Cascades and Fluctuations in an Economy with an Endogenous Production Network*

Mathieu Taschereau-Dumouchel†

Cornell University

January 2018

Abstract

This paper proposes a simple theory of production in which the network of input-output linkages is endogenously determined by the decision of the firms to operate or not. Since producers benefit from having multiple suppliers, the economy features complementarities between the operating decisions of nearby firms. As a result, tightly connected clusters of producers emerge around productive firms. In addition, after a firm is hit by a severe shock, a cascade of shutdowns might spread from neighbor to neighbor as the network reorganizes itself. While well-connected firms are better able to withstand shocks, they trigger larger cascades upon shut down, a prediction confirmed by U.S. data. The theory also predicts how the shape of the network interacts with the business cycle. As in the data, periods of low economic activity feature less clustering among firms, and are associated with thinner tails for the degree distributions. Finally, allowing the network to reorganize itself in response to idiosyncratic shocks leads to substantially smaller variations in aggregate output, suggesting that endogenous changes in the shape of the production network have a significant impact on the aggregation of microeconomic shocks into macroeconomic fluctuations.

JEL Classifications: E23, D85, C67

**First draft:* February 2017. I thank Enghin Atalay, Jeremy Greenwood, Ben Lester, Ezra Oberfield, Nicolas Petrosky-Nadeau, Carlos Ramírez, Gill Segal, Ali Shourideh and seminar participants at Wharton, the Philadelphia Fed, Drexel, FGV São Paulo, the Cornell/Penn State workshop, UT Austin, SED Edinburgh, SAET Faro, the ITAM-PIER Conference, the Tepper-LAEF Conference, the CIRANO – Sam M. Walton College of Business Workshop on Networks, LSE, Cornell, the annual meeting of the CMSG and Rutgers for helpful comments and suggestions.

†Economics Department, Cornell University, Uris Hall, Ithaca, NY 14850 (email: mathtd@cornell.edu).

1 Introduction

Production in modern economies involves a complex network of specialized producers, each using inputs from suppliers and providing their own output to downstream production units. In such an environment, the way in which shocks to individual producers aggregate to affect macroeconomic variables depends on the shape of the production network.[1] But the network itself is also constantly changing in response to these shocks. Cascades of firm shutdowns are a salient example of this process: if a severe shock pushes a producer to shut down, its neighbors, now missing a useful input or a valuable customer, might also decide to shut down, thereby pushing their own neighbors to do the same, and so on.[2] A single shock can thus trigger a cascade of shutdowns that spreads through the economy, changing the shape of the network along the way, and thus influencing how the economy handles further shocks. To properly understand the origin of aggregate fluctuations, we therefore need a joint theory of production and endogenous network formation.

This paper proposes such a theory. In the model, a large number of firms produce differentiated goods using labor and a set of inputs from other producers. Since production requires the payment of a fixed cost, firms operate or not as a function of economic conditions. When a firm operates, it makes an additional input available to all of its customers, thereby creating new input-output relationships. Together, the operating decisions of the firms therefore determine the structure of the production network.

Modeling the formation of the network in this way is motivated by the data. According to Factset Revere, a large dataset that covers firm-level input-output linkages in the U.S., about 70% of all link destructions occur when one of the two firms forming the relationship stops producing. It is therefore important to have a model that takes these operating decisions into account to properly understand the evolution of the production network.

At the heart of the model's mechanisms is a simple consequence of the firms' production function: it is beneficial to produce with a greater variety of inputs. As a result, firms with multiple neighbors, either upstream or downstream, are particularly valuable, and therefore more likely to operate in the efficient allocation. Extrapolating to multiple firms, this mechanism pushes for the creation of a network that features tightly connected clusters of producers. By organizing production in this way, firms take advantage of the many suppliers available in the cluster to become more efficient and, in turn, make all their downstream neighbors benefit from this higher efficiency.

[1] See, for instance, Acemoglu et al. (2012). Even when Hulten's (1978) theorem holds, the shape of the network affects the sales share of a producer and therefore how the shocks it receives influence aggregate output. See Baqaee and Farhi (2017a) for environments in which Hulten's theorem does not apply.

[2] Carvalho et al. (2014) find that firms affected by the Great East Japan Earthquake of 2011 that stopped production because of the catastrophe had a significant negative impact on their customers and suppliers that were outside of the affected zone. Another example of these cascades concerns the 2008 bailout of U.S. car manufacturers. At that time, the CEO of Ford advocated for the bailout of its competitors, General Motors and Chrysler. The rationale was that the bankruptcy of these carmakers would push their suppliers to shutter operations which would adversely affect Ford since suppliers are shared in the industry.

Since, in a cluster, firms are often downstream from themselves, this way of organizing production provides a self-reinforcing process that maximizes production.

The benefit provided by a greater diversity of inputs also affects how shocks propagate through the network. In particular, it generates complementarities between the firms' operating decisions that can trigger cascades of firm shutdowns that change the shape of the production network, a novel propagation mechanism explored in this paper. To see this, consider a firm that stops production after a severe shock. Its customers, having lost a valuable input, might also shut down. Similarly, its suppliers, now producing a less useful product, are also more likely to stop production. As the same logic applies to the firm's second neighbors, and so on, the initial shock can trigger a cascade that shuts down multiple producers as it propagates upstream and downstream through the network.

One last consequence of the complementarities at work in the economy is that a small change in the environment can trigger a large reorganization of the network. For instance, a small drop in the productivity of an influential firm can lead to the shutdown of its whole neighborhood while activity moves to another part of the network. Through this mechanism, large changes in firm-level distributions can be observed while macroeconomic aggregate are relatively unchanged.

Two features of the environment make the efficient allocation in this economy challenging to solve for. First, since the decision to operate a unit is binary, the optimization problem has a non-convex feasible set and belongs to the class of mixed integer nonlinear problems — a family of problems that are notoriously difficult to solve (Garey and Johnson, 1990). Second, the complementarities between producers make the objective function non-concave. As a result, usual numerical algorithms are ineffective and the Karush-Kuhn-Tucker conditions are not sufficient to characterize a solution. I propose instead a novel approach that involves *reshaping* the original optimization problem such that 1) the reshaped problem can be solved easily, and 2) the solutions to the reshaped and the original problems coincide. This approach is expected to work well if the network of *potential* connections between firms is densely connected. To verify its robustness, I extensively test the approach on sparsely connected networks and find that it performs very well in practice.

While the bulk of the paper focuses on the problem of a social planner, I also show that the efficient allocation can be implemented as a competitive equilibrium with input subsidies and a lump-sum tax on the representative household. Because of the strong complementarities at work in the economy, this equilibrium is however not unique and alternative allocations can also be sustained as equilibria.

Using rich data about firm-level connections, I calibrate the model to the United States economy. To understand how the forces at work in the environment operate to shape the organization of production, it is useful to compare the optimal network, arising from the efficient allocation, to a completely random one. There are several differences between these two networks. A first difference concerns the distributions of the number of suppliers (in-degree) and the number of customers (out-degree) that firms have. These two distributions have fatter right tails in the optimal network — a sign that the planner particularly values highly connected firms. A second difference concerns the amount of clustering that exists between firms, with the optimal network being the more clustered of the two. Together, these two differences confirm that production networks with tightly connected clusters of highly connected firms provide an efficient way of organizing economic activity. Finally, the optimal and the random networks differ in terms of the distributions of output and employment across firms. In the efficient allocation, the planner shuts down low-productivity producers and leverages high-productivity firms by building ecosystems of suppliers and customers around them. As a result, output and employment are more skewed in the optimal network than in its random counterpart.

I use the calibrated economy to investigate how cascades of firm shutdowns arise and propagate through the network. I find that highly connected firms are more resilient to shocks but that, upon shutting down, they create large cascades that spread through the economy. These findings, which are also visible in the data, can be interpreted through the lens of the model. The planner is willing to keep highly connected firms in business even under severe hardship but, if they must shut down, there is no longer a need to maintain the ecosystem around them.

The calibrated economy also features a positive correlation between the size of a cascade and its negative impact on output. While the average cascade has a negligible effect on macroeconomic aggregates, a cascade that originates from a highly connected firm can have a substantial impact on aggregate output.

Cascades, in this environment, are the manifestation of the optimal reorganization of the network in response to shocks. Indeed, through a process akin to creative destruction the planner moves the resources freed by the shutdowns to more productive parts of the network. Preventing this process, for instance by bailing out the distraught firms, increases by about 40% the negative impact of a firm-level shock on aggregate output.

One contribution of this paper is to highlight novel business cycle correlations between aggregate output and the shape of the production network. In the data and in the model, recessions are periods in which, 1) the tails of the degree distributions are thinner, and 2) the amount of clustering in the network is lower. These correlations can be explained through the lens of the model. Expansions are periods in which it is easy to leverage the complementarities at work in the economy by creating tight clusters of economic activity around highly-connected firms. In contrast, recessions are periods in which creating these clusters would be too costly, perhaps because a few influential firms are facing bad shocks, and in which production therefore involves a more diffused, and less efficient, network.

Finally, I consider how the forces at work in the economy interact with microeconomic shocks to influence aggregate fluctuations. To do so, I compare the benchmark economy, in which the production network reorganizes itself efficiently in response to shocks, to an alternative economy in which the network is forever fixed and therefore does not react to shocks in any way. I find that the standard deviation of output is about 30% larger in the fixed network economy. This finding highlights the importance of considering how the network structure of production adapts to shocks in order to properly understand the microeconomic origin of aggregate fluctuations.

Literature review

A large empirical literature documents that losing a supplier is disruptive to a firm's operation. For instance, Hendricks and Singhal (2005) find that firms facing supply chains disturbances face large negative abnormal stock returns. They further find that the impact of the disturbance is long lasting. Wagner and Bode (2008) survey business executives in Germany who report that supply-side disturbances, including supplier suddenly stopping production, were responsible for significant disturbances in production. The Zurich Insurance Group (2015) also conducted a global survey of executives in small and medium enterprises. Of all the respondent, 39% report that losing their main supplier would adversely affect their operation and 14% report that they would need to significantly downsize their business, require emergency support or that they would shut down.

This paper relates to a literature that studies how shocks to interconnected sectors contribute to aggregate fluctuations in exogenous networks (Long and Plosser, 1983; Horvath, 1998; Dupor, 1999).[3] Acemoglu et al. (2012) find that sectoral shocks can lead to large aggregate fluctuations if there is enough asymmetry in the way sectors supply to each other. Acemoglu et al. (2015) further show that inter-sectoral linkages can generate tail-risks in aggregate output. This literature emphasizes the importance of the shape of the network in transmitting idiosyncratic shocks. In contrast, the current paper studies how endogenizing the shape of the network affects the size of aggregate fluctuations.

This paper also contributes to a recent literature on endogenous network formation. Atalay et al. (2011) consider a model in which links between firms are created through random and preferential attachment. They find that this model is able to match properties of the U.S. network data. Closer to the current paper, Oberfield (2017) considers an economy in which producers optimally choose one input from a randomly evolving set of suppliers, thereby creating a production network. Acemoglu and Azar (2017) consider a network of competitive industries in which firms select a production technique that involves different sets of suppliers. They focus on how the evolution of the network can generate economic growth. In contrast, I consider how changes in the operating status of firms shapes the production network. I find that this margin of adjustment plays an important role in the formation of the firm-level production network in the U.S. data. It also

[3]See Carvalho (2014) for an overview of the literature on production networks. Recent contributions to the literature on macroeconomics and networks include di Giovanni et al. (2014), Barrot and Sauvagnat (2016), Atalay (2017), Baqaee and Farhi (2017b), Bigio and La'O (2016), Caliendo et al. (2017a), Caliendo et al. (2017b), Carvalho et al. (2016), Grassi (2017), Lim (2016), Ozdagli and Weber (2017) and Ramírez (2017).

allows cascades of shutdowns to arise, so that their impact on business cycle fluctuations can be studied. Carvalho and Voigtländer (2014) consider how inputs diffuses in a network through input adoption decisions. To do so, they build a model in which producers direct their search for new inputs to the suppliers of their current neighbors in the network. To my knowledge, the current paper is the first to consider how the shape of the production network changes endogenously over the business cycle.

Baqaee (2016) also studies the impact of cascades of firm shutdowns on the macroeconomy. To do so, he considers an exogenous sectoral network in which the mass of firm in each sector can vary. This adjustment margin can lead to further amplification of sectoral shocks in the presence of external economies of scale. In contrast, the current paper studies cascades and fluctuations when the production network evolves endogenously.

Most of the literature on production networks considers economies in which adjustments to changes in the environment are continuous rather than discrete. One exception is Bak et al. (1993). They consider, like in the present work, how discrete adjustment margins can lead to cascading effects. To do so, they propose a model of production and inventories with non-convex interactions between neighbors and find that it can generate large fluctuations from idiosyncratic shocks.[4]

The next section introduces the model and the solution approach. Section 3 discusses the forces at work in the economy. Section 4 calibrates the model and compares it to the data. The last section concludes. All proofs are in the appendix.

2 Model and solution approach

There are n units of production indexed by $j \in \mathcal{N} = \{1, \ldots, n\}$. Each of these units produces an intermediate good that can be used in the production of a final good or as an input in the production of other intermediate goods. A final good producer uses a CES production function with elasticity of substitution $\sigma > 1$,

$$Y = \left(\sum_{j \in \mathcal{N}} c_j^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

to convert intermediate goods $\{c_j\}_{j \in \mathcal{N}}$ into aggregate output Y. A representative household consumes the final good and supplies L units of labor inelastically.

Firm j has access to a production technology that converts a vector x_j of intermediate inputs

[4]See also Blume et al. (2011a) and Blume et al. (2011b) for studies of cascading failures and network formation.

and l_j units of labor into y_j units of good j according to the production function

$$y_j = \frac{A}{\alpha^{\alpha}\left(1-\alpha\right)^{1-\alpha}} z_j \left(\sum_{i\in\mathcal{N}} x_{ij}^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}\alpha} l_j^{1-\alpha}$$

where x_{ij} denotes the intermediate input from firm i, $0 < 1-\alpha < 1$ is the labor intensity, $\varepsilon > 1$ is the elasticity of substitution between intermediate inputs, $A > 0$ is aggregate productivity and $0 < \underline{z} < z_j < \overline{z}$ is firm-specific total factor productivity. Since the number of firms is finite, shocks to z generate aggregate fluctuations in this economy.

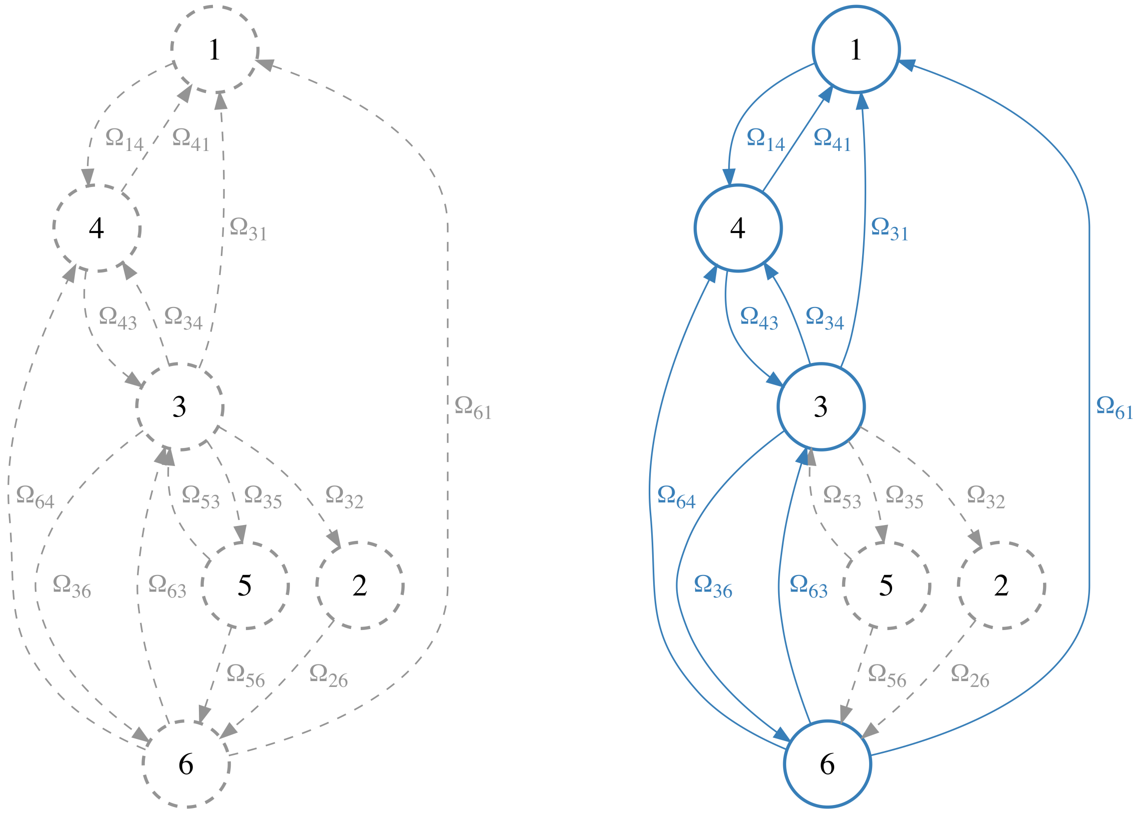
Firm j can use good i in production only if there exists a *connection* from firm i to firm j. The set of all connections is fully described by an $n \times n$ adjacency matrix Ω such that $\Omega_{ij} = 1$ if firm j can use inputs from i and $\Omega_{ij} = 0$ otherwise. Without loss of generality, assume that $\sum_{i\in\mathcal{N}} \Omega_{ij} \geq 1$ otherwise firm j cannot produce and we can redefine the economy without it. As it describes the set of inputs that are usable by each firm, Ω is part of the production possibility frontier of the economy.

A firm can only produce if it is *operating*, which requires the payment of a fixed cost $f \geq 0$, in units of labor. The vector θ describes the operating status of the firms, such that $\theta_j = 1$ if firm j is operating and $\theta_j = 0$ otherwise. This fixed cost captures overhead labor, such as managers and other non-production workers, that is necessary for production. A broader interpretation could include the cost of fixed structures or any other factor that does not scale with production.[5]

Together, Ω and θ describe the network structure of production in this economy. While the set of *possible* connections Ω is taken as exogenous, the set of *active* connections evolves endogenously as economic conditions affect which firms operate. To fix ideas, Figure 1a provides an example of the set of possible connections Ω in a simple economy with six firms. Each arrow represents a connection $\Omega_{ij} = 1$, with the direction of the arrow showing the potential movement of goods between the firms. The set of active connections, in blue in Figure 1b, is selected endogenously and depends on the set of active firms, also shown in blue.[6]

[5] Bresnahan and Ramey (1994) show that the decision to operate or not is responsible for about 80% of the fluctuations in output at the plant level in the automobile industry.

