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

Mathieu Taschereau-Dumouchel

Abstract

This paper studies an economy in which the production network is endogenously determined by the firms’ extensive margin of production. Because of the presence of fixed costs, a firm might decide not to produce, thereby severing connections with potential suppliers and customers. Every stable equilibrium in this economy can be characterized as the solution to a social planner’s problem. But because of the discrete decisions involved in the formation of the network, standard optimization techniques can take an infeasibly long time to find a solution. To overcome this issue, I propose a novel solution method that involves reshaping the planner’s problem. Analytic results and numerical simulations show that the method rapidly finds the planner’s solution in a class of economies that were considered particularly challenging. To illustrate how this approach works in practice, I show that a basic calibration of the model can capture how the U.S. production network changes over the business cycle. The calibrated model also features cascades of firm shutdowns that resemble those observed in the data. In addition, I find that the endogenous reorganization of the network leads to substantially smaller variations in aggregate output.

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 (Acemoglu et al., 2012). But the network itself is also constantly changing in response to these shocks. In the data, a key driver behind these changes is the firms’ entry and exit decisions. For instance, in the U.S. a large fraction of all link destructions

*This draft: July 2019. First version: February 2017. Cornell University, mt763@cornell.edu. I thank Enghin Atalay, Julieta Caunedo, Jeremy Greenwood, Christopher Huckfeldt, Ben Lester, Kristoffer Nimark, Ezra Oberfield, Nicolas Petrosky-Nadeau, Carlos Ramírez, Stephen Redding, Gill Segal, Ali Shourideh, Gianluca Violante, 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, Rutgers, the New York Fed, the Midwest Macro Meetings, Minnesota Macro, USC Marshall, Queen’s, the NYU Search Conference, Princeton and the 2019 AEA for helpful comments. Finally, I thank the anonymous referees and the editor for suggestions that have substantially improved the paper. I gratefully acknowledge support from Cornell’s Institute for the Social Sciences.
occur when one of the two firms forming the relationship stops production. Yet, these types of network formation problems have been neglected by the literature—perhaps because of technical issues that make standard solution techniques unhelpful.

This paper proposes a methodology to solve network formation problems driven by the firms’ extensive margin of production. The general setting is as follows: there is a finite number $n$ of firms each producing a differentiated good using labor and a set of inputs from other producers. Production requires the payment of a fixed cost so that 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.

Every stable equilibrium is efficient in this environment, and I therefore focus on the problem $\mathcal{P}$ of a social planner. A bare bones version of this problem involves choosing the operating status $\theta_j \in \{0, 1\}$ of the $n$ firms to maximize a welfare function $V_{\mathcal{P}}$. This problem is hard to solve for two reasons. First, the operating decisions of the firms are discrete so that $\mathcal{P}$’s feasible set, $\{0, 1\}^n$, is non-convex. Second, complementarities between operating decisions of nearby producers arise naturally in this environment—a firm will not produce without a supplier—and their presence breaks the concavity of the objective function. As a result, $V_{\mathcal{P}}$ features multiple local maxima and standard numerical algorithms are ineffective. Finding a solution, for instance by comparing the welfare level provided by each of the $2^n$ potential vectors $\theta$ in $\{0, 1\}^n$, can take an infeasibly long time.

To overcome these difficulties, this paper proposes a novel solution approach that involves reshaping the planner’s problem. Consider an alternative optimization problem, denoted by $\mathcal{R}$, that is constructed from $\mathcal{P}$ by first relaxing the feasible set $\{0, 1\}^n$ so that it now includes all the points in $[0, 1]^n$. This process introduces new points to optimize over but, importantly, these new points have no economic meaning in our environment. We can therefore modify the objective function over them to help us solve $\mathcal{P}$. To that end, consider an objective function $V_{\mathcal{R}}$ for $\mathcal{R}$ that coincides with $V_{\mathcal{P}}$ over the economically meaningful points, $\{0, 1\}^n$, while possibly differing elsewhere. The key idea behind the solution approach, is to construct $V_{\mathcal{R}}$ so that $\mathcal{R}$ is easy to solve and that its solution also solves the planner’s problem $\mathcal{P}$. I describe how to construct such a $V_{\mathcal{R}}$, and I establish sufficient conditions under which this approach is guaranteed to find the efficient allocation. But even when those conditions are not met, numerical simulations show that the proposed procedure provides a rapid and robust way of tackling network formation problems that were considered particularly challenging (Carvalho and Tahbaz-Salehi, 2018).

To illustrate how the approach can be useful in practice, I provide a basic calibration of the model to the United States economy. I find that the economic forces at work in the environment play an important role in shaping the production network. In particular, the economy features cascades of firm shutdowns that resemble those in the data. In addition, the mechanisms generate correlations between

---

\footnote{Appendix C.1 explores the importance of the extensive margin of production for link destruction in the U.S. data.}
the shape of the network and the business cycles that are also consistent with the data. Finally, I find that allowing the production network to reorganize itself in response to idiosyncratic shocks leads to substantially smaller variations in aggregate output. This last exercise highlights the importance of the formation of the network for the aggregation of microeconomic shocks into macroeconomic fluctuations.

**Literature.** This paper contributes to a recent literature in which production networks are built endogenously by the decisions of economic agents.\(^2\) Oberfield (2018) builds a model in which producers optimally choose one input from a randomly evolving set of suppliers, thereby creating the production network. He finds that star suppliers can emerge endogenously in equilibrium. Lim (2018) studies sourcing decisions in a model with sticky relationships. In independent contemporaneous work Acemoglu and Azar (2018) consider a network of competitive industries in which firms select a production technique that involves different sets of suppliers. They show that the endogenous evolution of the network can generate long-run growth. These papers feature a continuum of firms so that aggregate fluctuations do not arise from individual firm-level shocks. Tintelnot et al. (2018) build a model of endogenous network formation and international trade. In contrast to the current paper, they only consider acyclic networks. Boehm and Oberfield (2018) estimate a model of network formation using Indian micro data to study misallocation in the inputs market. In contrast to this literature, the current paper studies an economy in which the input-output network is built endogenously through the extensive production decisions of the firms—a margin that accounts for a large fraction of link changes in the U.S. data and that allows for cascades of firm shutdowns. This paper also relates to Baqae (2018) who studies cascades of firm shutdowns but in an economy with an exogenous production network.

One methodological contribution of this paper, that may be of independent interest, is a new solution technique for nonlinear optimization problems with binary variables. Several heuristics have been developed to handle these problems (Li and Sun, 2006). Closest to the present work are smoothing algorithms that attempt to get rid of the local maxima that emerge in the relaxed problem (Murray and Ng, 2010). In practice, finding an appropriate smoother is usually done through trial and error and there is no guarantee that the algorithm converges to a global maximum. In contrast, the current work explicitly describes how to reshape the objective function of the planner and proposes a rapid and robust solution method.

2 Model

The model is static, and there are three types of agents: firms, a final good producer and a representative household. There is a set \( \mathcal{N} = \{1, \ldots, n\} \) of firms, each of which produces a dif-

\(^2\)Another literature studies networks that are mostly formed through an exogenous random process. See for instance Atalay et al. (2011) and Carvalho and Voigtländer (2015). König et al. (2018) proposes an hybrid model in which a firm’s extensive production decision depends on its number of customers and an exogenous shocks.
ferentiated good that can be used as intermediate input by the final good producer and the other firms. The final good producer uses a CES production technology with elasticity of substitution \( \sigma > 1 \) and factor intensities \( \{\beta_j\}_{j \in \mathcal{N}} \) to convert intermediate inputs \( \{c_j\}_{j \in \mathcal{N}} \) into aggregate output

\[
C = \left( \sum_{j \in \mathcal{N}} \beta_j \frac{1}{\sigma} \frac{c_j^{(\sigma-1)/\sigma}}{\sigma/(\sigma-1)} \right)^{\sigma/(\sigma-1)}.
\]

The representative household consumes \( C \) and supplies \( L \) units of labor inelastically.

