactual distribution matches the data somewhat well, especially compared to the other counterfactuals. The mean is shifted left, with a slightly higher variance, with a similar right tail but longer left tail. Next, the herfindahl is slightly higher, 0.073 in the counterfactual to 0.046 in the data, implying volatility would increase if demand were the only firm variation in the economy. In addition, the percentile ratio is higher, with 54 in the data and 321 in the counterfactual, which is largely due to the long left tail of the Demand counterfactual distribution (see Figure 4). This significantly reduces the denominator of the 90/10 ratio. In spite of this drawback, the shape of the distribution of the demand counterfactual is very similar to the data. The Demand counterfactual does well explaining the individual firm sizes; the R^2 of a regression of $\log v_i$ on $\log v_{xi}$ gives an R^2 of 0.596, suggesting the demand measures alone explain 60% of the variation in the firm size distribution. In addition, removing higher order interconnections reduces the R^2 of the counterfactual sizes by 35 percentage points. Furthermore, the percentiles ratio increases substantially to an unreasonable number, again because of a very long left tail.

Although other studies of retail goods would consider the Λ and Γ parameters 'firm appeal,' and studies of production networks would call them direct-requirement or input-output parameters, they are conceptually the same. Here, an increase in γ_{ij} could mean an increase in preference by firm i for firm j's product, or a technical requirement for i to use j in production, and both are consistent with demand interpretations in other studies. The relevant distinction here is that these demand parameters are not constant within a firm j—different customers, both firms and final consumers, have different preferences for one firm's output. The interconnections between a firm's customer's preferences, and the preferences of their customers, and so on, have aggregate implications that single-firm measures cannot explain.

Finally, leaving geography as the sole source of variation in the model embodies the true meaning of 'counterfactual.' Relative to the data, the density has a very long and fat left tail, a short right tail, and a much shorter peak. The lack of a right tail results in a very small herfindahl, and the long right tail produces a very higher 90/10 percentile ratio. The visual dissimilarity is matched by the regression results;

Table 3: Counterfactual firm density statistics

	Herfindahl	90/10	Coef.	R^2
Data	0.048	54.0		
Demand (§4.4.1)	0.073	321.92	0.536	0.596
Geography (§4.4.2)	0.016	3598.82	−0.029	0.003
Higher Order (§4.4.3)	0.052	14683.71	0.220	0.245
Productivity (§4.4.4)	0.018	27.72	0.379	0.092

Notes: The coefficient and R^2 are from a regression of $\log v_i$ on $\log v_{xi}$, where v_{xi} is the predicted value of the firm size in counterfactual scenario x, where x can be 'Demand', 'Productivity' or 'Higher Order'. 'Demand', 'Geography,' and 'Productivity' counterfactuals are the resulting firm size density after removing all variation in the model except value added shares and x. 'Higher Order' is the resulting firm size density after *removing* the higher order interconnections between demand and productivity.

the implied firm sizes are actually negatively correlated with the data, making the R^2 comparison somewhat moot. Geography, on its own, does not help with explaining firm sizes at all.

Note that each counterfactual has drawbacks, and cannot explain the firm size distribution alone. Specifically, demand explains a lot of the firm size distribution, but the herfindahl is actually higher after removing variation in geography and productivity. This suggests the factors combine in complex ways, sometimes complementary (e.g., a firm with high demand characteristics is located close to its customers), sometimes not (e.g., a firm with higher than average productivity is in a remote area), to arrive at the final equilibrium.

5.2 Volatility

The contribution of microeconomic shocks to aggregate volatility depend on the skewness of the firm size distribution, and the skewness of the firm size distribution depends on the factors outined previously. Specifically, the contribution of idiosyncratic shocks to aggregate volatility can be calculated with the formula

$$\hat{\sigma}_{GDP} = \sum_i \left(\frac{s_i}{\sum_k \beta_k s_k} \right) \sigma_{zi}, \tag{23}$$

where the term in brackets is a firm-level Domar weight (sales over total value added), see Gabaix (2011) for a discussion of the justification Domar weights and Hulten's

Table 4: Microeconomics shocks and aggregate volatility in the data

Productivity, z	σ_z	$\hat{\sigma}_{GDP}$	Rel. S.D.
TFP (Naïve)	0.17	0.019	0.32
TFP (GNR)	0.27	0.031	0.51

Notes: σ_z is the weighted standard deviation of productivity shocks. I remove industry and region shocks from z in an attempt to approximate idiosyncratic productivity shocks. The sales Herfindahl in the data is $h = 0.048$, the share of value added in aggregate sales is $\beta = 0.41$. The implied volatility is defined as $\hat{\sigma}_{GDP} = \sigma_z h/\beta$. Actual value added volatility is 0.06.

theorem. Using the weighted standard deviation of productivity as a measure of σ_{zi}, and writing β as the share of total value added in total output, this equation can be rewritten

$$\hat{\sigma}_{GDP} = \left(\frac{h}{\beta}\right)\sigma_z, \tag{24}$$

which provides an easy estimate of the contributions of microeconomic shocks to aggregate volatility. Using data on h, β, and σ_z, Table 4 shows the relative contribution of microeconomic shocks to aggregate volatility. These results are consistent with other studies of aggregate volatility.