[6] While the model is kept simple for tractability, it can accommodate individual link formation as well. This can be done in two ways. First, one can simply install an artificial firm on a link. By operating that firm or not, the link becomes active or not. Another approach involves duplicating each firm in the economy so that each duplicate is connected to a different subsets of the firm's full connections. Then, the planner's problem can be augmented by an additional constraint such that only one of the duplicate operates (this constraint is linear in θ so that the first-order conditions are still sufficient).



(a) All firms and connections

(b) Active firms and connections

Figure 1: Operating decisions determine the shape of the network endogenously

2.1 Planner's problem

Consider the problem $\mathcal{P}_{SP}$ of a social planner in this economy. The planner maximizes the utility of the representative household

$$\max_{\substack{c\geq 0,\, x\geq 0,\, l\geq 0\\ \theta\in\{0,1\}^n}} \left(\sum_{j\in\mathcal{N}} c_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

subject to a resource constraint for each intermediate good $j\in\mathcal{N}$,

$$c_j + \sum_{k\in\mathcal{N}} x_{jk} \leq \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_j \left(\sum_{i\in\mathcal{N}} x_{ij}^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\alpha\frac{\varepsilon}{\varepsilon-1}} l_j^{1-\alpha}, \tag{2}$$

a resource constraint for labor,

$$\sum_{j\in\mathcal{N}} l_j + f\sum_{j\in\mathcal{N}} \theta_j \leq L, \tag{3}$$

an operation constraint for each firm $j \in \mathcal{N}$,

$$\{\theta_j = 0\} \Rightarrow \{l_j = 0\}, \tag{4}$$

such that a firm can only operate if the fixed cost f is paid, and a connection constraint for each pair of firms $(i, j) \in \mathcal{N}^2$

$$\{\Omega_{ij} = 0\} \Rightarrow \{x_{ij} = 0\} \tag{5}$$

so that input i can only be used by j if an appropriate connection exists.

It is convenient to integrate the operation constraints (4) and the connection constraints (5) directly into the goods resource constraints (2) such that they become

$$c_j + \sum_{k\in\mathcal{N}} x_{jk} \leq \theta_j \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z_j \left(\sum_{i\in\mathcal{N}} \Omega_{ij} x_{ij}^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\alpha \frac{\varepsilon}{\varepsilon-1}} l_j^{1-\alpha}, \tag{6}$$

for all $j \in \mathcal{N}$. Notice the addition of the terms θ_j and Ω_{ij} in the production function. The term θ_j imposes that firm j optimally produces goods only if $\theta_j = 1$. Otherwise it would be using valuable resources to generate no output. Similarly, the addition of Ω_{ij} imposes that firm j can only use goods x_{ij} if a connection $\Omega_{ij} = 1$ exists.

2.2 Solution approach

To isolate the difficulties that arise in solving $\mathcal{P}_{SP}$, it is useful to first fix the set of operating firms θ and to solve for the optimal allocation of resources in the exogenous network.

Planner's problem with exogenous firm status

For a given vector θ, maximizing (1) subject to the constraints (3) and (6) is a convex maximization problem and the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize its solution. Denote by w the Lagrange multiplier on the labor resource constraint (3) and by λ_j the Lagrange multiplier on the resource constraint for the jth good (6). As in **?**, it is convenient to define $q = w/\lambda_j$. Since, from the first-order condition on l_j, output can be written as $(1-\alpha)\, y_j = q_j l_j$, it is natural to interpret q_j as a measure of the labor productivity of firm j. I also refer to $1/q_j$ as the production cost of firm j.

Combining the first-order conditions of $\mathcal{P}_{SP}$ with the production function yields the following proposition.

Proposition 1. *The Lagrange multipliers associated with the efficient allocation satisfy*

$$q_j = z_j \theta_j A \left(\sum_{i\in\mathcal{N}} \Omega_{ij} q_i^{\varepsilon-1} \right)^{\frac{\alpha}{\varepsilon-1}}, \quad \forall j \in \mathcal{N} \tag{7}$$

Furthermore, there is a unique vector q that satisfies (7) such that $q_j > 0$ if firm j has access to a closed loop of active suppliers.

Equation (7) shows the importance of the production network in determining labor productivity in this economy. In particular, a firm that has access to many active suppliers benefits from a greater diversity of inputs and produces more efficiently. The importance of this benefit is inversely related to the elasticity of substitution ε. Similarly, a firm using cheap inputs (low $1/q$) is able to produce a cheap product.

Equation (7) can be solved easily by iterating on the mapping. Its solution q can then be used to fully characterize the optimal allocation of resources in the economy. For instance, aggregate output Y can be computed directly from q, as shown by the following lemma.

Lemma 1. *Aggregate output in the optimal allocation is given by*

$$Y = Q\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) \tag{8}$$

where $Q \equiv \left(\sum_{j\in\mathcal{N}} q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$.

This lemma shows that aggregate output is the product of aggregate productivity Q and the amount of labor available after fixed costs have been paid.

Similarly, the vector of labor l used by each firm satisfies

$$l = (1-\alpha)\left[I_n - \alpha\Gamma\right]^{-1}\left(\frac{q}{Q}\right)^{\circ(\sigma-1)}\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) \tag{9}$$

where $\circ$ denotes the Hadamard exponent (each element of q raised to the power $\sigma-1$), I_n denotes the $n\times n$ identity matrix and where Γ is an $n\times n$ matrix with elements $\Gamma_{jk} = \frac{\Omega_{jk}q_j^{\varepsilon-1}}{\sum_{i=1}^{n}\Omega_{ik}q_i^{\varepsilon-1}}$. Γ_{jk} captures the relative importance of firm j as a supplier to firm k. Since the Leontief inverse $\left[I_n-\alpha\Gamma\right]^{-1}$ can be written as $I_n + \alpha\Gamma + (\alpha\Gamma)^2 + \dots$, we see that the planner allocates a large fraction of the labor available to firms that are important suppliers in the production network.[7]

Planner's problem with endogenous firm status

We now turn to the full planner's problem in which the operational status θ of the firms is a choice variable. Combining (7) and (8) we can rewrite $\mathcal{P}_{SP}$ as

$$\mathcal{P}_{SP}\colon\ \max_{\theta\in\{0,1\}^n} Q\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right)$$

[7] See the proof of Lemma 1 for the derivation of (9). The proof also shows that all other relevant quantities can be computed analytically from q. See equations (19) for y and (20) for x.

where q solves, for each $j \in \mathcal{N}$,

$$q_j = z_j \theta_j A \left(\sum_{i \in \mathcal{N}} \Omega_{ij} q_i^{\varepsilon - 1} \right)^{\frac{\alpha}{\varepsilon - 1}}. \tag{7}$$

These two equations highlights the tradeoffs facing the planner. When deciding whether to operate firm j or not, the planner compares the cost of the decision with its benefit. On the cost side, operating j takes away f units of labor that could be employed by other firms. On the benefit side, operating j increases the aggregate labor productivity of the network Q. In particular, because of the recursive nature of (7), operating j improves the efficiency of all the firms that are downstream from it.

Two features of $\mathcal{P}_{SP}$ make this problem hard to solve. First, the vector θ is limited to the corners of the unit hypercube $\{0,1\}^n$, such that the constraint set is clearly not convex. Second, the objective function is not (quasi) concave. Each one of these features implies that the standard Karush-Kuhn-Tucker conditions cannot be used to solve this problem.[8]

There is however a naive way of solving $\mathcal{P}_{SP}$. Since there is a finite number of possible vectors θ, one can try them all. For each θ, q can be found by iterating on (7) and the objective function can then be computed using (8). This *exhaustive* approach is guaranteed to find the correct solution but is limited to economies with only a few firms. Indeed, since there are 2^n possible vectors θ, for n larger than a few dozens the huge number of possibilities become impossible to handle.[9]

Instead of relying on this computationally intensive approach, I propose to solve $\mathcal{P}_{SP}$ by considering an alternative optimization problem that is 1) easy to solve and 2) whose solution is also a solution to $\mathcal{P}_{SP}$. For that purpose, consider the following problem, denoted by $\mathcal{P}_{RR}$, which is obtained by *relaxing* and *reshaping* $\mathcal{P}_{SP}$,

$$\mathcal{P}_{RR}\colon \max_{\theta \in [0,1]^n} Q \left(L - f \sum_{j \in \mathcal{N}} \theta_j \right)$$

where the vector q solves, for each $j \in \mathcal{N}$,

$$q_j = z_j \theta_j^a A \left(\sum_{i \in \mathcal{N}} \theta_i^b \Omega_{ij} q_i^{\varepsilon - 1} \right)^{\frac{\alpha}{\varepsilon - 1}}. \tag{10}$$

$\mathcal{P}_{RR}$ differs from $\mathcal{P}_{SP}$ in two important ways. First, the binarity constraint $\theta \in \{0,1\}^n$ is now

[8] Using the standard classification of optimization problems, $\mathcal{P}_{SP}$ belongs to the class of Mixed Integer Nonlinear Problems (MINLP). Their combinatorial nature makes these problems notoriously challenging to solve and they are, from the perspective of computational complexity theory, NP-Hard (Karp, 1972).

[9] In some models, it is possible to order firms in some way and to progressively shut down the "worst" ones until the desired allocation is found. Here, however, firms differ along two dimensions (their position in the network Ω and their productivity z) so that this approach does not apply directly. Despite several attempts at finding a proper variable to rank the firms, I have not been able to make this approach work.

relaxed to $\theta \in [0,1]^n$, so that θ can now take values in the inside of the (convex) unit hypercube. Second, (7) now becomes (10) which includes the *reshaping parameters* $a > 0$ and $b \geq 0$. These parameters modify the shape of the optimization problem, but only away from potential solutions to $\mathcal{P}_{SP}$. To see this, remember that any solution to $\mathcal{P}_{SP}$ is such that $\theta_j \in \{0,1\}$ for all j, so that, for these points, $\theta_j^a = \theta_j$. Similarly, for the b parameter, if $\theta_i = 0$ then $q_i = 0$, and if $\theta_i^b = 1$ then $\theta_i^b q_i^{\varepsilon-1} = q_i^{\varepsilon-1}$. We see that (7) and (10), and therefore the objective functions of $\mathcal{P}_{SP}$ and $\mathcal{P}_{RR}$, coincide over the set $\theta \in \{0,1\}^n$.[10]

As mentioned earlier, the relaxed and reshaped problem $\mathcal{P}_{RR}$ is only useful if, unlike $\mathcal{P}_{SP}$, it is easy to solve and if a solution to $\mathcal{P}_{RR}$ is also a solution to $\mathcal{P}_{SP}$ — the problem we are actually interested in solving. It turns out that both of these objectives can be satisfied when the reshaping parameters take on the values

$$a = \frac{1}{\sigma - 1} \quad \text{and} \quad b = 1 - \frac{\varepsilon - 1}{\sigma - 1} \tag{$\star$}$$

which we assume from now on.

The following proposition establishes that $\mathcal{P}_{RR}$ is easy to solve when Ω belongs to a particular family of matrices.

Proposition 2. *If $\Omega_{ij} = c_i d_j$ for some vectors c and d then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{P}_{RR}$.*

For more general networks, a similar result can be established if the network is sufficiently connected.

Proposition 3. *Let $\sigma = \varepsilon$ and suppose that the fixed cost $f > 0$ and the dispersion in productivities $\overline{z} - \underline{z} > 0$ are not too big. If the network Ω is large enough and sufficiently connected with $\Omega_{ii} = 0$ for all i, then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{P}_{RR}$.*

Propositions 2 and 3 both show that, under certain conditions, any feasible point θ^* that satisfies the first-order conditions and the complementary slackness conditions solves $\mathcal{P}_{RR}$. As a result, standard numerical algorithms, such as gradient ascent, can be used to easily solve $\mathcal{P}_{RR}$ even when the network contains thousands of firms.[11]

Note that, while these two propositions rely on assumptions about Ω, they only provide *sufficient* conditions for the result. In particular, Section 2.3 below shows that, in practice, the reshaping approach finds the correct solution to the planner's problem for a broader class of matrix Ω.

The following proposition establishes under what conditions a solution to $\mathcal{P}_{RR}$ is also a solution to $\mathcal{P}_{SP}$.

[10] Note that this reshaping is *not* a standard change of variables.

[11] Appendix B provides a fast algorithm to solve $\mathcal{P}_{RR}$.

Proposition 4. *If a solution θ^* to $\mathcal{P}_{RR}$ is such that $\theta_j^* \in \{0,1\}$ for all j, then θ^* also solves $\mathcal{P}_{SP}$.*

The result follows directly from the fact that the feasible set of $\mathcal{P}_{RR}$ contains the feasible set of $\mathcal{P}_{SP}$ and that both of their objective functions coincide when $\theta_j \in \{0,1\}$ for all $j \in \mathcal{N}$.

Together, Propositions 2 to 4 offer a convenient way to solve $\mathcal{P}_{SP}$. One can simply solve the relaxed and reshaped problem $\mathcal{P}_{RR}$ and look at the solution θ^*. If θ^* is such that, for all j, θ_j^* is either 0 or 1, then θ^* also solves $\mathcal{P}_{SP}$. It is therefore important, for the reshaping approach to work, that the solution to $\mathcal{P}_{RR}$ is such that $\theta^* \in \{0,1\}^n$. Fortunately, this can be achieved in large connected networks when the reshaping parameters are such that condition ($\star$) is satisfied.

To see why, it is convenient to consider the first-order conditions of the equivalent problem $\mathcal{P}'_{RR}$ in which (10) is treated as an inequality constraint

$$\mathcal{P}'_{RR}\colon \max_{\theta\in[0,1]^n, q} \left(\sum_{j\in\mathcal{N}} q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}} \left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) \tag{11}$$

subject to, for all $j \in \mathcal{N}$,

$$q_j \leq A z_j \theta_j^a B_j^\alpha \tag{12}$$

where $B_j = \left(\sum_{i\in\mathcal{N}} \theta_i^b \Omega_{ij} q_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}$ is a composite of the productivity of firm j's suppliers.[12]

Combining the first-order conditions of this problem yields

$$\frac{\partial q_k}{\partial\theta_k}\frac{\partial Q}{\partial q_k}\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) - fQ + \sum_{j\in\mathcal{N}}\beta_j\left(\frac{\partial q_k}{\partial\theta_k}\frac{\partial B_j}{\partial q_k} + \frac{\partial B_j}{\partial\theta_k}\right)\frac{\partial q_j}{\partial B_j} = \bar{\mu}_k - \underline{\mu}_k \tag{13}$$

where β_j is the Lagrange multiplier on the jth inequality constraint (12) and where $\underline{\mu}_k$ and $\bar{\mu}_k$ are the Lagrange multipliers on $\theta_k \geq 0$ and $\theta_k \leq 1$.[13] We see from (13) that an increase in θ_k has a *direct* and and *indirect* impact on the objective function. The direct impact captures the increase in productivity of firm q_k which leads to an increase in aggregate productivity Q (first term in 13). Increasing θ_k also lead to higher costs of operation (second term in 13). The indirect impact of θ_k operates through its action on the other firms. First, the larger q_k leads to an increase in the labor productivity of all of k's customers (first term in the summed parenthesis). Second, increasing θ_k also increases the composite input of all of k's customers through the reshaping term θ_k^b (second term in the summed parenthesis).

Equation (13) determines the operational status θ_k of firm k such that if its left-hand side is larger than 0 the firm is active ($\theta_k = 1$) and, similarly, if it is smaller than 0 the firm is inactive ($\theta_k = 0$). The key to the reshaping approach is to make the left-hand side of (13) independent of

[12] $\mathcal{P}'_{RR}$ is equivalent to $\mathcal{P}_{RR}$ since the inequality constraint always binds at the optimum.

[13] The partial derivatives of q_j are to be understood for the binding inequality constraint, i.e. such that $\frac{\partial q_j}{\partial\theta_j} = Az_j a\theta_j^{a-1} AB_j^\alpha$ and $\frac{\partial q_j}{\partial B_j} = Az_j\theta_j^a A\alpha B_j^{\alpha-1}$.

any firm-specific endogenous variables. The following lemma shows how this can be done.

Lemma 2. *Under the condition* ($\star$)*, the first-order condition* (13) *for the operating decision of firm j only depends on θ_j through aggregates.*

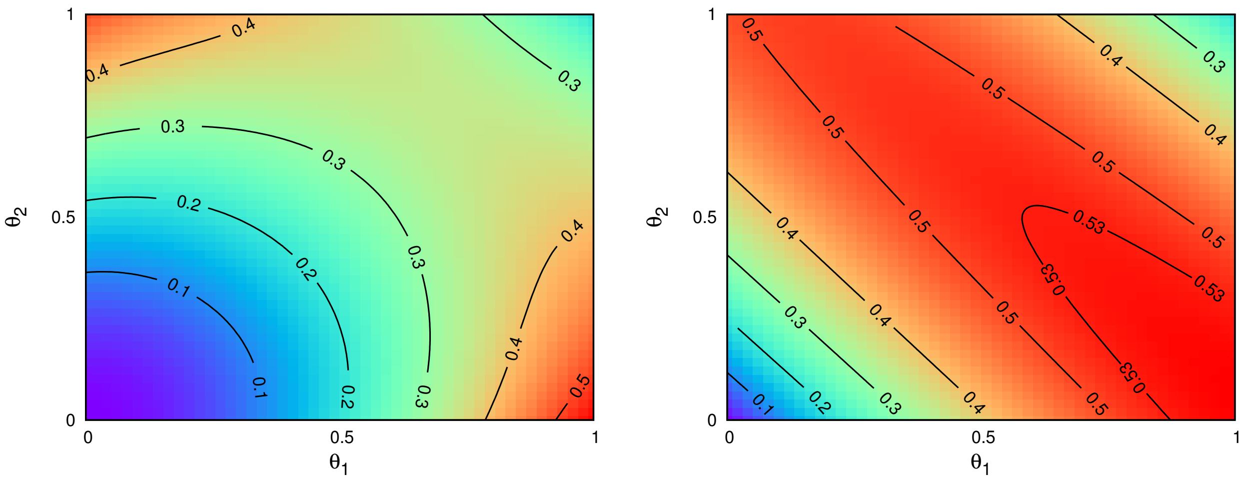
The proof of Lemma 2 shows that, when the parameters a and b satisfy the ($\star$) condition, the first-order equation (13) depends on θ_j only through $\{B_j\}_{j\in\mathcal{N}}$ and Q. But these variables are aggregate, in the sense that they are summations over a large number of firms. Therefore, as the number of firms increases, and the network becomes increasingly connected, they become more and more independent of θ_k and, as a result, so does equation (13). In the limit, the marginal benefit of increasing θ_k, the left-hand side of (13), is therefore independent of θ_k itself and depends only on the aggregate state of the network and on firm k's exogenous characteristics. If this marginal benefit is positive, the planner increases θ_k until it reaches 1. If it is negative, the planner decreases θ_k until it reaches 0. The solution to $\mathcal{P}_{RR}$ is therefore such that $\theta_k \in \{0,1\}$ for all k. Using Proposition 4 we can then conclude that the solution to $\mathcal{P}_{RR}$ also solves $\mathcal{P}_{SP}$.