To produce, a firm \( j \in \mathcal{N} \) must employ \( f_j L \geq 0 \) units of labor as a fixed cost, in which case we say that it is operating. The vector \( \theta \in \{0,1\}^n \) keeps track of the operating decisions of the firms, such that \( \theta_j = 1 \) if \( j \) operates and \( \theta_j = 0 \) otherwise. When it operates, firm \( j \) has access to a technology that converts \( l_j \) units of labor and a vector of intermediate inputs \( \{x_{ij}\}_{i \in \mathcal{N}} \) into \( y_j \) units of good \( j \) according to the production function

\[
y_j = \frac{A}{\alpha_j \sigma_j (1 - \alpha_j)^{1-\alpha_j}} z_j \theta_j \left( \sum_{i \in \mathcal{N}} \Omega_{ij}^{-1} \frac{\epsilon_j^{\alpha_j - 1}}{\epsilon_j^{\alpha_j} \alpha_j} x_{ij} \right) l_j^{1-\alpha_j} \tag{1}
\]

where \( \Omega_{ij} \geq 0 \) denotes the exogenous factor intensity of input \( i \), \( \varpsilon_j > 1 \) is the elasticity of substitution between inputs, \( 0 < 1 - \alpha_j < 1 \) is the labor intensity, and \( A > 0 \) and \( z_j > 0 \) are aggregate and firm-specific total factor productivities.\(^3\)

We see from (1) that a firm \( j \) can only use inputs from a supplier \( i \) if \( \Omega_{ij} > 0 \). As such, the matrix \( \Omega \) describes a network of potential connections between firms. A potential connection \((i,j)\) is active—with goods being traded—if firms \( i \) and \( j \) both operate, otherwise it is inactive. The production network is therefore jointly determined by \( \Omega \) and \( \theta \), and economic conditions, through their impact on the firms’ operating decisions, endogenously determine the shape of the network.

Figure 1a provides an example of the potential connections \( \Omega_{ij} > 0 \) in an economy with six firms. The set of active connections, in blue in Figure 1b, is determined by the set of operating firms.

While the paper focuses on the operating decisions \( \theta \) as the key determinants of the shape of the network, the model is general enough to accommodate the formation of individual links. For that purpose, we can think of a link as a pseudo firm whose good is not included in the production of the final good, and that has a single potential supplier and customer.

**Planner’s problem.** Consider the problem \( \mathcal{P} \) of a social planner that maximizes final good consumption

\[
\mathcal{P} : \max_{\mathcal{P} : \max_{c \geq 0, x \geq 0, L \geq 0}} \left( \sum_{j \in \mathcal{N}} \beta_j \frac{1}{\sigma_j} \frac{c_j^{(\sigma-1)/\sigma}}{\sigma/(\sigma-1)} \right)^{\sigma/(\sigma-1)} \tag{2}
\]

\(^3\)Without loss of generality, assume that \( \sum_{i \in \mathcal{N}} \Omega_{ij} > 0 \) for all \( j \in \mathcal{N} \), otherwise firm \( j \) cannot produce and we can redefine the economy without it. All quantities in this environment are common knowledge.
subject to a resource constraint for each good $j \in \mathcal{N}$,

$$c_j + \sum_{k \in \mathcal{N}} x_{jk} \leq y_j,$$

where $y_j$ is given by (1), and a resource constraint for labor,

$$\sum_{j \in \mathcal{N}} l_j + \sum_{j \in \mathcal{N}} \theta_j f_j L \leq L. \quad (4)$$

We say that an allocation is efficient if it solves $\mathcal{P}$.

The paper focuses on the planner’s problem, but Appendix A shows that any stable equilibrium in this environment is efficient. As such, by solving $\mathcal{P}$ we are also implicitly characterizing all equilibrium allocations.

3 Solving $\mathcal{P}$

To solve $\mathcal{P}$, it is useful to first find the best allocation when the production network, or equivalently $\theta$, is fixed. For a fixed $\theta$, $\mathcal{P}$ is a convex maximization problem and the usual Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize its solution. Denote by $\lambda_j$ the Lagrange multiplier on the resource constraint (3) for good $j$ and by $w$ the multiplier on the labor resource constraint (4). The first-order conditions imply that $(1 - \alpha_j) y_j \lambda_j = w l_j$ so that, as in Oberfield (2018), we can define $q_j = w/\lambda_j$ as a measure of firm $j$’s productivity.

The planner’s first-order conditions together with the production function yield the following result.
Proposition 1. In the efficient allocation, the productivity vector $q$ satisfies, for all $j \in \mathcal{N}$,

$$q_j = z_j \theta_j A \left( \sum_{i \in \mathcal{N}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}}. \tag{5}$$

There is a unique $q$ that solves (5) such that $q_j > 0$ if $j$ operates and has access to at least one input.\(^4\)

Proof. All proofs are in Appendix D. \qed

Equation (5) fully characterizes the firms’ productivities in this environment. Its recursive structure implies that any change in productivity propagates downstream—from customer to customer—through production chains. In addition, (5) implies that a firm that has access to a greater variety of active suppliers, leading to more terms in the summation in (5), is more productive—a standard consequence of the CES production function. This benefit from input variety has important implications for the shape of the network as we will see in Section 4.

Equation (5) can be solved easily by iterating on the mapping (Kennan, 2001). Once $q$ is known, the following lemma shows that aggregate output $C$ can be computed as the product of aggregate productivity $Q$ and the amount of labor available after fixed costs have been paid.

Lemma 1. In the efficient allocation, aggregate output is

$$C = Q \left( 1 - \sum_{j \in \mathcal{N}} \theta_j f_j \right) L \tag{6}$$

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

We can now take a step back to consider the full problem $\mathcal{P}$ in which the network itself is a choice variable. By combining Proposition 1 and Lemma 1, we can write $\mathcal{P}$ as the problem of finding a vector $\theta^*$ that maximizes (6), where $q$ solves (5). This problem is hard to solve for two reasons.\(^5\) First, $\theta$ is limited to the corners $\{0, 1\}^n$ of the $n$-dimensional unit hypercube, a non-convex set. But even if $\theta$ could move freely over $[0, 1]^n$, the fixed costs create firm-level increasing returns to scale that break the (quasi) concavity of the objective function. As a result, there are usually multiple local maxima, and the standard Karush-Kuhn-Tucker conditions are not helpful to find the global maximum.

I propose a novel method to solve this problem. The key idea is to find an alternative optimization problem that is easy to solve and whose solution also solves $\mathcal{P}$. This alternative problem, denoted

\(^4\)Define a cycle of operating firms as a sequence operating firms $\{s_1, \ldots, s_k\}$, for some $k \geq 1$, such that $\Omega_{s_i, s_{i+1}} > 0$ for $i \in \{1, \ldots, k - 1\}$ and $\Omega_{s_k, s_1} > 0$. A firm $j$ has access to at least one input if there exists a sequence of operating firms $\{t_1, \ldots, t_m\}$, for some $m \geq 1$, such that 1) $t_1$ is part of an operating cycle, 2) $\Omega_{t_i, t_{i+1}} > 0$ for $i \in \{1, \ldots, m - 1\}$ and 3) $t_m = j$.

\(^5\)\(\mathcal{P}\) 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 (Garey and Johnson, 1990).
by $R$, is obtained by *relaxing and reshaping $P$.\footnote{To my knowledge, reshaping the planner’s problem is the only practical way of solving these network formation problems when $n$ is large. An alternative algorithm that involves computing output for each possible network is discussed below but it is unhelpful for economies with more than twenty firms or so. Casual experimentation with genetic algorithms and branch-and-bound algorithms has found them slow and unreliable for these problems.} It is defined below alongside $P$ to highlight their differences.

**$P$: Original planner’s problem**

$$\max_{\theta \in \{0,1\}^n} Q \left( 1 - \sum_{j \in N} \theta_j f_j \right) L$$

where $q$ solves, for each $j \in N$,

$$q_j = z_j \theta_j A \left( \sum_{i \in N} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\alpha_j \varepsilon_j}$$  \hspace{1cm} (5)

**$R$: Relaxed and reshaped problem**

$$\max_{\theta \in [0,1]^n} Q \left( 1 - \sum_{j \in N} \theta_j f_j \right) L$$

where $q$ solves, for each $j \in N$,

$$q_j = z_j \theta_j A \left( \sum_{i \in N} \theta_i^{b_{ij}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\alpha_j \varepsilon_j}$$  \hspace{1cm} (7)

$R$ differs from $P$ in two important ways—emphasized in bold and blue above. First, the binarity constraint is relaxed and $\theta$ can now take values on the *inside* of the unit hypercube, $[0,1]^n$. While this relaxation has the advantage of convexifying $P$’s feasible set, it also augments the planner’s problem with points that have no real economic meaning. For instance, $\theta_j = 0.5$ does not correspond to any physical state in the environment. Since they have no interest on their own, we can change the value of the objective function over these new points—and only over these new points—to help us solve $P$. This is done in (7), a transformed version of (5), which includes the reshaping constants $a_j > 0$ and $b_{ij}$. These constants modify the shape of the optimization problem everywhere except over $P$’s original feasible set: $\{0,1\}^n$. Indeed, for $\theta \in \{0,1\}^n$, $\theta^{\varepsilon_j} = \theta_j$ for all $j$. Similarly for $b_{ij}$, if $\theta_i = 0$ then $q_i = 0$ anyway, and if $\theta_i = 1$ then $\theta_i^{b_{ij}} = 1$.\footnote{For (7) to be well-defined as $\theta \to 0^+$, I impose $b_{ij} \geq -a_j \left( \varepsilon_j - 1 \right)$ for all $i,j$. This inequality is satisfied by (\footnote) below.} In both cases, the term in the summation is unchanged. This reshaping procedure therefore preserves the ranking, in terms of utility, of the corners $\{0,1\}^n$—the only points with actual economic meaning—while elsewhere changing the shape of the optimization problem.