In addition, the formula gives an easy calculation of aggregate volatility using counterfactual estimates of h and β. The sales herfindahl implied by the productivity distribution alone is very low, 0.018, and the aggregate value added share is higher at 0.70, giving an implied idiosyncratic volatility of 0.004, which lowers aggregate volatility by 25% (assuming the macroeconomic factors remain the same). However, using only variation in demand actually raises the herfindahl to 0.073, raising aggregate volatility by 11% (after accounting for a slight increase in the value added share).

5.3 Robustness

The result that demand is driving the firm size distribution is robust across different specifications. When using *ad valorem* rates instead of estimated trade costs, demand explains an even higher percentage of the firm size distribution at 89.9%, with about 39 p.p. of that due to higher order demand network connections (in the specification using labour productivity). See Table 5 for more. Estimates using different measure of productivity give very similar results.

Table 5: Robustness of firm size counterfactuals to different measures of productivity and geography

A: Geography → *Ad valorem* rates						
	Productivity					
	LP (Pay.)		TFP (Naïve)		TFP (GNR)	
	Coef.	R^2	Coef.	R^2	Coef.	R^2
Demand	0.990	0.899	1.094	0.906	1.015	0.953
Geography	−0.045	0.006	−0.039	0.005	−0.049	0.008
Higher order	0.657	0.504	0.710	0.533	0.662	0.502
Productivity	0.466	0.020	1.619	0.143	0.915	0.022

B: Geography → Estimated trade costs						
	Productivity					
	LP (Pay.)		TFP (Naïve)		TFP (GNR)	
	Coef.	R^2	Coef.	R^2	Coef.	R^2
Demand	0.536	0.596	0.581	0.603	0.549	0.619
Geography	−0.029	0.003	−0.040	0.007	−0.035	0.005
Higher order	0.220	0.245	0.210	0.246	0.216	0.239
Productivity	0.379	0.092	0.475	0.146	0.388	0.089

Notes: Ad valorem rates and estimated trade costs are described in Section 2.3. Productivity measures are described in Section 3.1.1. The coefficient and R^2 are from a regression of $\log v_i$ on $\log v_{xi}$, where v_{xi} is the counterfactual firm size in each case x, where x can be 'Demand,' 'Geography,' 'Higher order,' or 'Productivity.' All coefficients are statistically significant with t-stats of less than 2×10^{-16}, so I omit standard errors from the table.

6 Conclusion

In this paper, I ask whether productivity, geography or network asymmetry provide better microfoundations for the propagation of idiosyncratic shocks. If granularity, a skewed firm size distribution, determines aggregate fluctuations, what determines granularity? Using detailed data on firm-firm trade in Canada, I study a firm-firm production network and its effect on aggregate volatility.

To differentiate between productivity, geography and the unobserved demand network, I use a model in which these factors vary independently and use the production network data to uncover the model parameters. I find three main results: first, the demand network explains approximately 60% of the observed firm size distribution. One dimensional firm demand measures can only explain about 25%, which leaves higher order interconnections between firms to account for 35 p.p. of the firm size distribution. This suggests the complex demand *network*, i.e., your customers and the customers of your customers, is a significant determinant of the firm size.

Second, I find that productivity only explains 10% of the firm size distribution. Productivity does not vary enough to explain the aggregate shape of the distribution and is not correlated enough with firm size to explain much of the individual sizes themselves. Third, despite substantial agglomeration and geographic variation across Canada, trade costs do not seem to explain any of the firm size distribution, although there are important interactions between geography and productivity. This may be due to endogenous firm location decisions, a potential avenue to explore further.

Finally, reasonable levels of idiosyncratic shocks can account for approximately 32% of aggregate volatility. Counterfactual estimates suggest that removing cross-sectional demand and geographic variation in the economy would reduce aggregate volatility by 25%, while removing productivity and geographic variation would increase it by 11%.

The major conclusion to draw from this paper, and something that sets the stage for future work, is that the empirical results confirm the idea that the demand network significantly determines the firm size distribution and aggregate volatility. Futhermore, higher order interconnections between firms explain a large part of the firm size distribution. Firm-firm trade is complex, and studying the implications of the production network for aggregate volatility, trade, transaction costs, vertical integration, and many other subjects, will require much more theoretical and empirical work.

References

Acemoglu, D., U. Akcigit, and W. Kerr (2015). Networks and the Macroeconomy: An Empirical Exploration. In *NBER Macroeconomics Annual 2015*, Volume 30. University of Chicago Press.

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The Network Origins of Aggregate Fluctuations. *Econometrica 80*(5), 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2013). Systemic Risk and Stability in Financial Networks.

Atalay, E., A. Hortaçsu, J. Roberts, and C. Syverson (2011). Network structure of production. *Proceedings of the National Academy of Sciences of the United States of America 108*(13), 5199–5202.

Barrot, J.-N. and J. Sauvagnat (2015). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2015). Production Networks, Geography and Firm Performance.

Carvalho, V., N. Makoto, and Y. U. Saito (2014). Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. *RIETI Discussion Paper Series* (14-E-035).

Carvalho, V. and N. Voigtländer (2014). Input Diffusion and the Evolution of Production Networks.

Di Giovanni, J. and A. Levchenko (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy 120*(6), 1083–1132.

Di Giovanni, J. and A. Levchenko (2013). Firm Entry, Trade, and Welfare in Zipf's World. *Journal of International Economics 89*(2), 283–296.