Example with two firms

To better understand how the reshaping approach works, it is useful to consider a simple economy with two firms and a complete set of potential connections between them. Consider first the relaxed problem without any reshaping ($a = 1, b = 0$). The contour plot of its objective function $V_R(\theta)$ is shown in Figure 2a where warmer colors represents higher utility for the planner. The horizontal and vertical axes refer to the operating decisions of firms 1 and 2, respectively. We see that V_R is shaped like a saddle with local maxima at $[\theta_1\ \theta_2] = [1\ 0]$ and $[0\ 1]$ and local minima at $[0\ 0]$ and $[1\ 1]$. The planner therefore prefers to have a single firm operating than to have both, or neither, firms operating. The global maximum is at $\theta = [1\ 0]$.

Since V_R is not concave, the first-order conditions are not sufficient to characterize the optimum and finding a solution to the FOCs does not guarantee that this solution is the global maximum. As a result, this problem cannot be solved reliably with standard numerical algorithms. Starting from an initial point, these algorithms generally operate by following the steepest slope, which can easily lead to the local (but not global) maximum at which only firm 2 operates.

Consider instead the objective function V_{RR} of the same optimization problem but reshaped with the parameters given by ($\star$). Figure 2b shows the contour plot of V_{RR}. Notice first that the reshaping has not changed the value of the objective function at the corners $\theta \in \{0,1\}^2$. As a result, the corner with the highest utility according to V_{RR} is also the corner with the highest utility according to V_R. Second, we see that the reshaping basically stretches the objective function between the corners so as to transform the non-concave function V_R into the concave function V_{RR}.



(a) The objective function $V_R(\theta)$ of the relaxed (but not reshaped) problem is not concave

(b) The objective function $V_{RR}(\theta)$ of the relaxed and reshaped problem is concave

Figure 2: Example: reshaping the objective function

The first-order conditions are now sufficient to characterize a solution and, starting from any point $(\theta_1, \theta_2) \in [0,1]^2$, standard numerical algorithms like gradient ascent will converge to the global maximum at which only firm 1 operates, which is also the solution to $\mathcal{P}_{SP}$.

From this example, one can get some intuition about how the reshaping approach makes the problem easier to solve. Let us first consider the role played by the parameter a. Suppose that we begin the optimization at the local maximum $\theta = [0\ 1]$ in Figure 2a, and are now looking for alternative allocations in its neighborhood to improve the objective function. Moving towards the global maximum at $\theta = [1\ 0]$ incurs a marginal cost proportional to fQ. The marginal benefit of this move is, however, proportional to $\partial Q/\partial\theta_1 \propto \theta_1^{\sigma-2}$ without the deformation. As a result, for θ_1 small, the marginal benefit is close to 0 and it declines with the elasticity σ — operating a new firm provides little benefit if the goods are highly substitutable. Increasing θ_1 is thus not *locally* advantageous and the optimization converges erroneously to the local maximum. Including the reshaping parameter a increase the local incentives to increase θ_1. In particular, the marginal benefit is now proportional to $\partial Q/\partial\theta_1 \propto \theta_1^{a(\sigma-1)-1}$. Under the parameter values ($\star$), this benefit is constant and of the same order of magnitude as the cost fQ, which helps the optimization algorithm to move towards the global maximum. A similar intuition applies for the reshaping parameter b, but the argument involves the addition of an input into the production of intermediate goods.

2.3 Testing the solution approach

The theoretical results of the last section show that the reshaping approach finds the correct solution to $\mathcal{P}_{SP}$ when the number of firms is large and the network is sufficiently connected. The approach also work very well when these assumptions are relaxed. To show this, I randomly generate a large set of economies with various parameters and matrices Ω. In each case, I find the optimal solution by comparing the utility provided by each of 2^n possible networks. I then compare this allocation to the one found by the reshaping approach. Since the exhaustive search always finds the correct solution to $\mathcal{P}_{SP}$, any difference between the two allocations is therefore indicative of a failure of the reshaping approach. Since, as explained earlier, the exhaustive search can only handle economies with few firms, the exercise is limited to environments with small n. Nevertheless, since small and sparse networks are the ones for which we expect the reshaping approach to fail, the exercise provides a useful test of its performance.

The results are presented in Table 1. We see that the reshaping approach (first two rows) works remarkably well even for small and sparsely connected networks. The algorithm attributes the correct operational status to more than 99.8% of the firms. It also finds output values that are within 0.002% of the correct value.[14] In contrast, attempting to solve for the efficient allocation by relaxing the problem but without reshaping it (last two rows) yields more than 15% of the firms with the wrong operational status and the average error in output can be close to 1%, a large number when considering business cycle fluctuations.

Table 1: Testing the reshaping approach for small networks

	Number of firms n			
	8	10	12	14
A. With reshaping				
Firms with correct θ	99.9%	99.9%	99.9%	99.8%
Error in output Y	0.00039%	0.00081%	0.00174%	0.00171%
B. Without reshaping				
Firms with correct θ	84.3%	83.2%	82.3%	81.3%
Error in output Y	0.84%	0.89%	0.93%	0.98%

Notes: Random networks with parameters $f \in \{0.05/n, 0.1/n, 0.15/n\}$, $\sigma_z \in \{0.34, 0.39, 0.44\}$, $\alpha \in \{0.45, 0.5, 0.55\}$, $\sigma \in \{4, 6, 8\}$ and $\varepsilon \in \{4, 6, 8\}$. For each combination of the parameters, 1000 different economies are created. For each economy, productivity is drawn from $\log(z_k) \sim$ iid $\mathcal{N}(0, \sigma_z)$ and Ω is drawn randomly such that a firm has on average five possible incoming connections. As a result, the networks Ω have different degrees of sparsity. The network is redrawn until it is strongly connected. A network is kept in the sample only if the first-order conditions converge to a solution for which θ hits the bounds.

It is also possible to test the reshaping approach on large networks. While the true solution to the planner's problem is unknown in this case, we can verify whether there exist welfare-improving deviations from the allocation found by the approach. In particular, I verify whether changing the operational status θ of a single firm improves the utility of the planner. I keep repeating this procedure as long as there are deviations to be found and then compare this deviation-free solution

[14] Note that since the problem is solved numerically, any algorithm is expected to converge on the wrong solution from time to time. For instance, the algorithm might converge to a point that is not a solution but for which the first-order conditions are satisfied within numerical tolerance.

to the original one given by the reshaping approach.

Since this procedure is computationally costly, I only consider large economies that follow the calibration presented in Section 4.2. The results are presented in Table 2. Again, the reshaping approach performs very well. After all the possible deviations are accounted for, more than 99.9% of the firms have kept their operational status and aggregate output has changed by a negligible amount.[15] In contrast, without the reshaping, the deviations reveal that more than 30% of the firms were assigned the wrong θ and the error in aggregate output amounts to about 0.7%. While this test does not guarantee that the reshaping approach finds the correct efficient allocation, it provides a good indication that there are no obvious mistakes in its solution.

Table 2: Testing the reshaping approach for large networks

	With reshaping	Without reshaping
Firms with correct θ	99.98%	69.3%
Error in output Y	0.00007%	0.696%

Notes: I simulate 100 different networks Ω and, for each Ω, draw 100 productivity vectors z that satisfy the properties of the calibrated economy (see Section 4.2). I run the procedure described in the text on each of them and report average results.

2.4 Equilibrium

The efficient allocation described in the previous sections can also be implemented as an equilibrium in which firms use monopoly power when selling their good. In such an environment, a firm j faces a wage W and prices of intermediate inputs $\{P_{ij}\}_{i\in\mathcal{N}}$. It then maximizes profits

$$\pi_j = P_j c_j + \sum_{i\in\mathcal{N}} x_{ji} P_{ji} - \sum_{i\in\mathcal{N}} x_{ij} P_{ij} - W f \theta_j - W l_j \tag{14}$$

subject to a feasibility condition, $c_j + \sum_{i\in\mathcal{N}} x_{ji} \leq y_j$, and while taking into account its pricing decisions on the demand of the final goods producer, $c_j = C\left(P_j/P\right)^{-\sigma}$, and on the demand of other intermediate goods producers, $x_{ji} = X_i\left(P_{ji}/\bar{P}_i\right)^{-\epsilon}$, where X_i and $\bar{P}_i$ are the usual Dixit-Stiglitz aggregators.

We can define an equilibrium in the usual way: an equilibrium is a set of prices for final goods $\{P_j\}_{j\in\mathcal{N}}$, for intermediate goods $\{P_{ij}\}_{i,j\in\mathcal{N}^2}$ and for labor W; and an allocation $\{c_j, l_j, \theta_j\}_{j\in\mathcal{N}}$ and $\{x_{ij}\}_{i,j\in\mathcal{N}^2}$ such that: 1) Given prices for labor W and intermediate inputs $\{P_{ij}\}_{i\in\mathcal{N}}$, each firm j maximizes profits (14) subject to the feasibility condition and the demand curves, and 2) The goods and the labor markets clear.

[15] When the reshaping approach fails it is in general because it gets the wrong operating status for a firm that is fairly isolated from the rest of the network. Since, these firms are in general small, they only have little influence on aggregate production, which explains why the error in output is very small in Table 2.

If we focus, as is common in the literature, on environments in which the elasticities of substitution σ and ε are equal, the following proposition shows that the efficient network can be interpreted as emerging from the individual decisions of agents interacting through markets.[16]

Proposition 5. *Suppose that the solution to $\mathcal{P}_{RR}$ also solves $\mathcal{P}_{SP}$ and that $\sigma = \epsilon$. Then the efficient allocation can be implemented as an equilibrium with a subsidy to the purchase of intermediate inputs and a lump-sum tax on the representative consumer.*

The proof of the proposition shows that, in equilibrium, a producer sells to different buyers at a common price, such that $P_j = P_{ji}$. In addition, the proof allows us to establish some intuitive links between the equilibrium and the efficient allocation. In particular, the price P_j is set at a constant markup over the social cost of production λ_j. To implement the efficient allocation, a subsidy to the purchase of intermediate inputs is therefore needed to align the incentives between the customer and the supplier. In addition, the link between prices and social costs of production allows us to interpret equation (7) in terms of prices. Since, as we will see in the next section, (7) embeds the key mechanisms through which shocks propagate in this economy, these mechanisms can also be understood in terms of price movements across producers.

The equilibrium of Proposition 5 is not unique. As the next section explains, this economy features strong complementarities between the firm operating decisions. These complementarities, which find support in the data, are crucial for the mechanisms of the model and can also give rise to multiple equilibria.

3 Economic forces

We now explore how the economic forces at work in the environment influence the shape of the production network, the propagation of shocks and the distributions of firm-level outcomes.

At the heart of the model's mechanisms is a benefit from producing with a greater diversity of inputs. This *diversity benefit* which, as we will see in the next section, has strong support in the data, is a consequence of how intermediate inputs are aggregated by the production functions of the firms. Indeed, from equation (7) we see that an additional supplier increases the firm's productivity q or, equivalently, that it decreases it marginal cost of production $1/q$. The elasticity of substitution ε is the key determinant of the importance of this mechanism. Intuitively, with an additional input, a firm now has access to a broader set of techniques with which to produce. In fact, it can simply decide to not use the new input at all and be at least as efficient as before.

[16] When $\sigma \neq \varepsilon$ the same good sells for different prices to the representative customer and to intermediate goods producers, and arbitrage opportunities might arise.

3.1 Firm selection and clustering

A direct consequence of this diversity benefit is that firms with multiple neighbors, upstream or downstream, are more likely to operate in the efficient allocation. Indeed, firms with many suppliers are particularly efficient, and the planner is likely to use them in production. Similarly, firms that supply to multiple producers improve the production process of their many customers and are also more likely to operate. Note that this mechanism extends to firms with many *indirect* neighbors (second neighbors, and so on). Because of the recursive nature of equation (7), operating a firm benefits all producers that are downstream from it. Similarly, a firm with many upstream producers enjoys a higher productivity q, which increases the incentives of the planner to operate it.

While it prefers to operate firms that have many neighbors, the planner is obviously also determining how many neighbors a firm actually has. As a result, the planner tends to create networks in which many firms have many neighbors. Organizing production in this way involves creating tightly connected clusters of operating firms. In this type of configuration, each firm j benefits from the presence of many neighbors that, in turn, benefit from a more productive firm j. This recursive process reinforces the productivity q of all the members of the cluster and makes organizing production in this way very efficient.

A simple example is useful to understand what features of the environment make clustering more or less desirable. Suppose that the planner wants to operate n firms and can decide to do so by operating clusters of m fully-connected firms, where each producer supplies to the $m-1$ other firms.[17] Which amount of clustering m is preferred? We can easily compute the productivity q of each firm in an m-cluster. Assuming that firms share the same $z=1$ for simplicity, symmetry imposes that all firms have the same q. From (7), we find that $q=(m-1)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}}$ and from (8) we can compute total output as

$$Y = n^{\frac{1}{\sigma-1}} (m-1)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}} (L - fn) .$$

Since this expression is increasing in m, the planner prefers to operate a unique tightly connected group of firms instead of a large number of connected pairs of firms. This preference for clustering is more important when the share α of intermediate inputs in production is large and when the elasticity of substitution ε between these inputs is low. Under these conditions, the planner values additional connections more, which leads to a larger desire for clustering.

When total factor productivity z varies across firms, the optimal network tends to cluster economic activity around the most productive firms. Doing so allows the planner to magnify the

[17] For simplicity, let us assume that n is large enough and that it is divisible by m. While the current example might seem abstract it fits in the framework outlined in the last section. The whole economy would be formed of groups of n firms which various degree of clustering. The exercise would then be about which of these groups the planner would prefer to operate.

impact of these high-z firms. Figure 3 presents an example of this process in an economy in which the network Ω consists of two groups of five fully-connected firms linked together by a single pair of connections. The figure shows the same economy under four different vectors z. Two properties of optimal networks are visible in the figure. First, to maximize the benefit from input diversity, the planner clusters economic activity in either the top or the bottom group of firms. Operating a few firms in each cluster is, in general, a poor way to organize production. Second, the cluster of firms that actually produces tends to be the one with the firm with the highest z, thereby leveraging the benefit of this high productivity.

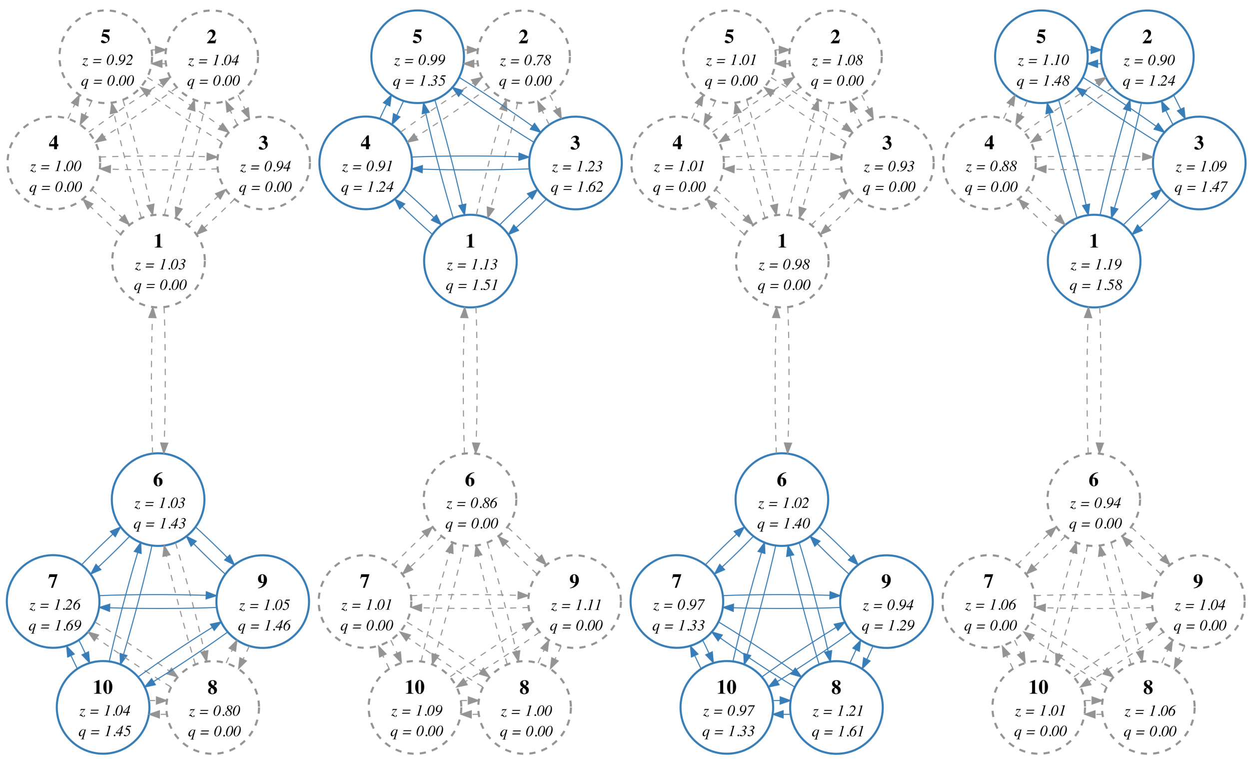


Figure 3: The optimal network features clusters of economic activity

3.2 Downstream and upstream cascades

In this economy, two mechanisms are responsible for the propagation of local shocks through the network. The first mechanism is at work even when the set of active firms is fixed. It operates, through the recursive structure of equation (7), by changing the firms' marginal costs of production. Consider for instance a firm facing a bad z shock. Since, the firm becomes less efficient at producing, its marginal cost increases and it passes that increase to its customers through a higher price. As a result, the customers face a higher cost of production which they, in turn, also pass on to their own customers, and so on. Through this first mechanism, shocks propagate *downstream* through the network.[18]

While this first mechanism operates even when the production network is fixed, a second mechanism operates precisely because the production network changes in response to shocks. Here, the benefit provided by a greater diversity of inputs plays an important role. In particular, it generates complementarities between the operating decisions of nearby firms. Indeed, if a firm loses a supplier or a customer, it suddenly has fewer neighbors and is therefore less likely to operate. As a result, firms that are close to each other in the network tend to operate, or not, together. When producers are subject to shocks, these complementarities can create cascades of shutdowns. Consider again a firm that stops production after suffering from a severe z shock. In response, its first neighbors, having lost a useful supplier or a valuable customer, are also likely to shutdown. Since the same logic then applies to *these* firms' first neighbors and so on, the initial z shock can trigger a cascade. Importantly, this second mechanism, in contrast to the first one, generates both *downstream* and *upstream* propagation of shocks. The endogenous evolution of the production network, one of the key new adjustment margin explored in this paper, is therefore crucial to obtain upstream propagation in this environment.