Crucially, we are free to pick the constants $a_j$ and $b_{ij}$ to help us solve the planner’s problem. In particular, we can pick these constants to increase the concavity of $R$ with the goal of removing the undesirable local maxima that prevent an easy resolution of the relaxed problem. On the other hand, too much concavity can create a new global maximum somewhere in the middle of $[0,1]^n$, in which case the solutions of $P$ and $R$ would clearly differ. Specific values for $a_j$ and $b_{ij}$ provide the right balance and are needed for the results of this section to hold. From now on, I therefore set

$$a_j = \frac{1}{\sigma - 1} \quad \text{and} \quad b_{ij} = 1 - \frac{\varepsilon_j - 1}{\sigma - 1}. \hspace{1cm} (\ast)$$

The following propositions establish conditions under which $R$ can be solved easily.
Proposition 2. Let $\varepsilon_j = \varepsilon$ and $\alpha_j = \alpha$. If $\Omega_{ij} = d_i e_j$ for some vectors $d$ and $e$ then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $P_{RR}$.

A similar result holds for a different set of $\Omega$'s. Define $\bar{\Omega} = \omega(\mathbf{1}_n - I_n)$ where $\mathbf{1}_n$ is the all-one matrix, $I_n$ is the identity and $\omega > 0$. The matrix $\bar{\Omega}$ describes a network of potential connections in which firms are connected to each other, but not with themselves, with the same intensity $\omega$.

Proposition 3. Let $\sigma = \varepsilon_j$ for all $j$. Suppose that the $\{\beta_j\}_{j \in \mathcal{N}}$ are not too far from each other and that the fixed costs $f_j > 0$ are not too big. If the matrix $\Omega$ is close enough to $\bar{\Omega}$, then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to $R$.

Propositions 2 and 3 establish conditions under which a feasible point $\theta^{*}$ that satisfies the first-order conditions and the complementary slackness conditions solves $R$. As a result, standard numerical algorithms, such as gradient ascent, can rapidly solve $R$. With a solution to $R$ in hand, the next proposition establishes conditions under which that solution also solves $P$.

Proposition 4. If a solution $\theta^{*}$ to $R$ is such that $\theta^{*} \in \{0, 1\}^n$ then $\theta^{*}$ also solves $P$.

This result follows directly from the fact that the feasible set of $R$ contains the feasible set of $P$ and that both of their objective functions coincide over $\{0, 1\}^n$ by construction.

Together, Propositions 2 to 4 offer a convenient way to solve $P$. First, find a solution $\theta^{*}$ to $R$ using a standard algorithm for convex problems. If $\theta^{*}$ belongs to $\{0, 1\}^n$ then it also solves $P$. This last condition can be tested in practice, but reshaping $P$ would not be very useful if $\theta^{*}$ rarely belonged to $\{0, 1\}^n$. Fortunately, condition $(\star)$ is such that solutions to $R$ are naturally pushed toward $\{0, 1\}^n$.

The following proposition is critical for this result.

Proposition 5. The (net) marginal benefit of increasing $\theta_j$ only depends on $\theta_j$ through aggregates.

These aggregates are summations over many firms. As the number of firms increases, they become more and more independent of $\theta_j$, and so does the marginal benefit. To see why this pushes solutions toward $\{0, 1\}^n$, consider a gradient ascent algorithm that begins at $\theta_j = 1/2$. If, for instance, the marginal benefit is positive, the planner increases $\theta_j$. But since the marginal benefit itself is independent of $\theta_j$, it remains positive and the planner keeps increasing $\theta_j$ until it reaches 1. The opposite happens if the marginal benefit is initially negative. As a result, the solution $\theta^{*}$ to $R$ is in $\{0, 1\}^n$ and Proposition 4 guarantees that $\theta^{*}$ also solves $P$.

---

8To be precise, let $\bar{\beta}$ be a $n \times 1$ vector such that $\bar{\beta}_i = \bar{\beta}_j$ for all $i, j$. Then there exists a ball $B = \{(\Omega, \beta): \| (\Omega, \beta) - (\bar{\Omega}, \bar{\beta}) \| < \delta \}$ for $\delta > 0$ such that the statement holds for $(\Omega, \beta) \in B$. An earlier version of this paper included a variant of Proposition 3 for binary matrices $\Omega$.

9The assumptions about $\Omega$ in Propositions 2 and 3 impose that the firms’ input sets are not too different from each other. For instance, $\Omega_{ij} = d_i e_j$ implies that the factor intensity vectors of the firms (the columns of $\Omega$) are proportional to each other. Appendix E provides a fast algorithm to solve $R$.

10The proof of Proposition 5 shows that the relevant aggregates are $Q$ and $B_j = \left( \sum_{i \in \mathcal{N}} \theta_i \Omega_{ij} q_{ij}^{(i)} \right)^{1/(e_j - 1)}$ for all $j$. As the number of firms increases, the number of positive entries in $\Omega$ must also grow otherwise the marginal benefit of increasing $\theta_j$ might still depend on $\theta_j$. Fortunately, these cases are identified easily in practice: the numerical solution to $R$ will be outside of $\{0, 1\}^n$. But even in these more pathological cases, Appendix B.3 shows that the solution approach makes minimal errors in aggregate output.
**Example with two firms.** To better understand how the solution approach works, consider an economy with two firms. The objective function $V(\theta)$ of the relaxed planner’s problem without any reshaping is shown in Figure 2a. $V$ 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 global maximum is at $(1, 0)$. Because of the non-concavity, standard algorithms can get trapped in a local maximum and cannot reliably solve this problem.

Figure 2b shows the objective function $V_R(\theta)$ of the reshaped problem. Three things are worth noticing. First, $V$ and $V_R$ coincide at the corners $\{0, 1\}^2$. As a result, the ranking of these corners, in terms of utility, is the same in both problems. Second, the reshaping stretches the objective function so that $V_R$ is concave. The first-order conditions are therefore sufficient. Third, the reshaping did not create another maximum somewhere in the inside of $[0, 1]^2$, and the maximum of $V_R$ is also the maximum of $V$. As a result, starting from any $\theta$ in $[0, 1]^2$, standard algorithms like gradient ascent will converge to the global maximum. This point also solves $P$ by Proposition 4.

![Figure 2](image.png)

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

Figure 2: Reshaping the planner’s problem in a simple economy

### 3.1 Numerical tests

The theoretical results of the last section provide sufficient conditions under which reshaping the planner’s problem solves $P$, but these conditions are far from being necessary. Here, I show through numerical simulations that the solution approach also works well when these conditions are relaxed. To do so, I randomly draw a large number of economies and compare, for each of them, the solutions to the reshaped problem $R$ and the original problem $P$. Since there is a finite number of vectors $\theta$ in the feasible set of $P$, we can try them all to find the correct solution. This brute-force approach is however limited to economies with only a few firms. Since there are $2^n$ possible vectors $\theta$ in $\{0, 1\}^n$, the large number of possibilities quickly becomes infeasible to handle as $n$ increases.

Appendix B provides the details of the simulations. They involve a broad range of economies that
cover all the dimensions of heterogeneity allowed by the model. They also cover matrices $\Omega$ with different shapes and various degrees of sparsity. The results are presented in Table 1. We see that reshaping the planner’s problem (first two columns) attributes the correct status $\theta$ to more than $99.9\%$ of the firms. It also finds output values that are within $0.001\%$ of their correct value. In contrast, without reshaping (last two columns) about $13\%$ of the firms are assigned the wrong operational status and the average error in output can reach above $0.8\%$, a large number when studying aggregate fluctuations.\(^\text{11}\)

Table 1: Testing the solution approach numerically

<table>
<thead>
<tr>
<th>$n$</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct $\theta$</td>
<td>Error in $C$</td>
</tr>
<tr>
<td>8</td>
<td>99.9%</td>
<td>0.001%</td>
</tr>
<tr>
<td>10</td>
<td>99.9%</td>
<td>0.001%</td>
</tr>
<tr>
<td>12</td>
<td>99.9%</td>
<td>0.001%</td>
</tr>
<tr>
<td>14</td>
<td>99.9%</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

Notes: See Appendix B for the details of the simulations.

4 Economic forces shaping the network

We now explore how economic forces shape the production network. Here, the benefit of producing with a broader variety of inputs is key.

Highly-connected firms are more likely to operate. As discussed earlier, (5) implies that having access to an additional supplier increases a firm’s productivity $q$. This gain in productivity is then propagated downstream through the production network. Through this mechanics, the benefit from input variety encourages the planner to operate firms that are well-connected in terms of both direct and indirect neighbors (second neighbors, etc.). Indeed, a firm with many upstream suppliers enjoys a larger $q$, which the planner values, while a firm with numerous downstream producers increases the productivity of its many (perhaps indirect) customers. As a result, these firms are valuable to the planner, which is then more likely to pay the fixed cost to operate them, as the following proposition shows.