Di Giovanni, J., A. Levchenko, and I. Méjean (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica 82*(4), 1303–1340.

Di Giovanni, J., A. Levchenko, and R. Rancière (2011). Power Laws in Firm Size and Openness to Trade: Measurement and Implications. *Journal of International Economics 85*(1), 42–52.

Dupor, B. (1999). Aggregation and irrelevance in multi-sector models. *Journal of Monetary Economics 43*(2), 391–409.

Elliott, M., B. Golub, and M. O. Jackson (2014). *Financial Networks and Contagion*, Volume 104.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. *Journal of Political Economy 119*(1).

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review 98*(1), 394–425.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica 79*(3), 733–772.

Gabaix, X. and R. Ibragimov (2011). Rank 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economic Statistics 29*(1), 24–39.

Gandhi, A., S. Navarro, and D. Rivers (2013). On the Identification of Production Functions: How Heterogeneous is Productivity?

Holmes, T. J. and J. J. Stevens (2014). An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade. *Journal of Political Economy 122*(2).

Horvath, M. (1998). Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks. *Review of Economic Dynamics 1*(4), 781–808.

Horvath, M. (2000). Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics 45*(1), 69–106.

Hottman, C., S. J. Redding, and D. E. Weinstein (2016). Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics*.

Hulten, C. (1978). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies 45*(3), 511–518.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Jessen, A. H. and T. Mikosch (2006). Regularly Varying Functions. *Publications de L'Institut Mathematique 80*(94), 171–192.

Jones, C. (2011). Intermediate Goods and Weak Links in the Theory of Economic Development. *American Economic Journal: Macroeconomics 3*(2), 1–28.

Lim, K. (2016). Firm-to-firm Trade in Sticky Production Networks. (January), 1–74.

Long, J. B. and C. I. Plosser (1983). Real Business Cycles. *Journal of Political Economy 91*(1), 39—-69.

Luttmer, E. G. J. (2007). Selection, Growth, and the Size Distribution of Firms. *Quarterly Journal of Economics* (August), 1103–1144.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695—-1725.

Oberfield, E. (2011). Business networks, production chains, and productivity: A theory of input-output architecture.

Shea, J. (2002). Complementarities and comovements. *Journal of Money, Credit, and Banking 34*(2), 412–433.

7 Appendix: Theory

See important model notation in Table 6.

Table 6: Table of Notation

R	$\triangleq$	Set of regions. Abusing notation, R is also the number of regions.
N	$\triangleq$	Set of plants. Abusing notation, N is also the number of plants.
G	$\triangleq$	$N \times N$ matrix of observed plant input shares. An element g_{ij} is the share of plant j's input in plant i's sales.
Γ	$\triangleq$	$N \times N$ matrix of exogenous plant input demand characteristics. An element γ_{ij} enters plant i's demand for plant j's output.
A	$\triangleq$	$R \times N$ matrix of observed region-plant demand shares. An element a_{ri} is the share of region r's total expenditure on plant i's output.
Λ	$\triangleq$	$R \times N$ matrix of exogenous region input demand characteristics. An element λ_{ri} enters region r's demand for plant i's output.
T	$\triangleq$	$(R+N) \times (R+N)$ matrix of trade costs. An element τ_{ri} is the cost of trade between region r and i, and an element τ_{ij} is the cost of trade between plants i and j.
z_i	$\triangleq$	Productivity of plant i.
ϵ	$\triangleq$	Final demand elasticity of substitution.
η	$\triangleq$	Intermediate elasticity of substitution.
β_i	$\triangleq$	Share of value-added in plant i's production.

7.1 Full model

7.1.1 Consumers

There are R regions, with a representative consumer in each with utility function $u_r(c_r)$,

$$u_r(c_r) = \left(\sum_{i \in N} \lambda_{ri}^{\frac{1}{\epsilon}} c_{ri}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{25}$$

Labour is inelastically supplied given the stock of labour in region r, L_r. Consumer r's problem is

$$\max_{c_r} u_r(c_r) \text{ s.t. } \sum_{i \in N} p_{ri} c_{ri} \leq w_r L_r \tag{26}$$

Consumer r must pay a trade cost τ_{ri} to buy from plant i, so that

$$p_{ri} = \tau_{ri} p_i \tag{27}$$

The solution gives r's price index

$$p_r = \left(\sum_{i \in N} \lambda_{ri} (\tau_{ri} p_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \tag{28}$$

7.1.2 Producers

There are N producers. Producer i's production function is

$$f_i(l_i, q_{i1}, \ldots, q_{iN}) = z_i l_i^{\beta_i} \left(\sum_{j \in N} \gamma_{ij}^{\frac{1}{\eta}} q_{ij}^{\frac{\eta-1}{\eta}} \right)^{\frac{(1-\beta_i)\eta}{\eta-1}} \tag{29}$$

Producer i's problem is to minimize cost

$$\min_{(l_i, q_{i1}, \ldots, q_{iN})} \sum_{i \in N} p_{ij} q_{ij} \text{ s.t. } f_i \geq \bar{q}_i \tag{30}$$

Producer i's input cost for one unit of the intermediate input is

$$p_{mi} = \left(\sum_{j \in N} \gamma_{ij} (\tau_{ij} p_j)^{1-\eta} \right)^{\frac{1}{1-\eta}} \tag{31}$$