3.3 Small shock and large reorganization

In the efficient allocation, a small variation in the environment can lead to a large reorganization of the network. This unusual feature of the model arises because of non-convexities in the problem of the planner that arise from the complementarities in operating decisions. To better understand how these large effects arise from small changes, consider that the planner essentially compares the 2^n possible network configurations and selects the one providing the highest utility. As the environment changes, say a firm's z increases, there is a point at which a previously optimal configuration is replaced by a new one. In turns out that this new configuration can be substantially different than the original one.

Figure 4 provides an example. The primitives of the two environments are identical except that, in the economy on the left, the central node is slightly more productive than in the economy on the right. While the drop in productivity from left to right is negligible, it triggers a large reorganization of the production network. On the left, the planner prefers to make goods transit through the productive central firm while on the right it creates a loop of active firms to avoid the slightly less productive central firm.

While the shape of the network changes substantially, aggregate output is barely affected by the

[18] Several recent papers document that shocks propagate through supply chains. Barrot and Sauvagnat (2016) find that suppliers affected by a natural disaster impose substantial output loses on their customers. Carvalho et al. (2016) show that the Great East Japan Earthquake shock propagated upstream and downstream through supply chains.

drop in productivity. Indeed, the planner reorganizes the network precisely so that the impact of the bad shock is as small as possible. Even though it leaves output mostly unaffected, the reorganization of the network can have a large impact on the distributions of firm-level variables. In this example, for instance, the dispersions in labor productivity, output and employment collapse after the shock. A negligible shock can therefore have a large impact on cross-producer distributions.

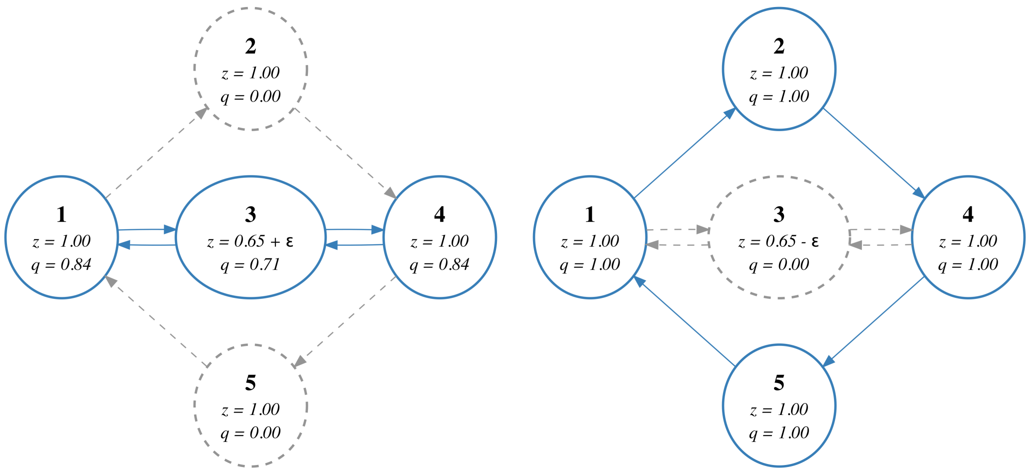


Figure 4: A small shock triggers a large reorganization of the economy

4 Quantitative exploration

To further investigate the importance of the economic forces at work in the environment, I now calibrate the model to the United States economy. Using this calibrated model, I first investigate how the optimal design of the network influences the amount of clustering between firms as well as the degree distributions and the distributions of firm-level outcomes. I then examine how cascades of shutdowns arise and spread in this economy and compare the results to the data. Finally, I investigate how the shape of the network varies over the business cycle and how, in turn, these changes affect aggregate fluctuations.

4.1 Network data

I use two datasets that document supply-chain relationships at the firm level.

Compustat The first dataset is the Compustat Segment database which covers public firms in the United States since 1976. Compustat gathers data about firms' major customers, defined as buyers of more than 10% of their sales, from annual financial statements. This reporting is mandated by an accounting rule: Financial Accounting Standards No. 131. One limitation of these data is that the customers are self-reported by the suppliers, so that General Motors might enter the database as "General Motors", "GM", "General Mtrs", etc. To address this issue, Cohen and Frazzini (2008) (CF) and Atalay et al. (2011) (AHRS), have used a combination of automatic algorithms and manual matching to properly identify each firm and then construct the production network. Their samples cover the periods 1980 to 2004 and 1976 to 2009, respectively. Since the two papers rely on different techniques to build the networks, I use both datasets to verify the predictions of the theory.[19]

Factset Revere I also use a second dataset, the Factset Revere Supply Chain Relationships Data, which covers the period from 2003 to 2016.[20] These data are gathered by analysts from a variety of sources such as 10-K and 10-Q fillings, annual reports, investor presentations, websites, press releases, etc. While it spans a shorter period than Compustat, Factset also includes private firms and less important relationships. It is therefore much more comprehensive such that, in an average year, it includes almost 13,000 firms and more than 46,000 relationships. In contrast, the Atalay et al. (2011) version of the Compustat dataset includes about 1,300 firms and 1,500 relationships in an average year.

The Compustat and the Factset datasets complement each other. Since Compustat is available for a longer period of time, I mostly use it when investigating properties of the model that are related to the business cycle. The higher coverage of the Factset dataset makes it ideal to investigate cascades of firm shutdowns.

Figure 5 shows the production network of the U.S. economy in 2016, as described by the Factset data. Each circle represents a firm with the size of the circle increasing in its degree. The darker colors denote firms with higher local clustering coefficient. The largest circles are all household names. In that year, the firm with the largest in-degree is Walmart with 448 suppliers and the firm with the largest out-degree is Microsoft with 332 customers.

We can use the Factset data to evaluate the importance of the firms' decisions to operate in shaping the production network. In this dataset, 70% of production relationships are destroyed when either the supplier or the customer (or both) stops producing.[21] This large number suggests that the firms' operating decisions play a first-order role in shaping input-output relationships, and it highlights the importance of modeling these decisions to properly understand the formation and the evolution of the production network.

[19] I thank the authors for sharing their data. The Compustat Segment data is also used by Kelly et al. (2013) and Wu (2016), among others.

[20] I restrict the sample to links in which at least one partner is located in the U.S.

[21] Appendix A.1 provides additional details about this exercise along with several robustness tests.

Notes: Larger circles represent higher-degree firms. Darker colors denote firms with higher clustering coefficients. Image generated using Gephi with the Yifan Hu layout. Vector image; zoom in for additional details. There are 20,702 firms and 62,474 links. Some firms at the outskirt are cropped out.

Figure 5: 2016 Factset Revere U.S. firm-level production network

4.2 Parametrization

The model is parametrized at an annual frequency and I normalize $A = 1$ and $L = 1$. For the share of intermediate goods in the production function, I follow Jorgenson et al. (1987) and set $\alpha = 0.5$. Jones (2011) surveys the literature on the share of intermediate goods in different countries and also suggests $\alpha = 0.5$ as a benchmark.

I assume that the log of the productivities z_{it} are drawn from independent AR(1) processes with persistence ρ_z and standard deviation of the ergodic distribution σ_z. Many studies document a large dispersion in producer productivity. For instance, Bartelsman et al. (2013) measure the dispersion in firm-level physical productivity in a number of countries and find $\sigma_z = 0.39$ for the U.S. I adopt this number for the calibration. Importantly for our purpose, their estimation technique controls for the usage of intermediate inputs. For the persistence, I follow Lucia Foster and Syverson (2008) and set $\rho_z = 0.81$.

There is no consensus in the literature about the cost of overhead labor f. Since employment in management occupations accounts for about 5% of total employment in the U.S., I set f so that $f \times n = 5\%$.[22] I set $n = 1000$ as a good trade-off between realism and computation time.[23]

In the data, the observed distributions of in- and out-degree roughly follow power laws. I therefore assume that the joint distribution of in- and out-degree of the *unobserved* set of potential links Ω follows a bivariate power law of the first kind and I set its shape parameter to $\xi = 1.85$ to roughly match the exponent of the *observed* in-degree distributions, averaged across datasets.[24]

Broda and Weinstein (2006) use disaggregated trade data for the U.S. to estimate the elasticity of substitution between product varieties. Their estimates vary greatly across sectors, levels of aggregation, as well as time periods. At the highest level of disaggregation (10-digit level), the mean estimate for the 1990-2001 period is 12.6 while the median is 3.1. At the 5-digit level, the mean estimate is 6.6 while the median is 2.7. I set $\sigma = 5$ as an average of these estimates. The empirical literature provides little guidance about the elasticity of substitution between intermediate inputs at the firm level. I therefore set $\sigma = \varepsilon$ as default and discuss alternative values for ε along the way.

Table 3 displays the parameters of the calibrated economy. So that the results do not hinge on one particular matrix Ω, I draw 20 different Ω's with the properties described above and, for each of them, simulate the economy for 100 periods.[25] The results of this section are averages over these 20 matrices.

Calibrated economy

Table 4 shows how the production network in the calibrated economy compares to the U.S. data. The first two rows show the exponents for the in-degree and out-degree distributions and the

[22] These fixed costs have often been estimated using models of firm turnover. For instance, Bartelsman et al. (2013) estimate a model of firm entry and exit on the manufacturing sector and estimate that overhead labor accounts for 14% of total employment in the industry. This number is likely to be lower for firms outside of manufacturing. Higher fixed costs do not affect the qualitative findings of this section but require solving economies with lager n, so that the number of active firms is reasonable.

[23] Experimentations with $n = 10,000$ firms show that the main results remain in larger economies.

[24] The joint density function is $f(x_{in}, x_{out}) = \xi(\xi - 1)(x_{in} + x_{out} - 1)^{-(\xi+1)}$ The marginal distributions follow power laws with exponents ξ. Generating Ω in that way creates a correlation between the observed in- and out-degrees of the firms of 0.67. Not far from the 0.43 correlation observed in the Factset data. An earlier version of this paper assumed that the in- and out-degree distributions of Ω were uncorrelated with similar results. I do not target the exponent of the out-degree distribution since data limitations in the Compustat dataset likely biases this number, as shown below.

[25] I discard and redraw simulations for which iterating on the first-order conditions does not converge to a point θ such that $\theta_j \in \{0, 1\}$ for all j. This rarely happens and, overall, the rejected networks do not seem different than the accepted ones. The algorithm sometimes fail to converge when Ω contains a firm that is particularly isolated from the rest of the network.

Table 3: Parameters of the calibrated economy

Parameter	Value
Time period	1 year
Average productivity	$A = 1$
Labor supply	$L = 1$
Number of firms	$n = 1000$
Intermediate goods intensity	$\alpha = 0.5$
Elasticity of substitution for final goods	$\sigma = 5$
Elasticity of substitution for intermediate goods	$\varepsilon = 5$
Standard deviation of individual productivities	$\sigma_z = 0.39$
Fixed cost of operation	$f = 0.05/n$
Shape of Ω	$\xi = 1.85$

third row shows the global clustering coefficient of each network.[26,27] The model fits the Factset data relatively well but there are some discrepancies with the Compustat datasets. Given that Factset provides the most comprehensive data, the large out-degree power law exponent and the low clustering coefficient in the Compustat data are likely to arise from its limitations. In particular, the truncation of customers purchasing less than 10% of sales artificially limits the out-degree of a firm and removes several links from important suppliers. As a result, the out-degree exponent is likely to be biased upwards and the clustering coefficient to be biased downwards.

Table 4: Production network in the calibrated economy and the data

	Model	Data		
		Factset	AHRS	CF
Power law exponents				
In-degree distribution	1.07	0.95	1.13	1.32
Out-degree distribution	1.02	0.81	2.24	2.22
Global clustering coefficient	0.51	0.64	0.013	0.014

Notes: Power law exponents are estimated using the approach of Gabaix and Ibragimov (2011). Global clustering coefficients are multiplied by the square roots of the number of nodes. See footnote 26 for details.

Figure 6 shows the degree distributions in the model and in the Factset data for 2016, the most recent year in the sample. To properly highlight the shape of these distributions, the figure

[26] All power law exponents are estimated using the approach of Gabaix and Ibragimov (2011). The global clustering coefficient equals three times the number of triangles (three *fully* connected nodes) divided by the number of triplets (three connected nodes). In power law graphs, the global clustering coefficient declines naturally with n. Following Ostroumova Prokhorenkova and Samosvat (2014), I therefore normalize the global clustering coefficients in Table 4 by multiplying them by the square root of the number of nodes. This normalization allows for a better comparison of networks of different sizes. It has no effect on the qualitative results of the paper.

[27] Another important measure often used in the literature is that of the centrality of a node. In the exercises of this section, the moments related to out-degree centrality are very similar to those related directly to the out-degree of a node.

uses a log-log scale and plots the complementary cumulative distributions (CCDF) on the vertical axis. The roughly linear shape of the distributions confirms that they are close to power laws. The model fits the in-degree distribution well but must approximate the out-degree distribution given the departure from the power law for high out-degree firms in the data.

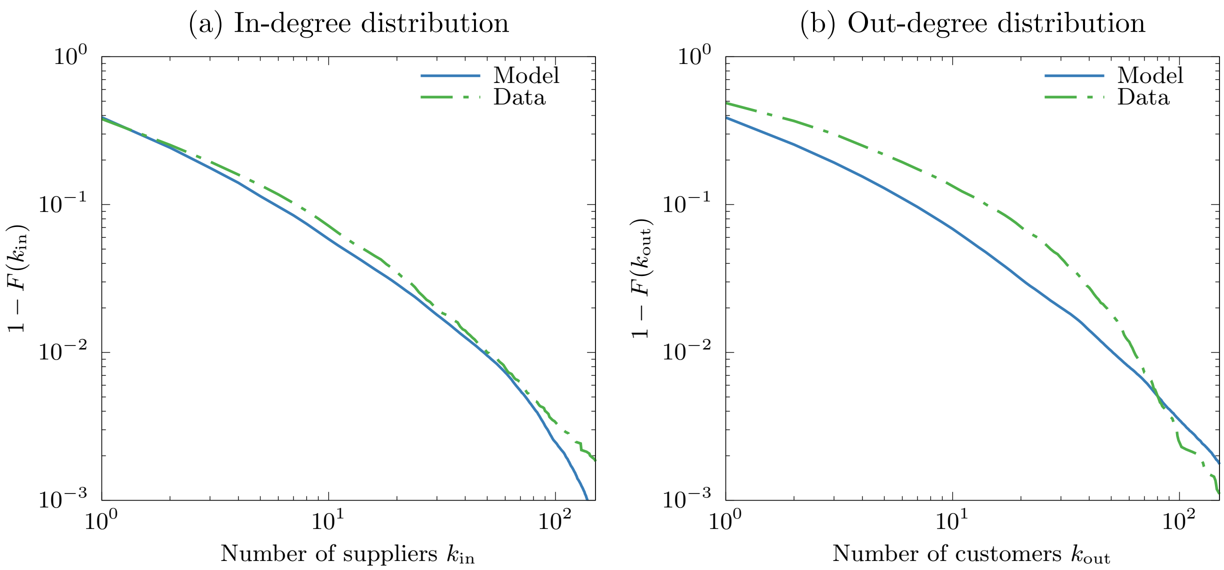


Figure 6: Distribution of the number of suppliers and customers

4.3 Comparison with a random network

To highlight what features of a network are more desirable for efficiency, it is useful to compare the *optimal* network, the one that arises in the efficient allocation, to a completely *random* one. This random network is built by operating each firm with some probability p, and by adjusting p such that the realized number of operating firms is the same as in the optimal allocation. Except for the operating decision θ, all other quantities are chosen optimally.

Table 5 shows the power law exponents of the in-degree and out-degree distributions, as well as the clustering coefficient, in the random and the optimal networks. We see that the power law exponents are smaller in the optimal network, indicating that these distributions have fatter tails than in the random network. The optimal network therefore features a larger fraction of highly connected suppliers and customers than the random network. The clustering coefficient is also larger in the optimal network. These moments highlight the planner's preferred way of organizing production: it creates tightly connected clusters of economic activity centered around firms with many connections. By building the network in this way, the planner takes full advantage of the production complementarities present in the environment.

To further understand how economic forces endogenously shape the production network, it is useful to consider the influence of certain parameters on the degree distributions and the clustering coefficient. Here, the key parameters are the elasticities of substitution as well as the share of intermediate inputs into production. For instance, a higher σ concentrates the economy towards the most productive firms which leads to degree distributions with fatter tails and higher clustering. A higher α, by increasing the importance of the network in the production process, has a similar effect.

Table 5: Shape of the optimal and random networks

	Model	
	Optimal network	Random network
Power law exponents		
In-degree distribution	1.07	1.22
Out-degree distribution	1.02	1.21
Global clustering coefficient	0.51	0.30

Notes: Power law exponents are estimated using the approach of Gabaix and Ibragimov (2011). Global clustering coefficients are multiplied by the square roots of the number of nodes. See footnote 26 for details.

4.4 Firm-level outcomes

The theory has several implications for firm-level outcomes. First, firms with many neighbors employ more labor and produce more. This can be seen in Table 6 which shows the correlation between a firm's in- and out-degree, and its employment and sales. All correlations in the simulated economy are positive. This finding is not surprising in view of equation (9). Firms with many suppliers are, overall, more efficient (high q/Q) and the planner takes advantage of these high productivities by providing these firms with more labor. Firms with many customers are also allocated more labor to fulfill their role as important suppliers in the production network (term Γ in the equation). Table 6 shows that the correlations are also positive in the data which, given the origin of the correlations in the theory, reinforces our choice of a production function with gains from input variety.[28]

The endogenous formation of the network also has implications for the distribution of firm-level outcomes. Two main forces are at work here. First, the planner tends to shut down firms with low labor productivity q. As a result of this *selection* mechanism, firms with low total factor productivity z, or those that are poorly located in the network, do not operate. Second, the planner takes advantage of the complementarity between nearby firms to *magnify* the impact of important producers on the network. In particular, the planner builds clusters of suppliers and customers

[28] Saito et al. (2007) and Bernard et al. (2015) find that firms with many suppliers and many customers are also larger in terms of sales in the Japanese data.