**Proposition 6.** In a large economy, operating a firm (weakly) increases the incentives to operate its (perhaps indirect) suppliers and customers in $\Omega$.\(^\text{12}\)

\(^\text{11}\)To further test the robustness of the solution approach, Appendix B provides several additional exercises that include economies 1) with a large number of firms, 2) with individual link formation, and 3) for which the solution to $R$ is not in \{0, 1\}\(^n\). Together with the results of Table 1, these extensive tests show that the solution approach performs well in a broad class of economic environment.

\(^\text{12}\)An economy is large if the operating decision of any individual firm has no impact on the Lagrange multiplier $w$ associated with the labor resource constraint (4). An indirect customer (supplier) in $\Omega$ is a firm that can be reached by moving downstream (upstream) in the network described by $\Omega$. 
Cascades of firm shutdowns can arise. One immediate consequence of this result is that the efficient allocation exhibits complementarities between the operating decisions of nearby firms. These complementarities, in turn, can generate cascades of firm shutdowns. Consider for instance a firm \( j \) 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 shut down. But then \( j \)'s second neighbors are also losing a neighbor and are at a greater risk of shutting down themselves. Since the same logic applies to further neighbors of \( j \), the initial \( z \) shock can trigger a wave of shutdowns that spreads through the economy.

Clustering of economic activity. Since having many neighbors is valuable from an efficiency perspective, the planner tends to organize production into tightly connected clusters of operating firms. Figure 3a presents an example of this process by showing the efficient allocation in an economy under two typical random productivity vectors \( z \). We see that, to maximize the benefit from input variety, the planner clusters economic activity in either the top or the bottom group of firms. In addition, the active cluster tends to include the firm with the highest \( z \). By creating an ecosystem around an important producer, the planner maximizes its positive impact on the economy.

Figure 3: Economic forces shape the network

Proposition 7 shows that the planner prefers to operate groups of firms that are highly connected.

**Proposition 7.** The incentives of the planner to operate a group of firms (weakly) increase with additional potential connections between them.
Small shocks can lead to large reorganizations. In the efficient allocation, a small change in the environment can trigger a large reorganization of the network. When designing the network, the planner compares the $2^n$ possibilities and selects the one providing the highest utility. As, say, a firm’s TFP $z$ declines there is a point at which the planner shuts the firm down. But because of the complementarities in operating decisions, it might be better to shut down the whole cluster around the firm and to move production elsewhere. Figure 3b provides an example. Both economies are identical except for the productivity $z$ of the red firm which is slightly larger in the economy on the left. While the drop in $z$ from left to right is negligible, it triggers a large reorganization of the network. Aggregate output, however, is barely affected. Indeed, the planner reorganizes the network precisely to limit the negative impact of the shock on output.\footnote{This is a standard result from the Theorem of the Maximum. Note that while aggregate output barely changes, firm-level distributions can change substantially. The discontinuous behavior of the model would be mitigated if the fixed costs were paid when changing the status of a firm instead.}

5 Quantitative example

To illustrate how the solution approach can be used in practice, this section provides a basic calibration of the model and shows that it captures salient features of the data, such as cascades of firm shutdowns and changes in the shape of the production network over the business cycle.

The calibration uses data about the U.S. firm-level production network from the Factset Revere Supply Chain Relationships Data. These data cover the period from 2003 to 2016, and include almost 13,000 firms and more than 40,000 relationships in a typical year. The details of the calibration are in Appendix C but the main ingredients are as follows. The number of firms is set to $n = 1000$ as a good trade-off between realism and computation time. Firms share the same parameters but they differ in their productivity $z$, which follows an AR(1) process. Since the model itself is static, the persistence of that process is the only inter-temporal linkage in the economy. All parameters are taken directly from the literature with the exception of the matrix $\Omega$, which is drawn randomly to match properties of the in-degree (number of customers) distribution in the data.

5.1 Cascades of firm shutdowns

Cascades of firm shutdowns arise in this economy because of the complementarities between operating decision of nearby firms. To evaluate the importance of these cascades, I regress the fraction of each firm’s neighbors that shut down in a given period on whether the firm itself shuts down.\footnote{In the data, I consider that a firm shuts down during the last year that it is in the sample.} I run separate regressions for upstream and downstream neighbors at various distances from the original firm. The estimated coefficients capture the increase in shutdown probability associated with an exiting neighbor.\footnote{To be precise, denote by $\text{DX}_{jdt}$ the fraction of firm $j$’s downstream neighbors located at a distance $d$ that exit between $t$ and $t + 1$. I regress $\text{DX}_{jdt} = a_d\text{Exit}_{jt} + \text{Controls}_{jt} + \varepsilon_{jdt}$, where $\text{Exit}_{jt}$ is 1 if $j$ exits between $t$ and $t + 1$ and 0 otherwise.}
Figure 4 shows the estimated coefficients in the Factset data (green dashed lines). We see that the shutdown of a firm is associated with about a 10% increase in the probability that one of its direct neighbors also exits. This number falls to about 2% for the second neighbors and keeps declining afterwards. The model (solid blue lines) is roughly able to match these patterns, suggesting that it broadly captures the joint operating decisions of nearby firms.$^{16}$

![Figure 4: Cascades of firm shutdowns in the model and in the data](image)

Notes: Factset data. Estimated coefficients from regressing the fraction of exiting neighbors on whether a firm exits. Time, in-degree and out-degree controls are included. The distance is the smallest number of connections between two firms.

Cascades can arise from any firm, but the shutdown of well-connected producers trigger larger reorganizations of the network. To see this, we can first compute the size of a cascade as the total number of shutdowns, summed up to the second neighbors, associated with the exit of a firm, and then compare this statistic across firms with different numbers of connections (degrees). The first column of Table 2 shows that in the data firms that are above the 95th percentile of the degree distribution are associated with cascades that are about three times larger than those associated with the average firm. High-degree firms are, however, less likely to actually shut down in response to shocks, as the third column of Table 2 shows. In the data, an average firm has a 11.8% chance of exiting in a given year, while this number drops to 2.0% for high-degree firms. The model, despite its simplicity, does well in terms of the size of the cascades and is also able to roughly replicate the exit probabilities. In the model, high-degree firms are particularly valuable to the planner, which keeps them in operation after most negative shocks. When they do shut down, however, the planner reorganizes the whole cluster of producers that was built around them, which explains the large cascades they trigger.

---

$^{16}$One possible worry is that Figure 4 captures common shocks across firms instead of propagation through the network. For instance, since trading partners are likely to be geographically close to each other, a local shock like an earthquake could directly affect both of them at the same time. To alleviate this concern, I run the same regressions on supplier/customer pairs located in different zip codes. Reassuringly, the outcome of these regressions are close to those of Figure 4. In careful empirical work, Barrot and Sauvagnat (2016) and Carvalho et al. (2016) also show clear propagation of shocks through supply chains in the data.
Table 2: High-degree firms are more resilient but create larger cascades

<table>
<thead>
<tr>
<th></th>
<th>Size of cascades</th>
<th>Probability of exit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>Average firm</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>High-degree firm</td>
<td>3.1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

*Notes:* “High degree firms” are above the 95th percentile of the degree distribution. “Size of cascades” corresponds to the sum of exiting firms up to the second neighbors downstream and upstream, computed by multiplying the regression coefficients in Figure 4 by the number of neighbors at a given distance.

5.2 Aggregate fluctuations

There are no aggregate shocks in the model but, since the number of firms is finite, aggregate fluctuations arise from the idiosyncratic shocks to $z$. In this section, we consider how these microeconomic shocks give rise to correlations between the shape of the network and aggregate output. We also investigate how the endogenous reorganization of the network affects the aggregation of these shocks into macroeconomic fluctuations.\(^\text{17,18}\)

**Comovements.** I focus on three key moments to describe how the shape of the network changes over the business cycle. The first two moments are the in-degree (number of suppliers) and out-degree (number of customers) distributions. In the model and in the data, these distributions are close to power laws so that their exponent parameters provide a good description of the full distributions. As is now well-known, these exponents have an important influence for the aggregate impact of idiosyncratic shocks (Acemoglu et al., 2012).\(^\text{19}\) The third moment is the global clustering coefficient which measures how tightly connected firms are with one another—a key moment given the importance of clustering for productivity.\(^\text{20}\)

Table 3 shows that in the data the exponent parameters of the degree distributions are negatively correlated with output, which indicates thicker tails and an abundance of well-connected firms during expansions. The economy also features more clustering during booms. The model is able to roughly match these patterns. When well-positioned firms receive good $z$ shocks, the planner builds dense

\(^{17}\)In a richer model, firms could have to pay a cost to change their operating status. These costs would limit the reaction of the network in response to temporary $z$ shocks. They would also make the model dynamic, adding significant technical difficulties given the size of the state space in this economy.