Given perfect competition, plant i's price is (including wages),

$$p_i = \beta_i^{-\beta_i} (1-\beta_i)^{\beta_i - 1} z_i^{-1} p_{mi}^{1-\beta_i} \tag{32}$$

7.1.3 Market clearing

Labour is free to migrate between regions. Total labour in the economy is

$$\sum_{r \in R} L_r = L \tag{33}$$

Now, each plant i is in a region r, and the total value added produced by those plants in r add up to total income in that region,

$$\sum_{i \in r} \beta_i s_i = w L_r \tag{34}$$

For goods, producer i supplies the other producers $j \in N$, and each region $r \in R$, giving market clearing

$$\sum_{r \in R} c^s_{ri} + \sum_{j \in N} q^s_{ji} = q^s_i, \text{ for } i \in N \tag{35}$$

Iceberg trade costs mean producer i ships $c^s_{ri} = \tau_{ri} c_{ri}$ to region r and $q^s_{ji} = \tau_{ji} q_{ji}$. Replacing those terms and multiplying all terms by p_i,

$$\sum_{r \in R} p_i \tau_{ri} c_{ri} + \sum_{j \in N} p_i \tau_{ji} q_{ji} = p_i q^s_i, \text{ for } i \in N \tag{36}$$

7.1.4 Equilibrium

Equilibrium in the economy means two sets of prices $\{p_r : r \in R\}$, $\{p_i : i \in N\}$, wage w normalized to 1, and labour stocks by region $\{L_r : r \in R\}$, that solve the consumer's and producer's problems for each region and producer, and the labour and goods markets clear.

7.1.5 Solving the model given data

Given data on T, G, A, w, β, solve for Λ and Γ. Must also solve for p_r and p_i, and normalize $w = 1$. Make assumptions about the elasticities η and ϵ, then solve. Have price equations:

$$p_r = \left(\sum_{i \in N} \lambda_{ri} (\tau_{ri} p_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}, \text{ for } r \in R \tag{37}$$

$$p_{mi} = \left(\sum_{j \in N} \gamma_{ij} (\tau_{ij} p_j)^{1-\eta} \right)^{\frac{1}{1-\eta}}, \text{ for } i \in N \tag{38}$$

$$p_i = z_i^{-1} \beta_i^{-\beta_i} (1-\beta_i)^{\beta_i - 1} w^{\beta_i} p_{mi}^{1-\beta_i}, \text{ for } i \in N \tag{39}$$

And share equations:

$$a_{ri} = \lambda_{ri}\tau_{ri}^{1-\epsilon}\left(\frac{p_i}{p_r}\right)^{1-\epsilon}, \text{ for } r \in R, i \in N \tag{40}$$

$$g_{ij} = (1-\beta_i)\gamma_{ij}\tau_{ij}^{1-\eta}\left(\frac{p_j}{p_{mi}}\right)^{1-\eta}, \text{ for } i \in N, j \in N \tag{41}$$

$$\lambda_{ri} = a_{ri}\tau_{ri}^{\epsilon-1}\left(\frac{p_i}{p_r}\right)^{\epsilon-1}, \text{ for } r \in R, i \in N \tag{42}$$

$$\gamma_{ij} = (1-\beta_i)^{-1}g_{ij}\tau_{ij}^{\eta-1}\left(\frac{p_j}{p_{mi}}\right)^{\eta-1}, \text{ for } i \in N, j \in N \tag{43}$$

And region income equations,

$$\beta_i s_i = w l_i \tag{44}$$

$$\sum_{i \in R}\beta_i s_i = w L_r \tag{45}$$

$$\sum_{r \in R} L_r = L \tag{46}$$

And finally, sizes:

$$wA'\vec{L} + G's = s, \text{ or} \tag{47}$$

$$s = w(I - G')^{-1}A'\vec{L} \tag{48}$$

How many unknowns? $p_r \to R$, $p_i, p_{mi} \to 2N$, $\Lambda \to RN$, $\Gamma \to N^2$, $s \to N$, $L_r \to R$, w. So $R + 2N + RN + N^2 + N + R + 1$. How many equations? $R + 2N + RN + N^2 + R + 1 + N$.

7.1.6 Solving the model given parameters

Do the opposite. Same equations, start with z, β, T (data), η (maximization), Γ, Λ, and so on. Solve the same equations for s, all p, A, G.

7.1.7 Solve for Λ, Γ

Given data on T, G, A, w, β, solve for Λ and Γ. Must also solve for p_r and p_i, and normalize $w = 1$. Make assumptions about the elasticities η, ϵ, then solve. We have price equations:

$$p_r = \left(\sum_{i \in N} \lambda_{ri} (\tau_{ri} p_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}, \text{ for } r \in R \tag{49}$$

$$p_{mi} = \left(\sum_{j \in N} \gamma_{ij} (\tau_{ij} p_j)^{1-\eta} \right)^{\frac{1-\beta_i}{1-\eta}}, \text{ for } i \in N \tag{50}$$

$$p_i = \beta_i^{-\beta_i} (1 - \beta_i)^{\beta_i - 1} w^{\beta_i} z_i^{-1} p_{mi}, \text{ for } i \in N \tag{51}$$

$$\lambda_{ri} = a_{ri} \tau_{ri}^{\epsilon-1} \left(\frac{p_i}{p_r} \right)^{\epsilon-1}, \text{ for } r \in R, i \in N \tag{52}$$

$$\gamma_{ij} = (1 - \beta_i)^{-1} g_{ij} \tau_{ij}^{\eta-1} \left(\frac{p_j}{p_{mi}} \right)^{\eta-1}, \text{ for } i \in N, j \in N \tag{53}$$