Table 6: Firm-level correlations

	Model		Data	
	Employment	Sales	Employment	Sales
In-degree	0.42	0.42	0.47	0.51
Out-degree	0.43	0.43	0.11	0.16

Notes: All variables are in logs. Sales in the model are y/q and employment is l. The data columns refer to the Factset network data merged with the Compustat sales and employment data. All correlations are also positive in the AHRS network data.

around these producers. As a result, these high-performing firms become even more productive and have a disproportionate effect on the economy.

Table 7 shows the mean, the standard deviation and the skewness of the labor productivity and the employment distributions in the optimal network economy. To properly evaluate the importance of the selection and magnification mechanisms, the table also shows these quantities in the random network economy, where they are inactive. Unsurprisingly, by comparing the two economies, we see that the average firm has a higher productivity q in the optimal network economy while it employs the same amount of labor.[29] Here, the two forces operate together, by truncating the bottoms of the distributions and pushing the firms at the top further to the right, to generate this outcome.

The two forces push, however, in different directions to impact the standard deviations of the distributions. We see from Table 7 that the selection of operating firms, by compressing the distributions from the left, dominates so that the optimal network economy features less dispersed productivity and employment distributions.

The selection and magnification mechanisms also work hand in hand to generate substantial positive skewness in the optimal network distributions. The theory therefore provides an explanation for the fatter right tail of the firm-size distribution, a feature that leads microeconomic shocks to have a meaningful impact on the aggregate economy (Gabaix, 2011; Acemoglu et al., 2012).

Table 7: Moments of firm-level outcomes in the optimal and random network economies

	Optimal network economy		Random network economy	
	Productivity q	Employment l	Productivity q	Employment l
Mean	2.40	0.0018	1.75	0.0018
Std. dev.	0.30	1.26	0.46	1.94
Skewness	0.53	1.04	-0.05	0.05

Notes: To better highlight the asymmetry generated by the model, the standard deviations and the skewness are computed on the log of q and l. The means are computed on q and l directly.

[29] The random network economy has, by construction, the same number of active firms as the optimal network economy. As a result, the amount of productive labor $L - f\sum_j \theta_j$ per firm must be the same in both economies.

4.5 Cascades of firm shutdowns

Because of the complementarities at work in the environment, a severe productivity shock to an important firm in the network can trigger a cascade of firm shutdowns. Indeed, if productivity falls too much, the planner prefers to shut the firm down and to reallocate the labor from the fixed cost to more productive parts of the network. The neighbors of the firm than lose a customer or a supplier, which lowers their contribution to aggregate output, and might lead to them shutting down as well. The same reasoning applies to the firm's second neighbors and so on. As this cascade of shutdowns propagates through the network it can affect several firms in the economy.

To evaluate the importance of these cascades, I regress the fraction of each firm's neighbors that shuts down in a given period on whether the firm itself shuts down.[30] To allow for a better description of these cascades, I run separate regressions for upstream and downstream neighbors at various distances for the original firm.[31] The coefficients estimated from these regressions capture the increase in shutdown probability associated with an exiting neighbor.

Figure 7 shows the estimated coefficients in the model (solid blue lines) and in the Factset data (green dashed lines). We see that, in the data, the shutdown of a firm is associated with about a 10% increase in the probability that its direct customers also exits. This increase in probability falls down to about 2% for the second customers and keeps declining for further neighbors. Upstream, a firm shutdown is also associated with about a 10% increase in the probability that its first suppliers shut down as well. The model is roughly able to match the shape and the magnitude of the cascades suggesting that it properly captures the joint operating decisions of nearby firms.

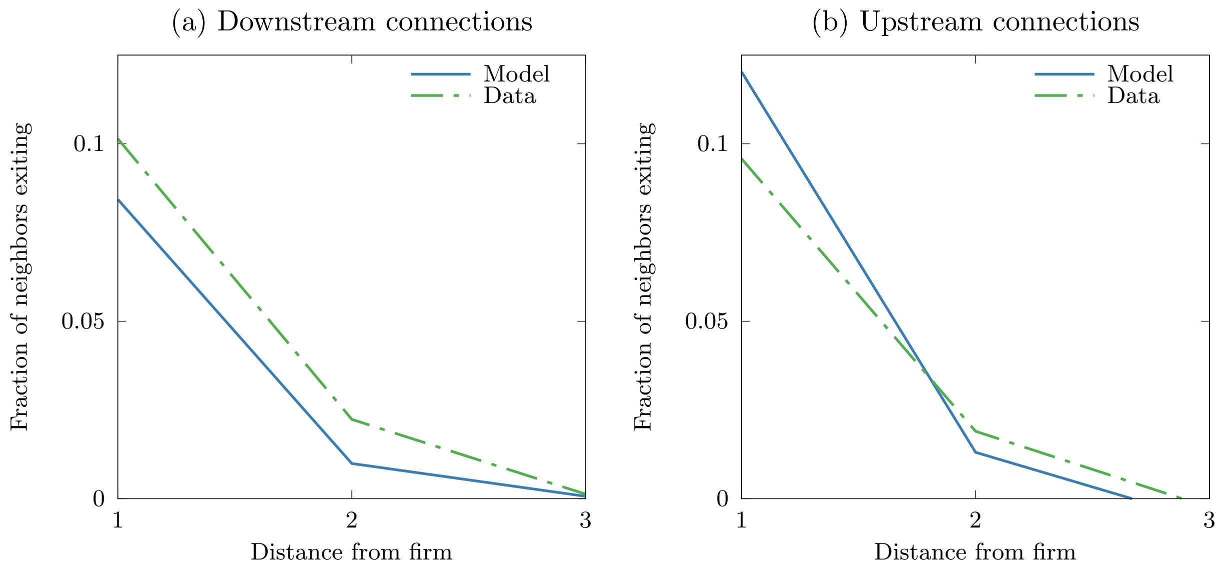
These cascades of firm shutdowns are akin to a process of creative destruction. When shutting down a cluster of poorly performing firms, the planner uses the saved labor to start the operation of firms elsewhere in the network. This process, can explain the negative coefficients attached to the third upstream neighbors in Figure 7. These negative numbers, which are statistically different from zero, indicate that the planner is more likely to keep operating distant suppliers of a firm that exits.

While cascades can arise from any firm, the shutdown of a producer with many neighbors is more likely to trigger a large reorganization of the network. To highlight this fact, the first two columns of Table 8 show the total number of shutdowns, summed up to the third neighbors upstream and downstream, associated with the exit of firms with average and with high degrees, where the latter is defined as above the 90^{th} percentile of the degree distribution.[32] We see that, in the model and

[30] In the data, I consider that a firm shuts down if it drops from the sample. Mergers and acquisition are therefore counted as shut downs. While this misclassification is unfortunate, it leads to an *underestimation* of the the impact of cascades. As a result, the results presented in this section can be interpreted as lower bounds.

[31] I also include time indicator variables in the regressions to control for potential aggregate shocks in the data. Control variable for the in-degree and the out-degree of a firm are also included to mitigate selection issues.

[32] This measure is computed by multiplying the regression coefficients by the average number of neighbors at a given distance and summing, downstream and upstream, up to the third neighbors.



Notes: Factset data. Estimated coefficients from regressing the fraction of exiting neighbors on whether a firm exits. Time indicator variables as well as in-degree and out-degree controls are included as regressors. The distance from firm measures the shortest path, in terms of numbers of connections, from the original firm.

Figure 7: Cascades of firm shutdowns in the model and in the data

in the data, the cascades triggered by high degree firms are much larger than those triggered by the average firm.

While high degree firms are likely to create bigger cascades upon shutting down, they are also less likely to actually shut down. The last two columns of Table 8 show the probability that a firm shuts down based on its degree. In the data, an average firm has about a 12% chance of shutting down in a given year. The same number is much lower if we consider only high degree firms, in which case the probability drops to 3.4%. The model is roughly able to match these numbers but it overestimates the exit probability of the average firm and underestimates the size of the cascades.

Table 8: Impact of firm degree on origin and size of cascades

	Size of cascades		Probability of exit	
	Model	Data	Model	Data
Average firm	0.4	0.9	16.3%	12.2%
High degree firm	5.9	7.9	2.3%	3.4%

Notes: "High degree firms" refers to the 90$^{\text{th}}$ percentile of the degree distribution. "Size of cascades" corresponds to the sum of exiting firms up to the third neighbors downstream and upstream.

That highly connected firms are less likely to stop production is not surprising in view of the theory. These firms, by their advantageous position in the network, are particularly valuable to the planner, and they are therefore kept in operation even after a low productivity shock. When they do shut down, however, the planner is likely to reorganize the whole cluster of producers that was built around them, which explains the importance of the cascades they trigger.

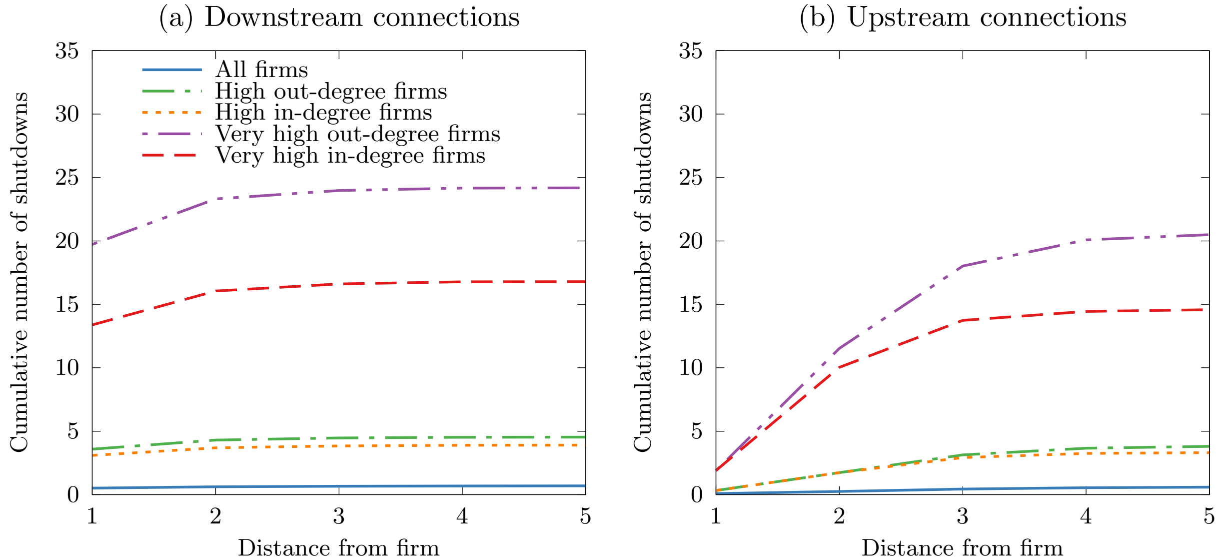
While the regression coefficients of Figure 7 are useful to highlight the co-movement between nearby firms — and they also have the advantage that they can be computed in the model and in the data — they do not inform us directly about the *causal* impact of the shutdown of a firm on its neighbors. Thankfully, we can use the model to get that information through the following exercise. Select a random firm in the calibrated economy and set its productivity to $z = 0$ to shut it down. Then compute the new efficient allocation and count the neighbors of the firm that have shut down. Figure 8 shows the outcome of this exercise. The left panel considers the downstream neighbors of the firm, and the vertical axis shows the cumulative number of these neighbors that shut down as we move further and further away from the shuttered firm. The right panel is similar but considers upstream neighbors. The figure also differentiates between cascades originating from the average firm or from firms with high (above 90^{th} percentile) or very high (above 99^{th} percentile) number of neighbors. We see that the average firm is likely to create very small cascades: about 0.6 of the firm's downstream neighbors, and even fewer of its upstream neighbors, shut down. As we move to high degree and very high degree firms, however, the cascades become much bigger: for very high out-degree firms above 24 downstream neighbors are swiped by the cascade and the economy is extensively reorganized. There is also a positive correlation between the size of a cascade and its negative impact on output. While the average cascade has a negligible effect on macroeconomic aggregates, a cascade that originates from a very high degree firm is responsible, on average, for a 1.9 percent drop in output. This large number highlights the importance of highly connected firms for economic fluctuations.

Figure 8 also shows that cascades mostly propagate downstream, from customer to customer, instead of upstream: a consequence of the mechanisms embedded in equation (7).[33]

In the model, the resilience of firms, the magnitude of the cascades and their direction of propagation depend critically on the elasticities of substitution σ and ε, as well as on the importance of intermediate inputs in the production function α. Table 11 in Appendix A shows how the exit probabilities evolve when these parameters change. In particular, a lower ε makes the network more rigid and firms are less likely to shut down. Alternatively, a higher elasticity σ, since it makes inputs in the production of the final good more substitutable, leads to higher likelihoods of shutting down. In terms of the size and the propagation of the cascades, Figure 9 in the appendix shows the cascades that arise when $\varepsilon = 3$ and how they propagate. In contrast to the benchmark economy, they are much bigger and they also have substantial upstream propagation.

Cascades in this environment are fully efficient. They are the manifestation of the optimal reorganization of the network in response to shocks. As such, bailing out firms is inherently welfare reducing. In fact, for the average firm, a bailout increases the contraction in aggregate output

[33] Many shutdowns that happen upstream in Figure 8 are actually about firms that are closer downstream to the original firm.



Notes: Cumulative sum of exiting firms. 10,000 simulations on 100 different matrices Ω. "High degree" refers to the 90^{th} percentile of the distribution. "Very high degree" refers to the 99^{th} percentile.

Figure 8: Cumulative cascades by type of originator

created by the shock by about 40%. However, if a policymaker would like, for whatever reason, to prevent large cascades from arising, the results of this section suggests to pay particular attention to firms with many neighbors, and that evolve in industries that rely heavily on specialized intermediate inputs for which good substitutes are unavailable.[34]

4.6 Aggregate fluctuations

We now consider how the network structure of production and aggregate fluctuations interact in this economy. The first part of this section investigates how the shape of the network evolves with the business cycle. To do so, it documents several correlations that appear in the simulated economy and in the data. The second part of the section considers how reorganizing the network in response to shocks affects the size of aggregate fluctuations.

Comovements

Table 9 shows the correlation between aggregate output and several moments that characterize the shape of the network. As can be seen in the first column, which reports these correlation in the simulated data, the clustering coefficient is pro-cyclical and the tails of the in- and out-degree distributions become fatter during expansions.[35] This cyclical evolution of the production network

[34]This result is reminiscent of work by Barrot and Sauvagnat (2016) that finds that a shock to a supplier has a stronger impact on a customer if the traded input is highly specific. See Baqaee (2016) for an environment with inefficient cascades.

[35]Carvalho and Grassi (2017) find that the tail of the firm size distribution is also fatter during expansions than during recessions in the U.S. data.

suggests that booms, in this economy, are periods with a relative abundance of highly connected firms surrounded by tight clusters of producers. To understand why the network evolves in this way, consider the origin of fluctuations in this economy. Since the model does not feature aggregate shocks, expansions are mostly driven by positive idiosyncratic shocks to influential firms in the network — those in the right tail of the degree distributions — and by the ability of the planner to cluster production around them. Inversely, recessions are periods in which it is too costly to organize these productive clusters — perhaps because a few important firms face low z's — and economic activity is therefore more dispersed through the network.

These correlations are also visible in the U.S. data, as can be seen in the last three columns of Table 9. In Factset, as well as in the two Compustat databases, the fatness of the tails of the degree distributions and the clustering coefficient are all pro-cyclical.

Table 9: Correlations between aggregate output and the production network

	Model	Data		
		Factset	AHRS	CF
Power law exponents				
In-degree distribution	-0.57	-0.85	-0.35	-0.12
Out-degree distribution	-0.67	-0.94	-0.30	-0.11
Global clustering coefficient	0.46	0.68	0.17	0.20

Notes: All time series are in logs. In the data, output is annual real gross domestic product. Output is detrended linearly in sample. Since there are only 13 years in the Factset database I use the CBO 10-year projection for real GDP growth at the beginning of the sample in 2003 (2.58%) to detrend the series. Global clustering coefficients are multiplied by the square roots of the number of nodes. See footnote 26 for details.

The results presented in this section so far provide information about the ability of the model to capture important features of the data that relate to the evolution of the network. Tables 8 and 9 are particularly useful for that purpose. Indeed, Table 8 shows that the operation decisions of firms, the margin through which the network evolves in the theory, are similar in the model and in the data. In addition, we see from Table 9 that the impact of these decisions on the shape of the network is also roughly consistent with the data. Overall, these findings therefore show that the endogenous network theory developed in this paper, while simple, is able to capture features of the data that are key to understand the relationship between the evolution of the production network and aggregate fluctuations.

Size of fluctuations

The ability of the network to reorganize itself is critical in determining how microeconomic shocks aggregate into macroeconomic fluctuations. To highlight this fact, it is convenient to compare the optimal allocation, in which the network is efficiently reorganized in response to shocks, to an alternative economy in which the network is drawn randomly and then kept fixed. Table

10 shows the standard deviation of output Y and its two components, labor productivity Q and productive labor $L - f\sum_j \theta_j$, in these two economies. We see that the standard deviation of output is about 30% larger in the economy with a fixed network.[36] To pinpoint the origin of this difference, the last two columns of the table show that the variation in productive labor accounts for a negligible part of the fluctuations in Y and that, instead, the productivity Q of the network is responsible for the bulk of the difference. The changes in output are therefore driven not by the number of firms that operate but by the way these firms are organized in the production network. We can conclude from these results that the ability of the production network to respond to idiosyncratic shocks has an important impact on the size of macroeconomic fluctuations.

To understand why fluctuations are smaller in the optimal network, it helps to think of the planner as picking the best network out of the 2^n possibilities (recall that a network here is characterized by a vector θ). Each of these network is essentially a function Y_k that maps a given realization of z into some aggregate output. We can therefore write total production in this economy as

$$Y(z) = \max_{k \in \{1,\ldots,2^n\}} \{Y_k(z)\}$$

where the notation emphasizes that the chosen optimal network depends on the stochastic realization of z. Now, by itself, each network k is associated with a distribution for Y_k. The means and the variances of these distributions vary with k but, for the networks that are actually selected by the planner, the differences are limited. As a result, these distributions have a substantial overlap with one another such that when the planner selects the best network k^* for a given z, the associated optimal output Y tends to be in the right tail of the Y_{k^*} distribution, but also in the right tail of all the other Y_k distributions. Repeating this process for all z's, the whole distribution of Y will therefore be concentrated in the right tails of the Y_k distributions, which leads to a smaller variance for Y than for each individual Y_k.[37]

Table 10: Standard deviation of aggregates in the optimal and fixed network economies

	Output Y	$\approx$	Labor Prod. Q	$+$	Prod. labor $L - f\sum_j \theta_j$
Optimal network	0.10		0.10		0.009
Fixed network	0.13		0.13		0

Notes: All variables are in log. In the fixed network economy, the standard deviation of productive labor is 0 by construction.