\(^{18}\)Adding aggregate productivity shocks to the model does not change the main conclusions of this section. Proposition 9 in Appendix C.5 shows that the network itself is invariant to these shocks and that they simply add an exogenous term to output volatility. In addition, Online Appendix C.5 provides a version of the calibrated economy with a much larger number of firms and aggregate shocks. The results are similar to those in the main text.

\(^{19}\)The out-degree distribution in the calibrated economy is closely related to the centrality distribution, which is the focus of Acemoglu et al. (2012). Conducting the exercises of this section with the centrality distribution instead yields similar results.

\(^{20}\)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 clustering coefficients by multiplying them by the square root of the number of nodes. This normalization allows for a better comparison of networks of different sizes.
ecosystems of suppliers and customers around them. These clusters are highly productive which generates the correlations between output, clustering and the degree distributions. Inversely, in recessions it might be too costly to organize these productive clusters—perhaps because a few critical firms face low $z$’s. As a result, economic activity is more dispersed and output is lower.\footnote{Appendix C.5 shows that these correlations are also in the Compustat data, an alternative dataset which covers fewer firm-level relationships but over a longer time period.}

Table 3: Correlation between the network and aggregate output

<table>
<thead>
<tr>
<th>Network</th>
<th>In-degree</th>
<th>Out-degree</th>
<th>Clustering coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>-0.87</td>
<td>-0.97</td>
<td>0.72</td>
</tr>
<tr>
<td>Model</td>
<td>-0.59</td>
<td>-0.70</td>
<td>0.53</td>
</tr>
</tbody>
</table>

\textit{Notes:} All time series are in logs. In the data, output is annual real GDP. Output is detrended linearly. Since there are only 13 years in the Factset data 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.

**Level and volatility of output.** We now consider the impact of the endogenous formation of the network on aggregate output. To do so, it is useful to compare the efficient allocation, in which the network is constantly reorganized in response to shocks, to an alternative economy in which the network is designed efficiently in the first period but then kept completely fixed. The differences between these two economies capture the role played by the endogenous network mechanisms. Doing this exercise, I find that aggregate output is 11\% lower and 20\% more volatile when the network is kept fixed.\footnote{There exist parametrizations under which the flexible network economy is more volatile than its fixed network counterpart. For reasonable changes in the parameters of the calibration, however, the flexible network economy remains the least volatile. Note also that the 20\% difference in volatility is conservative in the sense that other parametrizations, in the range of values found in the literature, lead to much bigger differences.}\footnote{When, in addition to the network itself, all the inputs of the firms are kept fixed, output volatility increases by 40\% compared to the flexible network benchmark. On its own, the endogenous formation of the network is therefore able to explain half of the reduction in volatility generated by all the adjustment margins together.} These numbers highlight the importance of the mechanism for the aggregation of microeconomic shocks into macroeconomic fluctuations. They also suggest that policies or frictions that would impede the reorganization of the network might lead not only to a loss in output but also to an increase in its volatility.\footnote{Additional exercises can be found in the Appendix. Appendix C.4 shows that the endogenous formation of the network leads to degree distributions with thicker tails and to more clustering. Online Appendix G discuss the impact of $\Omega$ on the shape of the network.}

6 Conclusion

U.S. data suggest that the firms’ extensive margin of operation plays an important role in shaping the production network. This paper proposes a novel approach to solve these network formation problems. The methodology involves reshaping the problem of the social planner such that it can be solved...
easily using standard techniques. Analytical results and numerical simulations find that this approach is successful in a broad family of economies. A basic calibration shows that the model generates cascades of firm shutdowns and correlations between aggregate output and the shape of the network that are consistent with the data. It also suggests that considering the impact of microeconomic shocks on the shape of the production network might be of first-order importance for aggregate volatility. The solution approach introduced in this paper might be of independent interest and could be useful to tackle other non-convex optimization problems with discrete choices.

References


Appendix

A Equilibrium

This section shows that the efficient allocation chosen can also be sustained as an equilibrium. The key equilibrium concept is that of stability: an allocation is a stable equilibrium if no coalition of firms wishes to deviate. This equilibrium definition has proven to be particularly convenient in network economies (Jackson and Wolinsky, 1996; Hatfield et al., 2013). The approach followed here is most closely related to Oberfield (2018).

Before formally defining an equilibrium, we describe the market structure in this economy. A contract between two firms $i$ and $j$ is a pair $\{x_{ij}, T_{ij}\}$ where $x_{ij}$ is a quantity shipped from $i$ to $j$, and $T_{ij}$ is a payment from $j$ to $i$. An arrangement is a contract between all possible pairs of firms $\{x_{ij}, T_{ij}\}_{i,j \in N^2}$.

Under a given arrangement, a firm $j$ decides on a price $p_j$ to charge the household, an amount $c_j$ to sell to the household, how much labor $l_j$ to employ and whether to produce or not ($\theta_j$). It picks these quantities to maximize its profits

$$\pi_j = p_j c_j - w l_j + \sum_{i \in N} T_{ji} - \sum_{i \in N} T_{ij} - w \theta_j f_j L,$$

subject to a technology constraint

$$c_j + \sum_{k \in N} x_{jk} \leq y_j,$$

and the household demand for its good

$$c_j = \beta_j C \left( \frac{p_j}{P} \right)^{-\sigma}$$

where $P = \left( \sum_j \beta_j P_j^{1-\sigma} \right)^{1/(1-\sigma)}$ is the usual price index and where $w$ is the wage. An allocation is feasible if the technology constraints (9) and the labor resource constraint $\sum_j l_j + \sum_j \theta_j f_j L \leq L$ are satisfied.

A coalition is a set of firms $J$. A deviation for a given coalition $J$ consists of (i) dropping any contracts that involve at least one firm in $J$ and (ii) altering the terms of any contract involving a buyer and a supplier that are both members of the coalition. A dominating deviation for a given coalition is a deviation that delivers at least the same amount of profits to all members of the coalition and strictly greater profits to at least one member.

We can now define an equilibrium in this environment.

**Definition 1.** A stable equilibrium is an arrangement $\{x_{ij}, T_{ij}\}_{i,j \in N^2}$, firms’ choices $\{p_j, c_j, l_j, \theta_j\}_{j \in N}$ and a wage $w$ such that (i) given the wage, total profits, and prices, the consumption choices $\{c_j\}_{j \in N}$ maximize the utility of the representative household; (ii) for each $j \in N$, $\{p_j, c_j, l_j, \theta_j\}$ maximizes the
profits of $j$ given the arrangement, the wage, the household demand and the technology constraint; (iii) labor and final goods market clear; (iv) there are no dominating deviations available to any coalition; and (v) the equilibrium allocation is feasible.

The following proposition shows how stable equilibria and the efficient allocation are related.

**Proposition 8.** Every stable equilibrium is efficient.

This proposition shows that the efficient network can be thought of as arising from the individual decisions of private agents, without the need for taxes or subsidies.

## B Numerical tests

This appendix provides the details of the numerical simulations of Section 3.1 as well as several additional exercises to show the robustness of the solution approach.

### B.1 Details of the simulations of Section 3.1

The numerical simulations of Section 3.1 involve a large number of economies that are generated randomly from a broad set of parameters. The aggregate parameters are selected from $n \in \{8, 10, 12, 14\}$ for the number of firms and $\sigma \in \{4, 6, 8\}$ for the elasticity of substitution. The matrix $\Omega$ is such that each firm has on average 3, 4, 5, 6, 7 or 8 potential incoming connections (non-zero $\Omega_{ij}$). The corresponding average numbers of active incoming connections are 2.1, 3.0, 3.8, 4.5, 5.3 and 5.8, respectively. Each non-zero element in $\Omega$ is drawn from $\Omega_{ij} \sim U([0, 1])$. Appendix E.1 describes the precise algorithm used to build $\Omega$.

The firm-level parameters are drawn from $\log(z_k) \sim \text{iid } \mathcal{N}(0, 0.25^2)$ for the productivities, $f_j \sim \text{iid } U([0, 0.2/n])$ for the fixed costs, $\alpha_j \sim \text{iid } U([0.25, 0.75])$ for the intermediate input shares, $\varepsilon_j \sim \text{iid } U([4, 8])$ for the elasticities of substitutions between inputs, and $\beta_j \sim \text{iid } U([0, 1])$ for the factor intensities of the final good producer.