7.2 Derivation of influence vector

Using the definition of observed expenditure shares,

$$g_{ji} = \frac{\tau_{ji} p_i q_{ji}}{p_j q_j} \tag{54}$$

Rewrite the system of market clearing equations

$$\sum_{r \in R} \tau_{ri} p_i c_{ri} + \sum_{j \in N} \tau_{ji} p_i q_{ji} = p_i q_i, \text{ for } i \in N \tag{55}$$

Data: G, A, T^r, T^i, z, β

Result: $\Lambda, \Gamma, p^i, p^r$ that solve Equations 49–53, with $w = 1$.

begin

$p_0^r \leftarrow \mathbf{0}, p_1^r \leftarrow I$;

$\Lambda_0 \leftarrow [0], \Lambda_1 \leftarrow [1]$;

$\Gamma_0 \leftarrow [0], \Gamma_1 \leftarrow [1]$;

while $(||p_1^r - p_0^r||_f + ||p_1^i - p_0^i||_f + ||\Lambda_1 - \Lambda_0||_f + ||\Gamma_1 - \Gamma_0||_f > \text{TOL})$ **do**

$p_0^r \leftarrow p_1^r$;

$\Lambda_0 \leftarrow \Lambda_1$;

$\Gamma_0 \leftarrow \Gamma_1$;

$p_1^r \leftarrow ((\Lambda_0 \odot (T^r \cdot p_0^i)^{1-\epsilon}) \cdot \mathbf{1})^{1/(1-\epsilon)}$;

$p_{mi} \leftarrow ((\Gamma_0 \odot (T^i \cdot p_0^i)^{1-\eta}) \cdot \mathbf{1})^{(1-\beta)/(1-\eta)}$;

$p_1^i \leftarrow p_{mi} \odot z^{-1}$;

`//` p_1^r `needs to be pre-multiplying the rest:`

$\Lambda_1 \leftarrow [p_1^r]^{1-\epsilon} \cdot A \odot (T^r \cdot p_1^i)^{\epsilon-1}$;

`//` β `is a vector, and` p_{mi} `needs to be pre-multiplying the rest:`

$\Gamma_1 \leftarrow (p_{mi}^{(1-\eta)} \odot (1-\beta)) \cdot G \odot (T^i \cdot p_1^i)^{\eta-1}$;

`// Element-wise inverse to normalize sum of demand shares to 1 by row:`

$\Lambda_1 \leftarrow \Lambda_1 \odot (\Lambda_1 \cdot \mathbf{1})^{-1}$;

$\Gamma_1 \leftarrow \Gamma_1 \odot (\Gamma_1 \cdot \mathbf{1})^{-1}$;

end

end

Algorithm 1: Solve for demand parameters, given observed demand, productivity and value-added share data.

as

$$\sum_{r\in R} \tau_{ri} p_i c_{ri} + \sum_{j\in N} g_{ji} p_j q_j = p_i q_i, \text{ for } i \in N \tag{56}$$

Then replace $\tau_{ri} p_i c_{ri} = a_{ri} p_r c_r = a_{ri} w L_r$ and define total sales as $s_i = p_i q_i$,

$$\sum_{r\in R} a_{ri} w L_r + \sum_{j\in N} g_{ji} s_j = s_i, \text{ for } i \in N \tag{57}$$

Rewrite in vector form, using $L = (L_1, \ldots, L_R)'$, write $a_{\cdot i}$ as the i-th column of A and $g_{\cdot i}$ as the i-th column of G,

$$w a'_{\cdot i} L + g'_{\cdot i} s = s_i, \text{ for } i \in N \tag{58}$$

Now stack those N equations on top of each other, which stacks the vectors $g'_{\cdot i}$ (now the *row* vectors of G'), which gives

$$wA'L + G's = s \tag{59}$$

Rearrange and factor out s,

$$s - G's = wA'L \tag{60}$$

$$(I - G')s = wA'L \tag{61}$$

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

$$s = w(I - G')^{-1} A'L \tag{62}$$

To get the influence vector, use $w\mathbf{1}'L = \beta \sum_{i\in N} s_i$ and $v_i = s_i / \left(\sum_{j\in N} s_j\right)$, and finally normalize wages to 1 ($w = 1$) and take the transpose of both sides:

$$v' = \left(\frac{\beta}{\mathbf{1}'L}\right) L'A(I - G)^{-1} \tag{63}$$

If value-added varies across plants, the relevant equation is

$$A'\overrightarrow{(\beta' v)_r} + G'v = v \tag{64}$$

Or,

$$A'\overrightarrow{(\beta' v)_r} = (I - G')^{-1} v \tag{65}$$

7.3 Asymptotic Theory

Asymptotic results are key to the arguments for and against the microfoundations of aggregate shocks.[3] 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 asymmetric production 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 ζ_X and ζ_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 Gabaix, 2009; Jessen and Mikosch, 2006). Using Remark 3, we are interested in explaining the tail parameter of the size distribution, β_v, given the tail parameters of the distributions of observed outdegree (ζ_d) and productivity (ζ_z).