The importance of the reduction in aggregate fluctuations generated by the reorganization of

[36] Unsurprisingly, the planner can also reach a higher level of output by adjusting the shape of the network in response to shocks. The average difference between output in the optimal network and the random network is about 31%, suggesting that impeding the reorganization of the network has a substantial negative effect on welfare.

[37] This intuition is reminiscent from results from extreme value theory that show, for instance, that the variance of the maximum of a large number m of independent standard normal random variables declines rapidly with m.

the network depends on several features of the environment. For instance, in an economy with a lower elasticity of substitution between inputs ε, the number of suppliers available to a producer has more importance and the shape of the network is mostly determined by the matrix Ω with little influence from productivity. As a result, the network reacts less to shocks to z and the reduction in the variance of output generated by the reorganization of the network is less important. To the contrary, in industries were intermediate inputs are highly substitutable, any friction in the reorganization of input-output relationships could have important consequences on the variability of the industry's production.

Overall, the results of this section have several implications for our understanding of business cycles and for the design of policy. First, in this economy, the shape of the network does not simply react to aggregate fluctuations but its evolution plays an integral role in actually shaping them. As a result, understanding how the production network responds to shocks is necessary to properly understand the microeconomic origin of macroeconomic fluctuations. Second, policies that impede the reorganization of the network might lead not only to a loss in output but also to an increase in its variance. Too-big-to-fail interventions might therefore inadvertently exacerbate business cycle fluctuations.

5 Conclusion

This paper provides a first attempt at explaining the joint dynamic of the production network and the business cycle. To do so, it proposes a theory of endogenous network formation and aggregate fluctuations. Because of a complementarity between firms' operating decisions, production tends to be organized in tightly connected clusters of producers centered around the most productive firms. The complementarities also give rise to cascades of firm shutdowns. As in the data, highly connected firms are more resilient to shocks but trigger larger cascades upon shutdown. The theory also predicts how the shape of the network evolves with aggregate output. In particular, expansions feature more clustering among firms as well as fatter tails for the degree distributions. These correlations are also present in the U.S. data. Finally, the theory predicts that the optimal reorganization of the network in response to shocks is responsible for a substantial decline in the size of aggregate fluctuations. These findings highlight the importance of considering the evolution of the production network to further understand the origin of aggregate fluctuations.

One contribution of the paper is to provide a novel reshaping approach to solve optimization problems with discrete adjustment margins and complementarities. Using this approach, it is possible to quickly find the optimal allocation in economies that were previously impossible to solve. This new tool could potentially be useful in a broad range of economic environments, including models with menu costs, or fixed costs of investment or hiring.

The theory also highlights many interesting future research avenues. One such project could study industry clusters such as Silicon Valley for technology, and New York and London for finance. In particular, the theory could shed light on how the appearance or the disappearance of influential firms affect other participants in the cluster and therefore shed light on how these clusters emerge in the first place. Another possible research project would involve documenting if the predictions of the theory regarding the link between flexibility of the network and the parameters of the production function hold in the cross section of industries. The theory also highlights several important mechanisms that should be considered by policy makers when debating whether to bailout distressed firms. As such, the model proposes a framework that could be extended to provide a more quantitative assessment of the benefits of such a bailout. Finally, it would be interesting to investigate whether the prediction of the theory that small changes in fundamentals can have large impacts on the allocation has support in the data.

For the sake of clarity and tractability, the mechanisms of the model were exposited in a simple framework but several extensions may be worth investigating. In particular, it would be interesting to understand how requiring the payment of a fixed cost to *change* the operating status of a firm would affect the main mechanism of the model. In general, considering the formation of production networks when firms are subject to dynamic concerns seems like a promising, if challenging, future avenue for research.

References

Acemoglu, D. and P. D. Azar (2017): "Endogenous Production Networks," Working Paper 24116, National Bureau of Economic Research.

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

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015): "Microeconomic Origins of Macroeconomic Tail Risks," Working Paper 20865, National Bureau of Economic Research.

Atalay, E. (2017): "How Important Are Sectoral Shocks?" *American Economic Journal: Macroeconomics*, 9, 254–80.

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

Bak, P., K. Chen, M. Woodford, and J. Scheinkman (1993): "Aggregate Fluctuations from Independent Sectoral Shocks: Self-Organized Criticality in a Model of Production and Inventory Dynamics," *Ricerche Economiche*, 47, 3–30.

Baqaee, D. R. (2016): "Cascading Failures in Production Networks," working paper.

Baqaee, D. R. and E. Farhi (2017a): "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," Working Paper 23145, National Bureau of Economic Research.

——— (2017b): "Productivity and Misallocation in General Equilibrium." Working Paper 24007, National Bureau of Economic Research.

Barrot, J.-N. and J. Sauvagnat (2016): "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks," *The Quarterly Journal of Economics*, 131, 1543–1592.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013): "Cross-Country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, 103, 305–34.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2015): "Production Networks, Geography and Firm Performance," Working Paper 21082, National Bureau of Economic Research.

Bigio, S. and J. La'O (2016): "Financial Frictions in Production Networks," Working Paper 22212, National Bureau of Economic Research.

Blume, L., D. Easley, J. Kleinberg, R. Kleinberg, and E. Tardos (2011a): "Network formation in the presence of contagious risk," in *Proceedings of the 12th ACM conference on Electronic commerce*, ACM, 1–10.

Blume, L., D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos (2011b): "Which networks are least susceptible to cascading failures?" in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, IEEE, 393–402.

Bresnahan, T. F. and V. A. Ramey (1994): "Output Fluctuations at the Plant Level," *The Quarterly Journal of Economics*, 593–624.

Broda, C. and D. E. Weinstein (2006): "Globalization and the Gains From Variety," *The Quarterly Journal of Economics*, 121, 541–585.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017a): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," Working paper.

Caliendo, L., F. Parro, and A. Tsyvinski (2017b): "Distortions and the Structure of the World Economy," Working Paper 23332, National Bureau of Economic Research.

Carvalho, V. M., , M. Nirei, and Y. Saito (2014): "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," Working Paper 14-E-035, The Research Institute of Economy, Trade and Industry.

Carvalho, V. M. (2014): "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, 28, 23–48.

Carvalho, V. M. and B. Grassi (2017): "Large Firm Dynamics and the Business Cycle," Working paper.

Carvalho, V. M., M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi (2016): "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake," Working paper.

Carvalho, V. M. and N. Voigtländer (2014): "Input Diffusion and the Evolution of Production Networks," Working Paper 20025, National Bureau of Economic Research.

Cohen, L. and A. Frazzini (2008): "Economic Links and Predictable Returns," *The Journal of Finance*, 63, 1977–2011.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82, 1303–1340.

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

Gabaix, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 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, 24–39.

Garey, M. R. and D. S. Johnson (1990): *Computers and Intractability; A Guide to the Theory of NP-Completeness*, New York, NY, USA: W. H. Freeman & Co.

Grassi, B. (2017): "IO in I-O: Size, Industrial Organization and the Input-Output Network Make a Firm Structurally Important," Working paper.

Hendricks, K. B. and V. R. Singhal (2005): "An Empirical Analysis of the Effect of Supply Chain Disruptions on Long-Run Stock Price Performance and Equity Risk of the Firm," *Production and Operations Management*, 14, 35–52.

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

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

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

JORGENSON, D. W., F. GOLLOP, AND B. FRAUMENI (1987): *Productivity and U.S. Economic Growth*, Harvard University Press, reprinted- Universe, 1999.

KARP, R. M. (1972): *Reducibility among Combinatorial Problems*, Boston, MA: Springer US, 85–103.

KELLY, B., H. LUSTIG, AND S. V. NIEUWERBURGH (2013): "Firm Volatility in Granular Networks," working paper.

KENNAN, J. (2001): "Uniqueness of Positive Fixed Points for Increasing Concave Functions on R^n: An Elementary Result," *Review of Economic Dynamics*, 4, 893 – 899.

LIM, K. (2016): "Firm-to-firm Trade in Sticky Production Networks," working paper.

LONG, J. B. AND C. I. PLOSSER (1983): "Real Business Cycles," *Journal of Political Economy*, 91, 39–69.

LUCIA FOSTER, J. H. AND C. SYVERSON (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *The American Economic Review*, 98, 394–425.

OBERFIELD, E. (2017): "A Theory of Input-Output Architecture," *Working Paper.*

OSTROUMOVA PROKHORENKOVA, L. AND E. SAMOSVAT (2014): *Global Clustering Coefficient in Scale-Free Networks*, Cham: Springer International Publishing, 47–58.

OZDAGLI, A. AND M. WEBER (2017): "Monetary Policy through Production Networks: Evidence from the Stock Market," Working paper.

RAMÍREZ, C. (2017): "Inter-firm relationships and asset prices," Working paper.

SAITO, Y. U., T. WATANABE, AND M. IWAMURA (2007): "Do larger firms have more interfirm relationships?" *Physica A: Statistical Mechanics and its Applications*, 383, 158 – 163.

WAGNER, S. M. AND C. BODE (2008): "An Empirical Examination of Supply Chain Performance Along Several Dimensions of Risk," *Journal of Business Logistics*, 29, 307–325.

WU, D. A. (2016): "Shock Spillover and Financial Response in Supply Chain Networks: Evidence from Firm-Level Data," working paper.

ZURICH INSURANCE GROUP (2015): "Potential business impact of loss of the main supplier for small and medium enterprises (SMEs) in 2015," .

A Additional exercises

A.1 Details and robustness of the link creation and destruction analysis

This section provides additional details about the exercise at the end of Section 4.1 in which I compute the fraction of link destructions that can be attributed to the operating decisions of the firms.

For the purpose of that exercise, I consider that a firm shuts down if it stops being included in the Factset sample. While Factset covers a large number of firms, one worry is that it might remove firms for reasons other than the firm shutting down. I believe that these concerns are of second-order importance for three reasons. First, while the creation margin could be affected by the inclusion of firms at some point in their lifecycle, it is likely that Factset keeps following firms for as long as possible. Indeed, Factset's promotional material takes pride in its coverage of a large number of firms. Second, the results are robust if we only consider publicly traded firms. Factset covers all of these firms in the United States such that they are more likely to be included in the sample until they exit. I find that, in this case, 64% of link destructions can be attributed to the exit of the customer or the supplier. Third, the result are also robust if I limit the sample to the Great Recession, a period with a high exit rate for firms according to the U.S. Census. In this subsample, 71% of link destructions are associated with the exit of either the supplier or the customer. Overall, these robustness exercises reinforce the finding that the operating decisions of firms are responsible for a large fraction of link destructions in the U.S. economy.

The same exercise can be repeated for link creations if we consider that a firm begins production when it joins the sample. In this case, 84% of production relationships in the data are either created or destroyed through the customer and the supplier's operating decisions. It is, however, likely that firms are included in the Factset sample at some point after their creation so that this large number might be more indicative of the high persistence of relationships once they are created.

A.2 Cascades

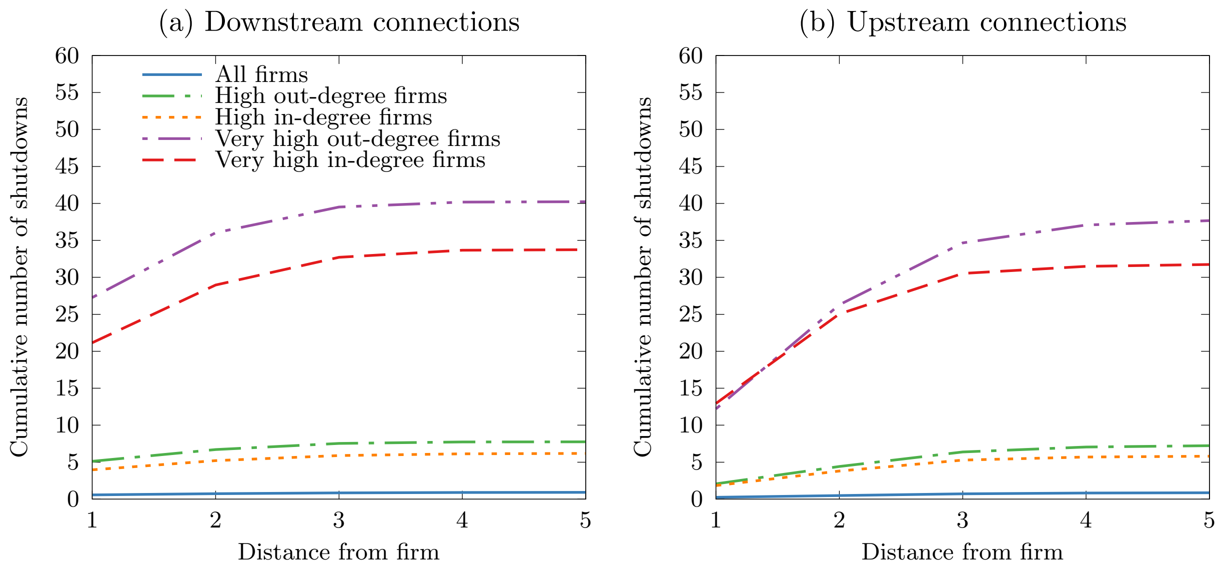
Table 11 shows how the probability that a firm shuts down vary with different parameters. See discussion in main text.

	Probability of exit			
	Benchmark	$\alpha = 0.75$	$\sigma = 7$	$\varepsilon = 4$
Average firm	16.3%	17.0%	25.2%	15.0%
High degree firm	2.3%	1.2%	2.8%	1.0%

Notes: "High degree firms" refers to the 90$^{\text{th}}$ percentile of the degree distribution. "Size of cascades" corresponds to the sum of exiting firms up to the third neighbors downstream and upstream.

Table 11: Probability of exit under different parameters

Figure 9 shows that lowering the elasticity of substitution to $\varepsilon = 3$ affects the cascades in important ways. In particular, shocks to the largest suppliers and customers can now trigger important cascades that have substantial upstream propagation (notice the scale difference between Figures 8 and 9). This stark difference with the benchmark simulations can be explained by understanding why the planner operates small firms in this economy. Since ε is small compared to σ the planner derives utility from them mostly as the providers of intermediate input to the high-q firms and not so much for their direct contribution to final goods production. As a result, if one of the largest customers (a high-q firm) shuts down, the planner is likely to shut its small suppliers down as well, thereby triggering an upstream cascade. In contrast, in the benchmark economy where $\varepsilon = \sigma$ the planner values the direct contribution to the final good from these small providers similarly to their contribution to the productivity of the high-q firms. They are therefore more likely to remain if one of their large customer shuts down.



Notes: Cumulative of exiting firms at distance from originator. "High degree" refers to the 90$^{\text{th}}$ percentile. "Very high degree" refers to the 99$^{\text{th}}$ percentile.

Figure 9: Cumulative cascades by type of originator when $\varepsilon = 3$

A.3 Aggregate Fluctuations

Table 12 shows how different parameters of the model affect the correlations between aggregate output and the shape of the production network. We see from the table that, with a smaller elasticity of substitution between intermediate goods ε or with a higher intensity of intermediate input in the production function α, the business cycle is more highly correlated with the fatness of the tail of the in-degree distribution. To understand why this is so, consider that, through equation (7) a lower ε or a higher α both increase the importance of additional suppliers on the productivity of a firm. As such, a change in the fraction of firms with multiple suppliers has a larger impact on output under these parameters.

Table 12 also shows that the correlation between the global clustering coefficient and aggregate output is relatively constant across parametrizations.

Table 12: Correlations between aggregate output and the shape of the production network

	Benchmark	$\varepsilon = 3$	$\alpha = 0.75$
Power law exponents			
In-degree distribution	-0.57	-0.72	-0.63
Out-degree distribution	-0.67	-0.75	-0.66
Global clustering coefficient	0.46	0.44	0.45

Notes: All time series are in logs. Global clustering coefficients are multiplied by the square roots of the number of nodes. See footnote 26 for details.

Table 13 shows how different parametrizations of the model affect the standard deviation of output. As the main text explains, when the elasticity of substitution between intermediate inputs ε is lower, the number of suppliers available to a producer has more importance and the shape of the network is mostly determined by the matrix Ω with little influence from productivity. As a result, the network reacts less to shocks to z and the reduction in the variance of output generated by the reorganization of the network is less important.

The table also shows that a higher α leads to a large increase in the standard deviation of output which is due to a smaller number of operating firm. Indeed, as α increases, the importance of intermediate inputs into production also increases but, since ε is relatively high, the planner prefers to provide these intermediate inputs using fewer larger firms. The idiosyncratic shocks to z are therefore less likely to average out and the fluctuations in aggregate output are more important. The fixed network economy also sees an increase in the standard deviation of output when α increases since it is constructed to have the same number of firms as the optimal network economy.

Table 13: Standard deviation of output with different parameters

	Benchmark	$\varepsilon = 3$	$\alpha = 0.75$
Optimal network	0.10	0.11	0.30
Fixed network	0.13	0.12	0.36
% difference	+30%	+9%	+20%

Notes: All variables are in log.

B Algorithms to find the efficient allocations

Exhaustive search

This algorithm performs an exhaustive search of the 2^n possible vectors $\theta \in \{0,1\}^n$. The algorithm is as follows:

1. Order all the possible θ, from θ^1 to θ^{2^n}, and initialize the first iteration with θ^1.
2. For the p-th iteration:
   (a) Using θ^p, iterate equation (7) to find the vector q^p.
   (b) Using equation (8) compute the consumption associated with θ^p.
3. Repeat step 2 above for all the possible vectors θ. The one with the highest consumption corresponds to the efficient allocation.

This algorithm is guaranteed to find the global maximum of $\mathcal{P}_{SP}$ but it is impossible to use for large n. In practice, it takes about 1 minute for $n = 17$ to check the 131072 possible vectors θ on a 4 GHz Intel Core i7 processor. This time roughly doubles with each additional producer.