For every potential combination of the aggregate parameters I run 200 simulations. In each case, the matrix $\Omega$ and the individual characteristics of the firms are drawn from the distribution described above. I then use the exhaustive search algorithm described in Appendix E.2 to compute the true solution to $\mathcal{P}$. I also use the algorithm of Appendix E.4 to compute solutions to the reshaped and non-reshaped versions of the planner’s problem. These two solutions are then compared to the true solution and the results are reported in Table 1. An economy is kept in the sample only if the first-order conditions of the reshaped problem yield a solution in $\{0, 1\}^n$.\(^{25}\)

---

\(^{25}\)More than 90% of the economies are kept in the sample. Appendix B.3 shows the outcome of these tests when no simulations are excluded.
B.2 Economies with a large number of firms

The simulations of Section 3.1 involve a small number of firms but the solution approach also works well when the economy features a large number of firms. While the true solution to the planner’s problem is unknown in these cases, we can verify whether there exist welfare-improving deviations from the solution to the reshaped problem. In particular, I verify whether changing the status \( \theta \) of each firm improves welfare. I keep repeating this procedure as long as there are deviations to be found, and then compare this deviation-free solution to the original one given by the reshaping approach. The precise algorithm is described in Appendix E.3.

Since this procedure is computationally costly, I only consider large economies that follow the calibration of Section 5. The results are presented in Table 4. Again, the reshaping approach performs very well. After all the possible deviations are accounted for, more than 99.9% of the firms have kept the same \( \theta \) and aggregate output has changed by a negligible amount.\(^{26}\) In contrast, without reshaping more than 30% of firms are assigned the wrong \( \theta \), and the error in aggregate output amounts to about 0.6%. While this test does not guarantee that the solution approach finds the correct efficient allocation, it provides a good indication that there are no obvious mistakes in its solution.

Table 4: Testing the reshaping approach for large networks

<table>
<thead>
<tr>
<th>n</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>&gt; 99.9%</td>
<td>&lt; 0.001%</td>
</tr>
</tbody>
</table>

**Notes:** The parameters are as in the calibrated economy of Section C.2. I simulate 100 different matrices \( \Omega \) and, for each \( \Omega \), draw 100 productivity vectors \( z \). I run the procedure described in Appendix E.3 on each of them and report average results. \( x < 0.001\% \) indicates that \( x > 0 \) but that proper rounding would yield 0, and similarly for \( x > 99.9\% \).

B.3 When the solution to the reshaped problem is not in \( \{0,1\}^n \)

The results presented in Table 1 exclude economies in which the first-order conditions of \( R \) are satisfied at a vector \( \theta \) such that \( 0 < \theta_j < 1 \) for at least one firm \( j \), which happens in less than a tenth of the simulations. This section shows that even when these cases are not excluded from the sample reshaping the problem of the planner proves to be a valuable tool. To see this, Table 5 shows the same tests as Table 1 but without excluding any economies. The table also breaks down the results by sparsity of the matrix \( \Omega \). “Less connected \( \Omega \)’s” refers to matrices \( \Omega \) with 3, 4 or 5 potential incoming connections on average while “More connected \( \Omega \)’s” refers to matrices \( \Omega \) with 6, 7 or 8 potential connections on average.\(^{27}\) We learn a few things from this table. First, even when no simulations are

\(^{26}\)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 4.

\(^{27}\)The corresponding average number of active connections is 2.9 for less connected \( \Omega \)’s and 5.2 for more connected \( \Omega \)’s.
excluded, the reshaping algorithm performs well. On average, the error in aggregate output \( C \) is only 0.009\%. In contrast, without reshaping the error is on average 0.77\%—86 times more. Second, the error in \( C \) is smaller when the matrix \( \Omega \) has fewer zeros in it. This is in line with the analytical results of Section 3 (see footnote 10). Third, as the number of firms \( n \) increases the errors in \( C \) stay roughly constant. In contrast, without reshaping the errors grow. This suggests that reshaping the planner’s problem becomes increasingly useful as the the economy becomes larger.\footnote{Finally, note that we can see whether \( \theta \in \{0,1\}^n \) or not when solving the problem. If extreme precision is needed, we can be extra careful when \( \theta \not\in \{0,1\}^n \) and use additional tests (for instance we can look for beneficial deviations) to check the robustness of the solution.}

Table 5: Testing the reshaping on small networks without exclusions: Error in aggregate output

<table>
<thead>
<tr>
<th>( n )</th>
<th>Reshaping?</th>
<th>All ( \Omega )’s</th>
<th>More connected ( \Omega )’s</th>
<th>Less connected ( \Omega )’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Yes</td>
<td>0.007%</td>
<td>&lt; 0.001%</td>
<td>0.014%</td>
</tr>
<tr>
<td>No</td>
<td>0.683%</td>
<td>0.640%</td>
<td>0.726%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>0.013%</td>
<td>&lt; 0.001%</td>
<td>0.027%</td>
</tr>
<tr>
<td>No</td>
<td>0.781%</td>
<td>0.739%</td>
<td>0.823%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Yes</td>
<td>0.008%</td>
<td>&lt; 0.001%</td>
<td>0.016%</td>
</tr>
<tr>
<td>No</td>
<td>0.799%</td>
<td>0.744%</td>
<td>0.853%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Yes</td>
<td>0.008%</td>
<td>0.001%</td>
<td>0.016%</td>
</tr>
<tr>
<td>No</td>
<td>0.831%</td>
<td>0.801%</td>
<td>0.862%</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Same as Table 1 except that all simulations are kept in the sample. “Less connected \( \Omega \)” refers to \( \Omega \)’s in which firms have, on average, 3, 4 or 5 potential incoming connections. “More connected \( \Omega \)” refers to \( \Omega \)’s in which firms have, on average, 6, 7 or 8 potential incoming connections. \( x < 0.001\% \) indicates that \( x > 0 \) but that proper rounding would yield 0.

Table 6 shows the percentage of correct \( \theta \) from the same simulations. Again the results show that the solution approach works best on more connected matrices \( \Omega \), but it always outperforms the no-reshaping numbers by large margins.

A similar exercise can also be done for economies with a large number of firms. This exercise is analogous to that of Table 4 and is presented in Table 7. We see that even when the first-order conditions of \( \mathcal{R} \) yield a solution \( \theta \) such that \( 0 < \theta_j < 1 \) for at least one \( j \) the errors in aggregate output are negligible. This last test suggests that the solution approach works particularly well in realistic economies with a large number of firms. Indeed, in general only a small number of firms are away from the \( \{0,1\} \) bounds. These firms are usually somewhat isolated in the network and their impact on aggregate outcome is limited.

### B.4 Formation of the network link by link

While the paper focuses on the role of the firms’ extensive margin of operation for the formation of the network, the model is general enough to accommodate networks that are formed link by link. In this context, it is useful to think of a link as a \textit{pseudo} firm that 1) has a single potential supplier
Table 6: Testing the reshaping on small networks without exclusions: Firms with correct $\theta$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Reshaping?</th>
<th>All $\Omega$’s</th>
<th>More connected $\Omega$’s</th>
<th>Less connected $\Omega$’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Yes</td>
<td>99.8%</td>
<td>99.9%</td>
<td>99.6%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>88.2%</td>
<td>89.1%</td>
<td>87.4%</td>
</tr>
<tr>
<td>10</td>
<td>Yes</td>
<td>99.7%</td>
<td>99.9%</td>
<td>99.5%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>86.5%</td>
<td>87.3%</td>
<td>85.8%</td>
</tr>
<tr>
<td>12</td>
<td>Yes</td>
<td>99.7%</td>
<td>99.9%</td>
<td>99.5%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>86.2%</td>
<td>87.0%</td>
<td>85.5%</td>
</tr>
<tr>
<td>14</td>
<td>Yes</td>
<td>99.7%</td>
<td>99.9%</td>
<td>99.4%</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>85.5%</td>
<td>86.1%</td>
<td>85.1%</td>
</tr>
</tbody>
</table>

Notes: Same as 1 except all simulations are kept in sample. “Less connected $\Omega$” refers to $\Omega$’s in which firms have, on average, 3, 4 or 5 potential incoming connections. “More connected $\Omega$” refers to $\Omega$’s in which firms have, on average, 6, 7 or 8 potential incoming connections. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0.

Table 7: Testing the reshaping on large networks without exclusions

<table>
<thead>
<tr>
<th>$n$</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct $\theta$</td>
<td>Error in $C$</td>
</tr>
<tr>
<td>1000</td>
<td>&gt; 99.9%</td>
<td>&lt; 0.001%</td>
</tr>
</tbody>
</table>

Notes: Same as Table 4 except all simulations are kept in the sample. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0, and similarly for $x > 99.9\%$.

and a single potential customer, such that $\sum_i (\Omega_{ij} > 0) = \sum_i (\Omega_{ji} > 0) = 1$, and 2) produces a good that is not included in the production of the final good, such that $\beta_j = 0$. In this context, the cost $f_j$ of this “link firm” is the fixed cost of operating the link, its productivity $z_j$ affects the variable cost of shipping goods through that link and $\alpha_j$ controls how labor intensive the shipping technology is.