Therefore, if the asymptotic results hold for this economy, network asymmetry cannot be the fundamental cause of the skewed firm 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 firm-level exogenous production network characteristics must determine the tail of the distribution of the observed firm-level production network characteristics. Second, the tail of the distribution of the observed production network determines the tail of the firm size distribution. If either of these arguments fail, it is

[3] In Appendix 7.4, 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.

unlikely the underlying demand characteristics are the cause of the skewed firm 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 firm size distribution's parameter of 1.04, which is consistent with a Zipf's law distribution of firm size. Therefore $\zeta_z < \zeta_d$ implies the degree distribution is dominated by some other firm characteristic, and thus does not determine firm 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 Luttmer, 2007; Gabaix, 2009), while Acemoglu et al. (2012) 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 production network data are necessarily the observed shares, and so depend on both the underlying demand characteristics and other firm characteristics, especially productivity.

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 7.1 *Suppose the distributions of outdegree and productivity both follow power laws with parameters* ζ_d *and* ζ_z*,*

$$P(d > x) = C_d x^{-\zeta_d} L_d(x), \tag{66}$$

$$P(z > x) = C_z x^{-\zeta_z} L_z(x) \tag{67}$$

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

$\min\{\zeta_d, \zeta_z/(\eta-1)\}$,

$$P(v > x) = C_v x^{-\min\left\{\zeta_d, \frac{\zeta_z}{(\eta-1)}\right\}} L_v(x) \tag{68}$$

Proof 1 *One element of the influence vector, v_i, is*

$$v_i = \frac{\beta}{N} + (1-\beta)\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right) \tag{69}$$

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

$$v_i \to \chi d_i z_i^{\eta-1} \tag{70}$$

$$F_v(x) = F_v\left(\chi d_i z_i^{\eta-1}\right) \tag{71}$$

$$P(v > x) \to P(\chi d z^{\eta-1} > x) \tag{72}$$

$$= P(d z^{\eta-1} > \chi^{-1} x) \tag{73}$$

$$P(v > x) = P(d z^{\eta-1} > \chi^{-1} x) \tag{74}$$

$$= \int_{\underline{d}}^{\infty} P\left(z > \left[\frac{x}{\chi d}\right]^{1/(\eta-1)}\right) \mathrm{d}F_d(d) \tag{75}$$

$$= \int_{\underline{d}}^{\infty} C_z \left[\frac{x}{\chi d}\right]^{-\zeta_z/(\eta-1)} \mathrm{d}F_d(d) \tag{76}$$

$$= \chi^{\zeta_z/(\eta-1)} C_z x^{-\zeta_z/(\eta-1)} \int_{\underline{d}}^{\infty} d^{\zeta_z/(\eta-1)} \mathrm{d}F_d(d) \tag{77}$$

For the integral to exist, we need $\zeta_z/(\eta-1) < \zeta_d$. If so, it is a constant (independent of x), so combine the other constants into $C_v = \chi^{\zeta_z/(\eta-1)} C_z \int_{\underline{d}}^{\infty} d^{\zeta_z/(\eta-1)} \mathrm{d}F_d(d)$, and write

$$P(v > x) = C_v x^{-\zeta_z/(\eta-1)} \tag{78}$$

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

$$P(v > x) = C_v x^{-\min\{\zeta_d, \zeta_z/(\eta-1)\}} \tag{79}$$

Or,

$$\log P(v > x) = \log C_v - \min\{\zeta_d, \zeta_z/(\eta-1)\} \log x \tag{80}$$

The distribution of productivity has a tail parameter of approximately 1.98, so for a suitable choice of η, it is easy to match the empirical tail parameter of the firm size distribution. In particular, if $\eta \approx 2.89$, 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 8.3 for an extension of the model with monopolistic competition and firm 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 (Di Giovanni et al., 2011; Di Giovanni and Levchenko, 2012, 2013). 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 Luttmer (2007) or Gabaix (2009).

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.

7.4 Aggregate volatility depends on the product of the distributions of outdegree and productivity

Aggregate volatility scales according to $||v||_2$, according to Hulten's Theorem (Hulten, 1978) and Theorem 1 of Acemoglu et al. (2012). 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

$$v_i = \frac{\beta}{N} + (1-\beta)\left(\frac{d_i z_i^{\eta-1}}{\sum_{k \in N} d_k z_k^{\eta-1}}\right) \tag{81}$$

Then the Euclidean norm of v can be written

$$||v||_2 = \sqrt{\sum_{i\in N}\left[\frac{\beta^2}{N^2} + (1-\beta)^2\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)^2 + 2(1-\beta)\left(\frac{\beta}{N}\right)\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)\right]} \tag{82}$$

$$||v||_2 = \sqrt{\frac{\beta^2}{N} + (1-\beta)^2\sum_{i\in N}\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)^2 + 2(1-\beta)\left(\frac{\beta}{N}\right)\sum_{i\in N}\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)} \tag{83}$$