Iterating on the first-order conditions

A convenient way to solve for the optimal allocation using the reshaping approach is to iterate on the first-order conditions of the log of $\mathcal{P}'_{RR}$, as defined by (11) and 12. In what follows ζ_k is the Lagrange multiplier on the k-th inequality constraint in $\mathcal{P}'_{RR}$. The algorithm is as follows:

1. Initialize the 0-th iteration with $\Delta\mu_k^0 = \underline{\mu}_j^0 - \bar{\mu}_j^0 = -1$ for all k.
2. For the p-th iteration:
   (a) Using the complementary slackness condition set $\theta_k^p = 1$ if $\Delta\mu_k^p \leq 0$ and $\theta_k^p = 0$ if $\Delta\mu_k^p > 0$
   (b) Using θ^p, Iterate on (10) to find the vector q^p
   (c) Find $\frac{\zeta_k^p q_k^p}{\theta_k^p}$ by solving the following system of linear equations derived from the first-order conditions:

$$\frac{(z_k A B_k^{\alpha})^{\sigma-1}}{\sum_{j=1}^n q_j^{\sigma-1}} + \alpha\,(z_k A)^{\varepsilon-1} B_k^{\alpha(\varepsilon-1)} \sum_{j=1}^n \Omega_{kj} \frac{\theta_j}{B_j^{\varepsilon-1}} \frac{q_j \zeta_j}{\theta_j} = \frac{\zeta_k q_k}{\theta_k}$$

   for each k and where $B_j = \left(\sum_{i=1}^n \theta_i^b \Omega_{ij} q_i^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}}$ is a composite of the productivities of j's suppliers.

(d) Compute $\Delta\mu_k^{p+1}$ using the following equation derived from the first-order conditions

$$\frac{f}{L - f\sum_{j=1}^{n}\theta_j} = \frac{1}{\sigma-1}\frac{q_k\zeta_k}{\theta_k} + \frac{\alpha}{\varepsilon-1}\left(1-\frac{\varepsilon-1}{\sigma-1}\right)(z_kA)^{\varepsilon-1}B_k^{\alpha(\varepsilon-1)}\sum_{j=1}^{n}\Omega_{kj}\frac{\theta_j}{B_j^{\varepsilon-1}}\frac{q_j\zeta_j}{\theta_j} + \Delta\mu_k.$$

3. Repeat step 2 above until convergence on $\Delta\mu$.

C Proofs

This section contains the proofs of the lemmas and propositions from the main text.

C.1 Results about q

An intermediate step in the proof of the uniqueness of q uses the following definition of R-concavity.

Definition. (Kennan, 2001) A function $g : \mathbb{R}^n \to \mathbb{R}^n$ is radially quasiconcave ("R-concave") if $g(x) = 0$ and $x > 0$ and $0 \le \lambda \le 1$ implies $g(\lambda x) \ge 0$. If (in addition) $0 < \lambda < 1$ implies $g(\lambda x) > 0$, then g is strictly R-concave.

In addition, we also need the concept of a quasi-increasing function.

Definition. (Kennan, 2001) A function $g : \mathbb{R}^n \to \mathbb{R}^n$ is quasi-increasing if $y_i = x_i$ and $y_j \ge x_j$ for all j implies $g_i(y) \ge g_i(x)$.

The following Lemma is used as an intermediate step to prove the uniqueness of the vector q.

Lemma 3. *Consider the mapping $g : \mathbb{R}^n \to \mathbb{R}^n$ defined as*

$$g_j(p) = (z_jA)^{\varepsilon}\left(\sum_{i\in\mathcal{N}}\Omega_{ij}p_i^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\alpha\frac{\varepsilon}{\varepsilon-1}} - p_j$$

for all $j \in \mathcal{N}$. Then g is strictly R-concave.

Proof. Suppose that there exists a $p^* > 0$ such that $g(p^*) = 0$. Then, for $0 \le \lambda \le 1$,

$$\begin{aligned}
g_j(\lambda p^*) &= (z_jA)^{\varepsilon}\lambda^{\alpha}\left(\sum_{i\in\mathcal{N}}\Omega_{ij}(p_i^*)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\alpha\frac{\varepsilon}{\varepsilon-1}} - \lambda p_j^* \\
&\ge (z_jA)^{\varepsilon}\lambda\left(\sum_{i\in\mathcal{N}}\Omega_{ij}(p_i^*)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\alpha\frac{\varepsilon}{\varepsilon-1}} - \lambda p_j^* \\
&\ge \lambda g_j(p^*) \\
&\ge 0
\end{aligned}$$

with strict inequality for $0 < \lambda < 1$ since $0 < \alpha < 1$ and $\sum_{i=1}^{n} \Omega_{ij} \geq 1$, which completes the proof. $\square$

Proposition 1. The Lagrange multipliers associated with the efficient allocation satisfy

$$q_j = z_j \theta_j A \left(\sum_{i \in \mathcal{N}} \Omega_{ij} q_i^{\varepsilon - 1} \right)^{\frac{\alpha}{\varepsilon - 1}}, \quad \forall j \in \mathcal{N} \tag{7}$$

Furthermore, there is a unique vector q that satisfies (7) such that $q_j > 0$ if firm j has access to a closed loop of active suppliers.

Proof. The first-order conditions with respect to l_j and x_{ij} are

$$\lambda_j (1 - \alpha) y_j = w l_j \tag{15}$$

$$\lambda_j \alpha \theta_j \frac{A z_j}{\alpha^{\alpha} (1 - \alpha)^{1 - \alpha}} \left(\sum_{i \in \mathcal{N}} (\Omega_{ij} x_{ij})^{\frac{\varepsilon - 1}{\varepsilon}} \right)^{\alpha \frac{\varepsilon}{\varepsilon - 1} - 1} l_j^{1 - \alpha} \Omega_{ij} x_{ij}^{\frac{\varepsilon - 1}{\varepsilon}} = \lambda_i x_{ij}.$$

Combining these first-order conditions with the production function yields

$$x_{ij} \lambda_i^{\varepsilon} = \lambda_j^{\varepsilon} \alpha \left(z_j \theta_j A \left(\frac{\lambda_j}{w} \right)^{1 - \alpha} \right)^{\frac{\varepsilon - 1}{\alpha}} y_j. \tag{16}$$

Plugging this equation and the first-order condition for l_j back into the production function yields (7).

To prove the uniqueness of q, I rely on a result by Kennan (2001). Consider the change of variable $p_j = q_j^{\varepsilon}$, and let us define a matrix $\tilde{\Omega}$ which is represents the same network as Ω but with only the firms that are 1) active, and that 2) have access to a closed loop of active suppliers. All the other firms are such that $p_j = 0$. The recursive equation becomes

$$p_j = (z_j A)^{\varepsilon} \left(\sum_{i \in \mathcal{N}} \tilde{\Omega}_{ij} p_i^{\frac{\varepsilon - 1}{\varepsilon}} \right)^{\alpha \frac{\varepsilon}{\varepsilon - 1}}. \tag{17}$$

Denote the right-hand side of that equation by $f_j(p)$ and define the function $g : \mathbb{R}^n \to \mathbb{R}^n$ as $g(p) = f(p) - p$. Then, from Lemma 3, we now that g is strictly R-concave (note that $\sum_{i=1}^{n} \tilde{\Omega}_{ij} \geq 1$ so the Lemma applies). Note also that g is quasi-increasing.

Consider the mapping $h : \mathbb{R}_+ \to \mathbb{R}^n$ defined as $h(s) = f(\mathbb{1} s)$ where $\mathbb{1}$ is a vector full of 1. Then $h(s)$ is strictly concave, strictly increasing and differentiable with $h(0) = 0$, $\lim_{s \to 0} h'(s) = \infty$ and $\lim_{s \to \infty} h'(s) = 0$. As a result, there exist $b > a > 0$ such that $h(a) > a$ and $h(b) < b$. Then Theorem 3.1 and Theorem 3.2 of Kennan (2001) apply and (17) has a unique positive fixed point p^*. There is therefore a unique positive q^* that satisfies (7). It is such that $q_j^* = \left(p_j^*\right)^{\frac{1}{\varepsilon}}$ for $\theta_j > 0$

and j having access to a closed loop of active suppliers, and $q_j^* = 0$ otherwise. Note that the proof goes through if we include the reshaping constant in (7). $\square$

C.2 Efficient allocation with exogenous firm status

Lemma 1. Aggregate output in the optimal allocation is given by

$$Y = Q\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) \tag{8}$$

where $Q \equiv \left(\sum_{j\in\mathcal{N}} q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$.

Proof. The first-order condition on c_j yields

$$c_j = \left(\frac{q_j}{w}\right)^{\sigma} Y.$$

Raising both sides of this equation to the power $\frac{\sigma-1}{\sigma}$ and summing across firm we find that $w = Q$, so that the aggregate labor productivity is equal to the shadow value of labor. Using the first-order conditions (15) and (16) into the resource constraints (2), we find

$$\begin{aligned} c_j + \sum_{k=1}^{n} x_{jk} - y_j &\leq 0 \\ c_j + \sum_{k\in\mathcal{N}} \Omega_{jk}\left(\frac{q_j}{q_k}\right)^{\varepsilon}\alpha\left(z_k\theta_k A\left(\frac{1}{q_k}\right)^{1-\alpha}\right)^{\frac{\varepsilon-1}{\alpha}}\frac{q_k l_k}{1-\alpha} - \frac{q_j l_j}{1-\alpha} &\leq 0 \\ c_j + \sum_{k\in\mathcal{N}} \Omega_{jk} q_j^{\varepsilon}\frac{\alpha}{1-\alpha}\left(\frac{z_k A\theta_k}{q_k}\right)^{\frac{\varepsilon-1}{\alpha}} l_k - \frac{q_j}{1-\alpha} l_j &\leq 0 \end{aligned}$$

which, taking the other variables as fixed, is a linear equation in l. Combining with the first-order condition on c_j we find

$$\left(\frac{q_j}{Q}\right)^{\sigma-1}\frac{Y}{Q} = \frac{1}{1-\alpha}\left[l_j - \sum_{k\in\mathcal{N}}\Omega_{jk} q_j^{\varepsilon-1}\alpha\left(\frac{z_k A\theta_k}{q_k}\right)^{\frac{\varepsilon-1}{\alpha}} l_k\right]. \tag{18}$$

Define the $n\times n$ matrix Γ with elements

$$\Gamma_{jk} = \Omega_{jk} q_j^{\varepsilon-1}\left(\frac{z_k A\theta_k}{q_k}\right)^{\frac{\varepsilon-1}{\alpha}}.$$

Using (7) we can write Γ_{jk} as

$$\begin{aligned}\Gamma_{jk} &= \Omega_{jk}\left(q_j\right)^{\varepsilon-1}\left(\frac{z_k A\theta_k}{\theta_k A z_k\left(\sum_i \Omega_{ik} q_i^{\varepsilon-1}\right)^{\frac{\alpha}{\varepsilon-1}}}\right)^{\frac{\varepsilon-1}{\alpha}} \\ &= \frac{\Omega_{jk} q_j^{\varepsilon-1}}{\sum_i \Omega_{ik} q_i^{\varepsilon-1}}.\end{aligned}$$

Summing (18) across firms, we therefore get

$$\begin{aligned}\frac{Y}{Q} &= \frac{1}{1-\alpha}\left[\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) - \alpha\sum_{k=1}^{n}\frac{\sum_j \Omega_{jk} q_j^{\varepsilon-1}}{\sum_i \Omega_{ik} q_i^{\varepsilon-1}} l_k\right] \\ &= \left(L - f\sum_{j\in\mathcal{N}}\theta_j\right).\end{aligned}$$

where I used the resource constraint on labor.

We can derive all the relevant quantities in the allocation by using (18). Beginning by rewriting the equation in matrix form

$$\left(\frac{q}{Q}\right)^{\circ(\sigma-1)}\frac{Y}{Q} = \frac{1}{1-\alpha}\left[I_n - \alpha\Gamma\right] l.$$

Inverting this equation[38], we find the labor allocation

$$l = (1-\alpha)\frac{Y}{Q}\left[I_n - \alpha\Gamma\right]^{-1}\left(\frac{q}{Q}\right)^{\circ(\sigma-1)}.$$

Returning to the first-order condition on labor we can find the output of each firm

$$y = Y\left[I_n - \alpha\Gamma\right]^{-1}\left(\frac{q}{Q}\right)^{\circ\sigma}. \tag{19}$$

and from (16) we find the amount of goods transiting between any two firms, if $\Omega_{ij} = 1$, is

$$x_{ij} = \left(\frac{q_i}{q_j}\right)^{\varepsilon}\alpha\left(z_j\theta_j A\left(\frac{1}{q_j}\right)^{1-\alpha}\right)^{\frac{\varepsilon-1}{\alpha}} y_j. \tag{20}$$

$\square$

[38] If no self-connection are allowed, $\Omega_{jj} = 0$ for all j, then $\left[I_n - \alpha\Gamma\right]^T$ is strictly diagonally dominant (since the columns of Γ sum to 1 and $0 < \alpha < 1$) and it is therefore invertible. Since the transpose of an invertible matrix is invertible, $\left[I_n - \alpha\Gamma\right]$ is also invertible. In the numerical exercises, $I_n - \alpha\Gamma$ is always invertible.

C.3 Reshaping approach

Preliminary results

The proof of Proposition 3 uses the following lemmas.

Lemma 4. *Let $F = A + \varepsilon B - \varepsilon C$ where $\varepsilon > 0$, A is the $n \times n$ matrix with every elements equal to 1, B is an $n \times n$ positive definite matrix and C is an $n \times n$ matrix with coefficients $c_{ij} \geq 0$. Let P denote an orthogonal projection onto the subspace $S : \sum_{i=1}^{n} x_i = 0$. If C is negative semi-definite on S then F is positive semi-definite for $\varepsilon > 0$ small enough.*

Proof. The positive definiteness of B implies that $z'Bz \geq c \|z\|^2$ for all z and some $c > 0$. Take a vector $z = x + y$ where $x \in S$ and $y \perp S$. Then, for ε small enough,

$$\begin{aligned} z' (A + \varepsilon B - \varepsilon C) z = z'Az + \varepsilon z'Bz - \varepsilon z'Cz = n \|y\|^2 + \varepsilon c \|z\|^2 - \varepsilon x'Cx - \varepsilon y'Cy - 2\varepsilon y'Cx \\ \geq (n - 1/2) \|y\|^2 + c\varepsilon \left(\|x\|^2 + \|y\|^2 \right) - 2\varepsilon \|C\| \|x\| \|y\| \\ \geq 0. \end{aligned}$$

so that F is positive semi-definite. $\square$

Lemma 5. *Let $X \subset \mathbb{R}^n$ be a nonempty convex set such that $int(X)$ is also nonempty and let $f : X \to \mathbb{R}$ be a continuous function. If f is convex on $int(X)$ then f is convex on X.*

Proof. Since $\operatorname{int}(X) \neq \emptyset$, then $\operatorname{cl}(\operatorname{int}(X)) = \operatorname{cl}(X)$. Therefore, for all $x \in X$ there exists a sequence $x_n \to x$ with $x_n \in \operatorname{int}(X)$. Take two points $x, y \in X$ and $x_n \to x$, $y_n \to y$ with $x_n \in \operatorname{int}(X)$ and $y_n \in \operatorname{int}(X)$. For any $\lambda \in [0, 1]$, by the convexity of f on $\operatorname{int}(X)$ we have

$$f(\lambda x_n + (1 - \lambda) y_n) \leq \lambda f(x_n) + (1 - \lambda) f(y_n).$$

Therefore, by continuity of f on X,

$$\begin{aligned} f(\lambda x + (1 - \lambda) y) &= \lim_{n \to \infty} f(\lambda x_n + (1 - \lambda) y_n) \\ &\leq \lim_{n \to \infty} \lambda f(x_n) + (1 - \lambda) f(y_n) = \lambda f(x) + (1 - \lambda) f(y) \end{aligned}$$

so that f is convex on X. $\square$

Proofs of concavity

Proposition 2. If $\Omega_{ij} = c_i d_j$ for some vectors c and d then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{P}_{RR}$.

Proof. The proof first finds conditions under which the objective function of $\mathcal{P}_{RR}$ is log-concave. Raising both sides of 10 to the power $\varepsilon - 1$, multiplying them by $c_j \theta_j^b$ and summing across firms we find that

$$\sum_{j\in\mathcal{N}} c_j \theta_j^b \left(q_j\left(\theta\right)\right)^{\varepsilon-1} = \left(\sum_{j\in\mathcal{N}} c_j d_j^\alpha \left(Az_j\right)^{\varepsilon-1} \theta_j^{a(\varepsilon-1)+b}\right)^{\frac{1}{1-\alpha}}$$

so that

$$q_j\left(\theta\right) = z_j \theta_j^a A d_j^\alpha \left(\sum_{i\in\mathcal{N}} c_i d_i^\alpha \left(Az_i\right)^{\varepsilon-1} \theta_i^{a(\varepsilon-1)+b}\right)^{\frac{\alpha}{1-\alpha}\frac{1}{\varepsilon-1}}.$$

Using the expression for $V_{RR}\left(\theta\right)$, we find that

$$\begin{aligned}
\log\left(V_{RR}\left(\theta\right)\right) = & \frac{1}{\sigma-1}\log\left(\left(\sum_{j\in\mathcal{N}}\left(z_j\theta_j^a A d_j^\alpha\right)^{\sigma-1}\right)\left(\sum_{i\in\mathcal{N}} c_i d_i^\alpha \left(Az_i\right)^{\varepsilon-1}\theta_i^{a(\varepsilon-1)+b}\right)^{\frac{\alpha}{1-\alpha}\frac{\sigma-1}{\varepsilon-1}}\right) \\
& + \log\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right) \\
= & \frac{1}{\sigma-1}\log\left(\sum_{j\in\mathcal{N}}\left(z_j\theta_j^a A d_j^\alpha\right)^{\sigma-1}\right) + \frac{1}{\varepsilon-1}\frac{\alpha}{1-\alpha}\log\left(\sum_{i\in\mathcal{N}} c_i d_i^\alpha\left(Az_i\right)^{\varepsilon-1}\theta_i^{a(\varepsilon-1)+b}\right) \\
& + \log\left(L - f\sum_{j\in\mathcal{N}}\theta_j\right).
\end{aligned}$$

If $0 < a \leq (\sigma-1)^{-1}$ and $-a\left(\varepsilon-1\right) \leq b \leq 1 - a\left(\varepsilon-1\right)$ (and in particular if a and b satisfy ($\star$)) the exponents on θ are all between 0 and 1 so that the first two summations are concave functions of θ. The last summation is linear in θ so it is concave. Since the log of a concave function is concave, we have that $\log\left(V_{RR}\right)$ is also a concave function of θ. Since, in addition, the constraint set $\theta \in [0,1]^n$ is convex and the Slater's qualification condition is obviously satisfied, the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize an optimal solution to the maximization of $\log\left(V_{RR}\left(\theta\right)\right)$ on the set $\theta \in [0,1]^n$. Since log is a monotone transformation, a solution to this problem also solves $\mathcal{P}_{RR}$. □

Proposition 3. Let $\sigma = \varepsilon$ and suppose that the fixed cost $f > 0$ and the dispersion in productivities $\overline{z} - \underline{z} > 0$ are not too big. If the network Ω is large enough and sufficiently connected with $\Omega_{ii} = 0$ for all i, then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $\mathcal{P}_{RR}$.