This appendix provides the results of two exercises that show that reshaping the planner’s problem is also useful in finding the correct production network when we allow for the formation of individual links. In both exercises, the economy contains $m$ real firms that are always active ($f_j = 0$). Any two of these real firms are connected to each other by a link with some probability $p > 0$ that varies across exercises (i.e. for any ordered pair of real firms $i, j$; with probability $p$ there exists a “link firm” $k$ such that $\Omega_{ik} > 0$ and $\Omega_{kj} > 0$). There are no other connections in $\Omega$.

Individual link formation in small networks

When the number $m$ of real firms is small, we can use the same approach as in Section 3.1 and find the true solution to the planner’s problem by comparing the welfare provided by each possible network $\theta$ (see algorithm in Appendix E.2). There are at most $m(m - 1)$ links in an economy, in

---

29 This assumption is made to focus on the link formation aspect of the problem. I have experimented with economies in which $f_j > 0$ for the real firms and the results are similar.
which case $2^m(m-1)$ networks have to be compared. Since this quantity grows rapidly with $m$, Table 8 shows the results of these tests when there are only $m \in \{3, 4, 5\}$ real firms. As before, the outcome of this exhaustive search is compared to the allocation found by reshaping the planner’s problem.

We see from Table 8 that the reshaping algorithm works well. Over all the simulations, more than 99.7% of the links are assigned the proper operating status $\theta$ and the errors in aggregate output are small. Without reshaping, large fractions of the links are assigned the wrong operating status and the errors in aggregate output can be sizable.

Table 8: Individual links formation with few firms

<table>
<thead>
<tr>
<th>Number of firms</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real firms $m$</td>
<td>Link firms $n - m$</td>
<td>Correct $\theta$</td>
</tr>
<tr>
<td>3</td>
<td>up to 6</td>
<td>99.9%</td>
</tr>
<tr>
<td>4</td>
<td>up to 12</td>
<td>99.7%</td>
</tr>
<tr>
<td>5</td>
<td>up to 20</td>
<td>99.7%</td>
</tr>
</tbody>
</table>

Notes: Real firms: $f_j = 0$, $\alpha_j = 0.5$, $\sigma = \varepsilon_j = 6$, $\sigma_z = 0.25$. Link firms: $f_{\text{link}} \in \{0.0, 0.1/n\}$, $\sigma_{\text{link}} \sim \text{iid } U([0.5, 1.0]), \sigma_{\text{link}} = 0.25$ and $p \in \{0.7, 0.8, 0.9\}$, where $p$ is the probably that a link exists between any two real firms. For simplicity all non-zero $\Omega_{ij}$ are set to 1. For each combination of $p$ and $f_{\text{link}}$ I randomly generate 200 economies and use the algorithm of Section E.4 to solve the planner’s problem. An economy is kept in the sample only if the first-order conditions converge to a point in $\{0, 1\}^n$. More than 80% of the economies are kept in the sample.

Individual link formation in large networks

For economies with a large number of firms, the true solution to the planner’s problem is unknown but we can check whether there exist welfare-improving deviations from the allocation found using the reshaped problem. The procedure is the same as in the Section B.2 above. The parameters of the tests are the same as in Table 8 but the economies feature $m \in \{10, 25, 40\}$ real firms and, at most, $n \in \{100, 625, 1600\}$ total firms (real plus links). The results are presented in Table 9. We see that reshaping the planner’s problem yields solutions with few welfare-improving deviations so that the vast majority of links are assigned the correct status and the errors in aggregate output are negligible. In contrast, a large fraction of the links are assigned the wrong status when the problem of the planner is not reshaped and the errors in aggregate output are significant.

One potential concern of using the reshaping method in this context is that the first-order conditions often converge on a vector $\theta$ such that $\theta_j \notin \{0, 1\}$ for at least one firm. Since, for the economies considered here, the matrix $\Omega$ is extremely sparse, the forces pushing the first-order conditions to hit the bounds are weakened (see footnote 10). In practice, however, this issue has limited implications. In most cases, only a small fraction of the links end up away from the $\{0, 1\}$ bounds, and their impact on aggregate output is minimal. Table 10 shows the outcome of the same simulations but without

---

30 In the simulations of Table 9, the first-order conditions converge to a corner for all $j$ in 51% of the simulations for $m = 10$, 13% for $m = 25$ and 5% for $m = 40$. 

22
Table 9: Individual links formation with a large number of firms

<table>
<thead>
<tr>
<th>Number of firms</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real firms $m$</td>
<td>Link firms $n - m$</td>
<td>Correct $\theta$</td>
</tr>
<tr>
<td>10</td>
<td>up to 90</td>
<td>99.7%</td>
</tr>
<tr>
<td>25</td>
<td>up to 600</td>
<td>99.9%</td>
</tr>
<tr>
<td>40</td>
<td>up to 1560</td>
<td>$&lt;99.9%$</td>
</tr>
</tbody>
</table>

Notes: The parameters of these tests, except for $m$, are the same as in Table 8. An economy is kept in the sample only if the first-order conditions converge to a point in $\{0,1\}^n$. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0, and similarly for $x > 99.9\%$.

excluding any simulations. We see that the results are essentially unchanged and that the solution approach also performs well in these situations.

Table 10: Individual links formation with a large number of firms and without exclusions

<table>
<thead>
<tr>
<th>Number of firms</th>
<th>With reshaping</th>
<th>Without reshaping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real firms $m$</td>
<td>Link firms $n - m$</td>
<td>Correct $\theta$</td>
</tr>
<tr>
<td>10</td>
<td>up to 90</td>
<td>99.7%</td>
</tr>
<tr>
<td>25</td>
<td>up to 600</td>
<td>99.9%</td>
</tr>
<tr>
<td>40</td>
<td>up to 1560</td>
<td>$&lt;99.9%$</td>
</tr>
</tbody>
</table>

Notes: The parameters of these tests are the same as in Table 8. No economies are excluded from the sample. $x < 0.001\%$ indicates that $x > 0$ but that proper rounding would yield 0, and similarly for $x > 99.9\%$.

C  Quantitative exercises

This appendix provides additional information about the calibrated model of Section 5 as well as robustness exercises.

C.1 Data

The main firm-level input-output dataset is the Factset Revere Supply Chain Relationships Data, which covers the period from 2003 to 2016. 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. I restrict the sample to links in which at least one partner is located in the U.S. and I use the SDC Platinum dataset to exclude firms that are the target of a merger or an acquisition. In an average year, the sample includes almost 13,000 firms and more than 40,000 relationships.

We can use these data to evaluate the importance of the firms’ extensive operating margin for the shape of the production network. In an average year, more than 40% of all link destructions occur at the same time as either the supplier or the customer (or both) stops producing.\(^{31}\) This number suggests

\(^{31}\)To remove high-frequency gaps in the data, I assume that a link is created during the first year it appears in the data and is destroyed during the last year it appears in the data (except for the first and last years in the sample, 2003
that the firms’ decisions to operate play a first-order role in shaping input-output relationships, and it highlights the importance of modeling these decisions to properly understand the formation of the production network.

To verify the robustness of some empirical patterns, I also use Compustat data in Appendix C.5. These data cover fewer firms—about 1,300 firms and 1,500 relationships in an average year—but over a longer time period. They are available since 1976. The longer time series can be used to better estimate business cycles correlations. Compustat gathers data about firms’ major customers, defined as buyers of more than 10% of their overall sales, from annual financial statements. Because of this truncation, we rarely see a firm supplying to more than 10 clients in the data. As a result, the tail of the out-degree distribution in Compustat is likely to be artificially thinner. Factset also relies on these data but supplements them using a variety of other sources. This issue is therefore likely to be less important in the Factset sample. Another limitation of the Compustat data is that the names of the customers are self-reported, so 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 to construct the production networks. Their samples cover the periods 1980 to 2004 and 1976 to 2009, respectively.\footnote{I thank the authors for sharing their data.}

\section*{C.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 $\alpha = 0.5$. I assume that the log of the productivities $z_{it}$ are drawn from independent AR(1) processes with persistence $\rho_z$ and standard deviation of the ergodic distribution $\sigma_z$. 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. For the persistence, I follow Foster et al. (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, I set $f$ so that $f \times n = 5\%$. For the number of firms, I set $n = 1000$ as a good trade-off between realism and computation time.\footnote{See Appendix C.5 for simulations with $n = 20,000$ firms and aggregate shocks. The results are similar.} Broda and Weinstein (2006) use disaggregated trade data for the U.S. to estimate the elasticity of substitution between product varieties. I set $\sigma = 5$ as an average of their estimates. The empirical literature provides little guidance about the elasticity of substitution between intermediate inputs at the firm level. I therefore set $\sigma = \varepsilon$.