Rewrite slightly,

$$||v||_2^2 = \frac{\beta^2}{N} + (1-\beta)^2\sum_{i\in N}\left(\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)^2 + 2(1-\beta)\left(\frac{\beta}{N}\right) \tag{84}$$

$$||v||_2^2 = \frac{\beta^2}{N} + 2(1-\beta)\left(\frac{\beta}{N}\right) + (1-\beta)^2 h_g^2 \tag{85}$$

$$||v||_2^2 = \frac{\beta(2-\beta)}{N} + (1-\beta)^2 h_g^2 \tag{86}$$

$$||v||_2^2 \geq (1-\beta)^2 h_g^2 \tag{87}$$

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

$$h_g^2 \geq \frac{1}{N}\left(\sum_{i\in N}\frac{d_i z_i^{\eta-1}}{\sum_{k\in N} d_k z_k^{\eta-1}}\right)^2 = \frac{1}{N} \tag{88}$$

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

$$||v||_2^2/h_g^2 = \frac{\beta(2-\beta)}{N h_g^2} + (1-\beta)^2 \tag{89}$$

Meaning

$$\limsup_{N\to\infty}\frac{||v||_2^2}{h_g^2} = \limsup_{N\to\infty}\left[\frac{\beta(2-\beta)}{N h_g^2} + (1-\beta)^2\right] \tag{90}$$

Using the result that $(Nh_g^2)^{-1}$ is bounded above by 1,

$$\limsup_{N\to\infty} \frac{||v||_2^2}{h_g^2} \leq \limsup_{N\to\infty} \left[\beta(2-\beta) + (1-\beta)^2\right] \tag{91}$$

$$\limsup_{N\to\infty} \frac{||v||_2^2}{h_g^2} \leq \beta(2-\beta) + (1-\beta)^2 < \infty \tag{92}$$

So $||v||_2^2 = \mathcal{O}\left(h_g^2\right)$, which combined with the Big-Ω result gives

$$||v||_2 = \Theta\left(h_g\right) \tag{93}$$

8 Appendix: Data and Empirics

8.1 Data sources

Additional descriptions of available data available at CDER: `http://www.statcan.gc.ca/eng/cder/data`.

8.1.1 Annual Survey of Manufacturing (ASM)

Also called the Annual Survey of Manufacturing and Logging (ASML). See `http://www.statcan.gc.ca/eng/survey/business/2103`, and an example survey at `http://www23.statcan.gc.ca/imdb-bmdi/instrument/2103_Q31_V3-eng.pdf`.

(1) Description: long panel of manufacturing establishments, covers most of Canada, x% of value added, etc. NAICS or SIC industry classifications, depending on the sample. (2) Gives province (or exports) of destination of shipments. (3) Gives detailed commodity inputs and outputs. (4) Postal code of establishment.

Years available: 1961-2012.

8.1.2 Surface Transportation File (STF)

Based on the Trucking Commodity Origin and Destination File and Railway Universe File. Transaction-level trade database with postal code geographical detail. Used to identify input shipments between establishments, and final demand shipments from establishments to regions. Includes information on carrier, mode, commodity classification (SCTG), value, tonnage, distance, and revenue to the carrier. Years available: 2004-2012.

8.1.3 Inter-provincial Trade Flows (IPTF)

CANSIM Tables 386-0001, 386-0002, 386-0003, 386-0004, `http://www5.statcan.gc.ca/cansim/a04`. I use the detailed-confidential versions of these tables in the paper. A province × province × commodity dataset of trade, including international imports, exports and re-exports. Years available: 2002-2012.

8.1.4 Input-Output Tables / Supply-Use Tables (IO)

CANSIM Tables 381-0033, 381-0034, 381-0035, `http://www5.statcan.gc.ca/cansim/a04`. I use the detailed-confidential versions of these tables in the paper. An province × industry × commodity dataset. Industry classification is IOIC, commodity classification is IOCC. Years available: 2002-2012.

8.1.5 Import-Export Registry (IER)

Records enterprise-product level imports and exports. I use this to impute the import share of each firm in order to generate an 'international' region. This process is in progress; for technical reasons, more recent data are not yet available, but I will update the paper with these data when I have them. For more information, see `http://www.statcan.gc.ca/eng/cder/data#a2`.

8.2 Predicting establishment links

8.2.1 National square IO table

Aggregate non-manufacturing into one industry. Trim possible industry pairs to those with direct-requirements of at least 0.01 (or maybe by absolute value—if a small industry supplies a big industry, the entire thing could be removed at 0.01 level).

8.2.2 Possible establishment-pair list

These are the establishments that could possibly be connected. Create pairwise establishment list (with more than 50,000 establishments, this is $50{,}000^2$ pairs), then drop ones whose industries aren't in the national square IO table.

8.2.3 Explanatory variables

The probability a directed edge between establishments exists depends positively on the variables in Table 7.

Table 7: Explanatory variables

`ind_io`$_{ij}$	The direct-requirement coefficient of the industry level IO table, IOT.
`asm_prov`$_{ij}$	In the ASM, does establishment j report shipments to establishment i's province?
`asm_io`$_{ij}$	The direct-requirement coefficient of the establishment level IO table, defined by the ASM commodity file.
`stf_pc`$_{ij}$	In the STF, are there shipments from establishment j's postal code to establishment i's postal code, in commodities that establishment j produces and establishment i uses?
`stf_oer_dpc`$_{ij}$	In the STF, are there shipments from establishment j's Economic Region to establishment i's postal code, in commodities that establishment j produces and establishment i uses?
`stf_opc_der`$_{ij}$	In the STF, are there shipments from establishment j's postal code to establishment i's Economic Region, in commodities that establishment j produces and establishment i uses?
`iptf_prov`$_{ij}$	In the IPTF, if establishment j produces a commodity and establishment i uses that commodity, is that commodity exported from establishment j's province to establishment i's province?