Proof. To simplify the notation, define $p_j = q_j^{\sigma-1}$ and

$$g^j = \frac{p_j}{z_j^{\sigma-1}\left(\sum_i \Omega_{ij} p_i\right)^\alpha}.$$

The optimization problem $\mathcal{P}_{RR}$ can be written as

$$\min_{p \in P} -\frac{1}{\sigma - 1} \log \left(\sum_{j \in \mathcal{N}} p_j \right) - \log \left(L - f \sum_{j \in \mathcal{N}} g^j (p) \right)$$

where P is defined as the set of vectors p such that

$$\begin{aligned} -p_j &\leq 0 \\ p_j &\leq z_j^{\sigma-1} \left(\sum_{i \in \mathcal{N}} \Omega_{ij} p_i \right)^{\alpha} \end{aligned}$$

for each $j \in \mathcal{N}$.

We first show that the objective function, which we denote by $\Lambda(p)$, is convex on the interior of P. To do so, consider P_μ, the set of all p such that

$$\begin{aligned} p_j &\geq \mu \\ p_j &\leq z_j^{\sigma-1} \left(\sum_{i \in \mathcal{N}} \Omega_{ij} p_i \right)^{\alpha} \end{aligned}$$

for $\mu > 0$ small. The hessian of Λ is such that each of its element is

$$\frac{\partial^2 \Lambda}{\partial p_k \partial p_l} = \frac{1}{\varepsilon - 1} \left(\sum_j p_j \right)^{-2} + f \frac{\sum_j g_{kl}^j (p)}{L - f \sum_j g^j (p)} + f^2 \frac{\left(\sum_j g_k^j (p) \right) \left(\sum_j g_l^j (p) \right)}{\left(L - f \sum_j g^j (p) \right)^2}.$$

The first term is a constant matrix. The last term can clearly be expressed as a Gramian matrix and its contribution to the hessian is therefore positive semidefinite. Using the derivatives of g explicitly, the second term becomes

$$f \frac{1}{L - f \sum_j g^j (p)} \left(\underbrace{\sum_{j \in \mathcal{N}} \frac{\alpha (\alpha + 1) p_j \Omega_{kj} \Omega_{lj}}{z_j^{\sigma-1} \left(\sum_i \Omega_{ij} p_i \right)^{\alpha+2}}}_{B_{kl}} - \underbrace{\left(\frac{\alpha \Omega_{lk}}{z_k^{\sigma-1} \left(\sum_i \Omega_{ik} p_i \right)^{\alpha+1}} + \frac{\alpha \Omega_{kl}}{z_l^{\sigma-1} \left(\sum_i \Omega_{il} p_i \right)^{\alpha+1}} \right)}_{C_{kl}} \right). \tag{21}$$

We will show that, when Ω is sufficiently connected, the matrix created by the first term in parenthesis (B) is positive definite and the matrix created by the last term in parenthesis (C) is negative semi-definite on the subspace $S : \sum_{i=1}^{n} x_i = 0$. By Lemma 4, the whole Hessian will therefore be positive semi-definite for f small enough.

Note first that B can be written as a Gramian matrix in which each element B_{kl} is the scalar product of a pair of vectors v_k and v_l defined as

$$v_m = \left[\begin{array}{ccccc} \sqrt{\frac{\alpha(\alpha+1)p_1}{z_1^{\sigma-1}\left(\sum_i \Omega_{i1} p_i\right)^{\alpha+2}}}\Omega_{m1} & \cdots & \sqrt{\frac{\alpha(\alpha+1)p_j}{z_j^{\sigma-1}\left(\sum_i \Omega_{ij} p_i\right)^{\alpha+2}}}\Omega_{mj} & \cdots & \sqrt{\frac{\alpha(\alpha+1)p_n}{z_n^{\sigma-1}\left(\sum_i \Omega_{in} p_i\right)^{\alpha+2}}}\Omega_{mn} \end{array} \right]'.$$

Now, define M as the matrix that has the vectors $v_1, \ldots, v_n$ as columns such that

$$M = \mathrm{diag}\left(\left[\begin{array}{ccccc} \sqrt{\frac{\alpha(\alpha+1)p_1}{z_1^{\sigma-1}\left(\sum_i \Omega_{i1} p_i\right)^{\alpha+2}}} & \cdots & \sqrt{\frac{\alpha(\alpha+1)p_j}{z_j^{\sigma-1}\left(\sum_i \Omega_{ij} p_i\right)^{\alpha+2}}} & \cdots & \sqrt{\frac{\alpha(\alpha+1)p_n}{z_n^{\sigma-1}\left(\sum_i \Omega_{in} p_i\right)^{\alpha+2}}} \end{array} \right]\right)\Omega.$$

When Ω is as connected as possible under the assumptions of the proposition ($\Omega_{ij} = 1$ for all i, j except for the empty diagonal), it is invertible such that M is also invertible (recall that $p_j > 0$ for all j). The vectors v_m are therefore linearly independent and B is therefore positive definite. Now, this result also stands for matrices Ω that are less connected. For instance, if we remove a random connection from Ω, we still have $\det(\Omega) \neq 0$ and B is still positive definite. This process can be repeated several times, in general, for a large network. If we keep removing connections, however, at some point Ω might become singular and B is then only positive semi-definite.

Now, consider the matrix C. By defining the vector b as having elements $b_j = \frac{\alpha}{z_j^{\sigma-1}\left(\sum_i \Omega_{ij} p_i\right)^{\alpha+1}}$, we can write

$$C = \Omega\mathrm{diag}(b) + (\Omega\mathrm{diag}(b))'$$

Suppose again that Ω is as connected as possible under the assumptions of the proposition such that $\Omega = \mathbb{1} - I_n$, where $\mathbb{1}$ is a matrix full of ones and I_n is the $n \times n$ indicator matrix. Take any vector x in S, then

$$\begin{aligned} x'Cx &= x'\left[(\mathbb{1} - I)\mathrm{diag}(b) + ((\mathbb{1} - I)\mathrm{diag}(b))'\right]x \\ &= x'\left[-2I\mathrm{diag}(b)\right]x \\ &< 0 \end{aligned}$$

so that C is negative definite. Again, this result also holds for matrices Ω that are less connected. For instance, if we repeat the earlier exercise and remove a connection which, without loss of generality, we pick as Ω_{12}. We write $\Omega = \mathbb{1} - I - \Omega_{12}$ where, by abuse of notation, Ω_{12} denotes a matrix full of zero except for the second column of the first row which equals one. Repeating the steps from above

$$x'Cx = -x'\left[(I + \Omega_{12})\mathrm{diag}(b) + ((I + \Omega_{12})\mathrm{diag}(b))'\right]x$$

and the matrix between square brackets is diagonally dominant, and C is therefore negative semi-definite, if $2b_1 > b_2$. In a large connected network in which the z's are close to each other, the b's cannot be far apart and this condition is satisfied (recall that the p's are bounded away from 0).

Again, one can repeat this operation several times, depending on which specific links are removed from Ω, and keep C negative semi-definite.

Summarizing the results so far, we have shown that in a large connected network Ω the terms B and C in (21) are positive definite and positive semi-definite, respectively. Using Lemma 4 we can conclude that the Hessian of the objective function is positive semi-definite for f sufficiently small and Λ is therefore convex on P_μ. This is true for every $\mu > 0$ so Λ is therefore convex on $\text{int}(P)$. Since Λ is continuous, it is also convex on P by Lemma 5. The constraint set of P is convex and the Slater condition is also satisfied so that the Karush-Kuhn-Tucker conditions are therefore necessary and sufficient to characterize a solution to $\mathcal{P}_{RR}$. □

Proof of the equivalence of the solutions

Denote by $V_{RR} : [0,1]^n \to \mathbb{R}^+$ the objective function of $\mathcal{P}_{RR}$ defined as

$$V_{RR}(\theta) = \left(\sum_{j\in\mathcal{N}} (q_j(\theta))^{\sigma-1}\right)^{\sigma-1} \left(L - f\sum_{j\in\mathcal{N}} \theta_j\right) \tag{22}$$

where q_j is implicitly defined as

$$q_j(\theta) = z_j \theta_j^a A \left(\sum_{i\in\mathcal{N}} \theta_i^b \Omega_{ij} (q_i(\theta))^{\varepsilon-1}\right)^{\frac{\alpha}{\varepsilon-1}}. \tag{23}$$

Similarly, denote by V_{SP} the objective function of $\mathcal{P}_{SP}$.

Proposition 4. If a solution θ^* to $\mathcal{P}_{RR}$ is such that $\theta_j^* \in \{0,1\}$ for all j, then θ^* also solves $\mathcal{P}_{SP}$.

Proof. From the assumptions of the proposition, assume that $\theta^* \in \{0,1\}^n$ solves $\mathcal{P}_{RR}$. By construction, the objective function V_{RR} of $\mathcal{P}_{RR}$ and the objective function V_{SP} of $\mathcal{P}_{SP}$ coincide over $\{0,1\}^n$. Therefore $V_{RR}(\theta^*) = V_{SP}(\theta^*)$. Since the feasible set of $\mathcal{P}_{RR}$, $[0,1]^n$, contains the feasible set of $\mathcal{P}_{SP}$, $\{0,1\}^n$, it must be that $V_{SP}(\theta^*) \geq V_{SP}(\theta)$ for $\theta \in \{0,1\}^n$, otherwise θ^* would not be a solution to $\mathcal{P}_{RR}$. □

Lemma 2. Under the condition ($\star$), the first-order condition (13) for the operating decision of firm j only depends on θ_j through aggregates.

Proof. Begin with equation (13) and consider the term

$$\begin{aligned}
\frac{\partial q_k}{\partial \theta_k}\frac{\partial Q}{\partial q_k} &= z_k a \theta_k^{a-1} A B_k^\alpha \times \left(\sum_{j\in\mathcal{N}} q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}-1} q_k^{\sigma-2} \\
&= z_k a \theta_k^{a-1} A B_k^\alpha \times Q^{2-\sigma} \left(z_k \theta_k^a A B_k^\alpha\right)^{\sigma-2}
\end{aligned}$$

Similarly, for the term in (13) that is inside the summation

$$\frac{\partial B_j}{\partial \theta_k}+\frac{\partial q_k}{\partial \theta_k}\frac{\partial B_j}{\partial q_k}=\frac{1}{\varepsilon-1}B_j^{2-\varepsilon}b\theta_k^{b-1}\Omega_{kj}q_k^{\varepsilon-1}+az_k\theta_k^{a-1}AB_k^{\alpha}\times B_j^{2-\varepsilon}\theta_k^{b}\Omega_{kj}q_k^{\varepsilon-2}$$
$$=z_k^{\varepsilon-1}\theta_k^{b-1+a(\varepsilon-1)}\Omega_{kj}B_j\left(\frac{AB_k^{\alpha}}{B_j}\right)^{\varepsilon-1}\left(a+b\frac{1}{\varepsilon-1}\right)$$

Substituting these two expressions in (13) we find that the first-order condition for θ_j can be written as

$$z_k a\theta_k^{a-1}AB_k^{\alpha}\times Q^{2-\sigma}\left(z_k\theta_k^{a}AB_k^{\alpha}\right)^{\sigma-2}\left(L-f\sum_{j\in\mathcal{N}}\theta_j\right)-fQ+$$
$$z_k^{\varepsilon-1}\theta_k^{b-1+a(\varepsilon-1)}\sum_{j\in\mathcal{N}}\beta_j\Omega_{kj}B_j\left(\frac{AB_k^{\alpha}}{B_j}\right)^{\varepsilon-1}\left(a+b\frac{1}{\varepsilon-1}\right)\frac{\partial q_j}{\partial B_j}=\overline{\mu}_k-\underline{\mu}_k.$$

When a and b take on the values of the condition ($\star$) we see that the first-order condition does not depend directly on θ_j. In fact, θ_j only enters the equations through the aggregates $\{B_j\}_{j\in\mathcal{N}}$ and Q. □

C.4 Equilibrium

Proposition 5. Suppose that the solution to $\mathcal{P}_{RR}$ also solves $\mathcal{P}_{SP}$ and that $\sigma=\epsilon$. Then the efficient allocation can be implemented as an equilibrium with a subsidy to the purchase of intermediate inputs and a lump-sum tax on the representative consumer.

Proof. Let us first consider the problem of a firm that faces an input subsidy $1-s_d=\frac{\sigma-1}{\sigma}$. Its problem is

$$\max\ P_jc_j+\sum_{i\in\mathcal{N}}x_{ji}P_{ji}-(1-s_d)\sum_{i\in\mathcal{N}}x_{ij}P_{ij}-Wl_j-Wf\theta_j$$

subject to the feasibility constraint

$$c_j+\sum_{k\in\mathcal{N}}x_{jk}\leq\frac{A}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\theta_j z_j\left(\sum_{i\in\mathcal{N}}\Omega_{ij}x_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\alpha\frac{\sigma}{\sigma-1}}l_j^{1-\alpha}$$

and the demand curves

$$P_j=P\left(\frac{c_j}{C}\right)^{-\frac{1}{\sigma}}\text{ and }P_{ji}=\bar{P}_i\left(\frac{x_{ji}}{X_i}\right)^{-\frac{1}{\sigma}}.$$

The first-order conditions yield the usual pricing equations $P_{ji}=\frac{\sigma}{\sigma-1}\beta_j$ and $P_j=\frac{\sigma}{\sigma-1}\beta_j$, where β_j is the Lagrange multiplier on the feasibility constraint (i.e. the marginal cost of production).

The first-order conditions on l_j and x_{ij} are

$$Wl_j = \beta_j (1-\alpha) y_j$$

and

$$\beta_i x_{ij} = \beta_j \alpha \frac{A}{\alpha^\alpha (1-\alpha)^{1-\alpha}} \theta_j z_j \left(\sum_{i \in \mathcal{N}} \Omega_{ij} x_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\alpha \frac{\sigma}{\sigma-1} - 1} \Omega_{ij} x_{ij}^{\frac{\sigma-1}{\sigma}} l_j^{1-\alpha}.$$

Notice that these two equations are the same as in the planner's problem, as shown by (15). Since these equations, together with the production function, yield the recursive equation for q (7) we have shown that the same recursive equation applies for W/β_j in the equilibrium. Since, by Proposition 1 this equation has a unique solution we have that, conditional on θ, $W/\beta_j = q_j$ and, if wages are equal in the equilibrium and in the efficient allocation , $\lambda_j = \beta_j$. As a result, we see that the labor allocation l and the inter-firm flows x_{ij} are the same, again conditional on θ, in the efficient allocation and in the equilibrium. Note that this also implies that the wage in the equilibrium W clears the labor market if it equals to the Lagrange multiplier w in the efficient allocation.

The last step is to show that the entry decisions θ are the same in the equilibrium and in the planner's allocation. In the equilibrium, simple algebra shows that the profits of a firm j can be written as

$$\pi_j = \frac{1}{\sigma - 1} \beta_j y_j - W f \theta_j$$

such that a firm optimally decides to operate if the total value of its output is high enough to compensate for the payment of the fixed cost. We will show that the same equation governs θ in the planner's problem $\mathcal{P}_{RR}$.

Recall that $\mathcal{P}_{RR}$ can be written as

$$\mathcal{P}'_{RR}\colon \quad \max_{\theta \in [0,1]^n, q} \left(\sum_{j \in \mathcal{N}} q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \left(L - f \sum_{j \in \mathcal{N}} \theta_j \right)$$

$$q_j \leq A z_j \theta_j^{\frac{1}{\sigma-1}} \left(\sum_{i \in \mathcal{N}} \theta_i^{1 - \frac{\varepsilon-1}{\sigma-1}} \Omega_{ij} q_i^{\varepsilon-1} \right)^{\frac{\alpha}{\varepsilon-1}}.$$

Denoting by δ_j the Lagrange multiplier on the jth inequality constraint, the first-order condition on q_j is

$$\left(\frac{q_j}{Q} \right)^{\sigma-1} \frac{Y}{Q} = \frac{\delta_j q_j}{Q} - \sum_{i \in \mathcal{N}} \alpha \Omega_{ji} q_j^{\varepsilon-1} \frac{1}{\sum_k \Omega_{ki} q_k^{\varepsilon-1}} \frac{\delta_i q_i}{Q} \tag{24}$$

and the first-order conditions on θ_j is

$$\frac{1}{\sigma-1}\delta_j q_j = \theta_j \Delta\mu_j + \theta_j f \left(\sum_{j\in\mathcal{N}} q_j^{\sigma-1}\right)^{\frac{1}{\sigma-1}}.$$

where, as before, $\Delta\mu_j$ is the difference between the multipliers on $\theta_j \leq 1$ and $\theta_j \geq 0$. Notice that (24) is essentially the same as (18) such that, given θ_j, $\delta_j q_j = \lambda_j y_j$. In other words, the value of endowing a firm with some productivity q_j is the same as the value of its production. Using this result in the first-order condition on θ_j, and recognizing that $\left(\sum_j q_j^{\sigma-1}\right)^{1/(\sigma-1)} = Q = w$, we find that

$$\frac{1}{\sigma-1}\lambda_j y_j - wf\theta_j = \theta_j \Delta\mu_j.$$

To show why this expression implies the same entry decision as in the equilibrium, consider the case in which $\theta_j = 1$ in the efficient allocation. Then

$$\frac{1}{\sigma-1}\lambda_j y_j - wf = \Delta\mu_j > 0$$

where the inequality follows from the complementary slackness condition. Since $w = W$ and $\lambda_j = \beta_j$, this expression implies that firm j also chooses $\theta_j = 1$ in the equilibrium. Inversely, if $\theta_j = 0$ in the efficient allocation,

$$\frac{1}{\sigma-1}\frac{y_j}{q_j\theta_j} - f = \frac{\Delta\mu_j}{w} < 0$$

and the firm prefers to set $\theta_j = 0$ in the equilibrium.[39] □

[39]Note that $\frac{y_j}{q_j\theta_j}$ is in finite and strictly positive in the limit as θ_j goes to 0. To see this, recall from (19) that the numerator, y_j, scales as $\theta_j^{\frac{\sigma}{\sigma-1}}$ which is the same scaling as the denominator, $q_j\theta_j$.