Notes: ASM: Annual Survey of Manufacturing, STF: Surface Transporation File, IPTF: Inter-provincial Trade Flow file. IOT: detailed-confidential IO Table (or Supply-Use Table). See Section 8.1.

8.3 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. To test this empirically, I use a similar decomposition to Di Giovanni et al. (2014).

First, write sales of plant i at year t as s_{it}. 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

Table 8: Intensive vs. Extensive Margin Volatility

	S.D.	Rel. S.D.
Aggregate Volatility, $\tilde{\sigma}_A$	0.065	1.00
Intensive Volatility, σ_A	0.066	1.02
Extensive Volatility, σ_ν	0.009	0.14

Notes: Aggregate volatility is the standard deviation of total manufacturing output. Intensive volatility is the standard deviation of total manufacturing output from firms that are alive in periods t and $t-1$. Extensive volatility is the standard deviation of total manufacturing output from firms that entered or exited in period t.

the log-difference aggregate growth rate of sales is

$$\tilde{g}_{At} \equiv \ln\left(\sum_{i\in I_t} x_{it}\right) - \ln\left(\sum_{i\in I_{t-1}} x_{it-1}\right) \tag{94}$$

$$= \ln\left(\frac{\sum_{i\in I_{t/t-1}} x_{it}}{\sum_{i\in I_{t/t-1}} x_{it-1}}\right) - \left[\ln\left(\frac{\sum_{i\in I_{t/t-1}} x_{it}}{\sum_{i\in I_t} x_{it}}\right) - \ln\left(\frac{\sum_{i\in I_{t/t-1}} x_{it-1}}{\sum_{i\in I_{t-1}} x_{it-1}}\right)\right] \tag{95}$$

$$= g_{At} - \ln\left(\frac{\nu_{t,t}}{\nu_{t,t-1}}\right) \tag{96}$$

where g_{At} is the intensive margin of growth and the other term is the extensive margin of growth. Now aggregate volatility is

$$\tilde{\sigma}_A^2 = \sigma_A^2 + \sigma_\nu^2 - 2\text{Cov}(g_{At}, g_\nu) \tag{97}$$

Calculating each of these in the data, we see that the extensive margin matters little (consistent with the results in Di Giovanni et al. (2014). Although large establishments do exit, it is more common for one to have large losses in one year, have a low value of output, and then exit the following year. This puts the volatility on the intensive margin, not extensive.

8.4 Simulated Method of Moments (SMM)

In a future version of the paper, I will apply SMM to estimate the elasticities and idiosyncratic shocks. This is a short description of the methodology.

Suppose there are only productivity shocks, $\log z_i' \sim \mathcal{N}(0, \sigma_z)$. What should σ_z

be in order to match aggregate volatility? To estimate this, I use simulated method of moments (SMM). The strategy is to (1) pick a σ_z, (2) draw productivity shocks z_i', (3) solve the model for plant growth rates, (4) calculate covariances and aggregate volatility. Lastly, (5) find the σ_z that best matches aggregate volatility in the data.

The estimated $\hat{\sigma}_z$ gives an idea of the magnitude of idiosyncratic productivity shocks that match aggregate volatility in the data. Of course, this doesn't mean that productivity is the source of aggregate fluctuations. These productivity shocks are transmitted through the economy via input-output links. For a minute, I leave for later the question of how the unobserved demand and productivity *levels* matter for the transmission of idiosyncratic shocks (of any kind).

8.4.1 Drawing random shocks

Before we start, draw $N \times T$ random variables x from a uniform distribution on [0,1]. After choosing σ_z, the random shocks are $\log z_{it}' = F^{-1}(x_{it})$, and $F = \Phi(0, \sigma_z)$. Quoting McFadden (1989), 'a simulator must avoid "chatter" as $[\sigma_z]$ varies; this will generally require that the Monte Carlo random numbers used to construct $[f(\sigma_z)]$ *not* be redrawn when $[\sigma_z]$ is changed.'

8.4.2 Rank condition in SMM

Suppose I wanted to separately identify σ_z and regional and plant demand shocks, λ_i' and γ_i'. If productivity and demand shocks affect the simulated moments in the same way, the algorithm cannot tell the difference between a change in the productivity shock parameter or a change in one of the demand parameters. In other words, the parameters must have different effects on the simulated moments. The formal statement of this is

$$\text{rank}\left[E\left(\frac{\partial \hat{\sigma}^2(\vec{\sigma})}{\partial \vec{\sigma}}\right)\right] = q \tag{98}$$

where $\vec{\sigma}$ is the length-q vector of parameters.

8.4.3 Program

Aggregate volatility in the data is σ, and simulated aggregate volatility is $\hat{\sigma}(\sigma_z)$,

$$\min_{\sigma_z} ||\sigma^2 - \hat{\sigma}^2(\sigma_z)||^2 \tag{99}$$

The minimization program is solved with simulated annealing using the R package `GenSA`. With more moments and parameters, the estimation program is very similar, with an added variance-covariance matrix of the moments weighting the squared errors.