I construct the matrix $\Omega$ by assuming that the number of potential incoming and outgoing connections, for any given firm, is drawn from a bivariate power law of the first kind. This family of
distributions is entirely described by a unique shape parameter $\xi$. I set $\xi = 1.79$ so that the distribution of active incoming connections generated by the model is close to its empirical counterpart. These two distributions are well approximated by power laws, with the empirical distribution close to following Zipf’s law. I therefore target a power law exponent of 1 for the distribution generated by the model. This indirect inference approach ensures that the calibrated economy is consistent with a key feature of the empirical production network. So that the results do not hinge on one particular matrix $\Omega$, I draw 20 different $\Omega$’s with these properties and, for each of them, simulate the economy for 100 periods. The results are averages over these 20 matrices.

C.3 Calibrated economy

Table 11 shows how the calibrated production network compares to the U.S. data. The first two rows show the power law exponents for the in-degree and out-degree distributions, and the third row shows the global clustering coefficients. The model fits the Factset data relatively well but there are some discrepancies with the Compustat datasets. These discrepancies are particularly important when looking at the out-degree distribution which is not surprising given the 10% truncation described above.

Table 11: Production network in the calibrated economy and in the data

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factset</td>
<td>AHRS</td>
</tr>
<tr>
<td>Power law exponents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-degree distribution</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>Out-degree distribution</td>
<td>0.96</td>
<td>0.83</td>
</tr>
<tr>
<td>Global clustering coeff</td>
<td>0.70</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Notes: Power law exponents are estimated following Gabaix and Ibragimov (2011). Global clustering coefficients are multiplied by the square roots of the number of nodes. See footnote 20 for details.

Figure 5 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 uses a log-log scale and plots the complementary cumulative distributions (CCDF) on the vertical axis. The roughly linear shapes confirm that they are close to power laws. The model fits the in-degree distribution well but the fit of the out-degree distribution is less precise. Again, the data limitations discussed above are the likely culprit for the departure from the power law observed in the right tail

---

34Online Appendix G shows how the network reacts to changes in $\xi$.
35The out-degree distribution is not targeted since the 10% reporting threshold from accounting rules distorts this distribution the most, as explained earlier. All power law exponents are estimated following Gabaix and Ibragimov (2011). Appendix E.5 provides the algorithm used to construct $\Omega$.
36I discard and redraw simulations for which iterating on the first-order conditions does not converge to a point $\theta$ in $\{0, 1\}$$. This rarely happens and, overall, the rejected networks do not look different than the retained ones. Keeping all the simulations in the sample yields very similar results.
37See footnote 20 in the main text for a definition of this coefficient.
of the out-degree distribution.\textsuperscript{38}

![Graphs showing in-degree and out-degree distributions](image)

Figure 5: Distribution of the number of suppliers and customers

C.4 Comparison with a neutral network

To highlight which network features are desirable for efficiency, it is useful to compare the efficient network to a neutral benchmark built by operating each firm with some probability \( p > 0 \), where \( p \) is set so that both networks have the same number of active firms. All other quantities are chosen optimally. Since it is completely random, any discrepancies between this neutral benchmark and the efficient networks are design decisions taken by the planner to improve efficiency.

Table 12 shows how both networks differ. The power law exponents are smaller in the efficient network, indicating thicker tails than in the neutral network. The efficient network therefore features a larger fraction of highly connected suppliers and customers. The clustering coefficient is also larger in the efficient network.\textsuperscript{39} 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 gains from input variety present in the environment.

\textsuperscript{38}The model generates similar in- and out-degree distributions. In countries with better firm-level network data such as Japan, the in- and out-degree distributions also look similar (Bernard et al., 2015).

\textsuperscript{39}This exercise shows that a large amount of clustering among firms comes from the endogenous mechanisms of the model. To further disentangle how much clustering comes from the exogenous matrix \( \Omega \) versus the endogenous forces, I also compute the local clustering coefficient of each firm in the endogenous production network and compare it to its local clustering coefficient in the exogenous network \( \Omega \). That clustering coefficient is 40% larger in the endogenous production network, such that the forces at work in the economy generate substantial clustering on their own.
Table 12: The efficient network has more highly-connected firms and clustering

<table>
<thead>
<tr>
<th>Network</th>
<th>Power law exponents</th>
<th>Clustering coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-degree</td>
<td>Out-degree</td>
</tr>
<tr>
<td>Efficient</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Neutral</td>
<td>1.16</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Notes: Power law exponents are estimated following Gabaix and Ibragimov (2011). Global clustering coefficients are multiplied by the square roots of the number of nodes (see footnote 20).

C.5 Aggregate Fluctuations

This section contains additional results related to the aggregate fluctuations exercises of Section 5.2.

Invariance of the network to aggregate shocks. The following lemma establishes conditions under which changes in aggregate productivity have no effect on the network.

Proposition 9. If \( \alpha_j = \alpha \) for all \( j \in \mathcal{N} \), then the efficient network is invariant with respect to \( A \).

The proof of the proposition essentially shows that we can write the planner’s problem as

\[
\max_{\theta \in \{0,1\}^n} A^{1 \times 1 - \alpha} Q \left( 1 - \sum_{j \in \mathcal{N}} \theta_j f_j \right) L, \text{ where } q_j = z_j \theta_j \left( \sum_{i \in \mathcal{N}} \Omega_{ij} q_i^{\frac{\epsilon_j}{\epsilon - 1}} \right)^{\frac{\alpha}{\epsilon + 1}} , \forall j. \tag{11}
\]

We see that \( A \) only enters the problem by multiplying aggregate output. As a result, it has no impact on the optimal choice of \( \theta \) and, therefore, on the production network itself. We can also use (11) to compute the impact of aggregate shocks on output volatility. If the variance of log output was \( x \) without fluctuations in \( A \), then the overall variance of log output would simply be \( x + (1 - \alpha)^{-2} y \) if we added shocks (independent of \( z \)) with a variance of \( y \) to log \( (A) \).

Proposition 9 shows that we can first abstract from variations in aggregate productivity to focus on firm-level shocks when studying the joint movements in aggregate output and the shape of the network. Aggregate shocks can be easily added afterwards.

Aggregate fluctuations in the Compustat data. While the Factset data covers a large number of input-output relationships in a typical year, its timespan is relatively short. To make sure that the business cycle moments documented in the main text also hold over longer periods, Table 13 extends Table 3 with the network data from the Compustat datasets put together by Cohen and Frazzini (2008) (CF) and Atalay et al. (2011) (AHRS), as described in Section C.1. We see that the movements between the production network and aggregate output go in the same direction as in the Factset data although the magnitudes differ.
Table 13: Correlation between the network and the business cycle

<table>
<thead>
<tr>
<th>Power law exponents</th>
<th>Model</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factset</td>
<td>Compustat CF</td>
</tr>
<tr>
<td>In-degree distribution</td>
<td>-0.59</td>
<td>-0.87</td>
</tr>
<tr>
<td>Out-degree distribution</td>
<td>-0.70</td>
<td>-0.97</td>
</tr>
<tr>
<td>Global clustering coefficient</td>
<td>0.53</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Notes: All time series are in logs. In the data, output is annual real GDP. Output is detrended linearly in sample. Since there are only 13 years in the Factset data 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.

D Proofs

This section contains the proofs.

D.1 Equilibrium

The following lemma is useful to show that every stable equilibrium is efficient.

Lemma 2. Let $X, Y \subset \mathbb{R}^n$. Define Problem $A$ as

$$\max_{x \in X, y \in Y} f(x, y) \text{ subject to } \sum_j y_j \leq 0$$

and Problem $B$ as

$$\max_{x \in X, y \in Y} g(f(x, y)) - \lambda \left(\sum_j y_j\right)$$

where $g$ is a strictly increasing function and where $\lambda$ is such that $\sum_j y_j = 0$ at any solution. Suppose that for any solution to Problem $A$ the constraint binds, then Problems $A$ and $B$ have the same solutions.

Proof. Take a point $(x^A, y^A)$ that solves Problem $A$. Then, since the constraint binds $\sum_j y^A_j = 0$. Suppose $(x^A, y^A)$ does not solve Problem $B$. Then there is another point $(\tilde{x}, \tilde{y})$ such that $\sum_j \tilde{y}_j = 0$ (by the definition of $\lambda$) and such that $g\left(f(\tilde{x}, \tilde{y})\right) - \lambda \left(\sum_j \tilde{y}_j\right) > g\left(f(x^A, y^A)\right) - \lambda \left(\sum_j y^A_j\right)$. Since $g$ is strictly increasing this implies that $f(\tilde{x}, \tilde{y}) > f(x^A, y^A)$ but, since $(\tilde{x}, \tilde{y})$ is in the feasible set of Problem $A$, this implies that $(x^A, y^A)$ was not a solution to Problem $A$, which is a contradiction.

Conversely, take a point $(x^B, y^B)$ that solves Problem $B$. Then by the definition of $\lambda$ it must be that $\sum_j y^B_j = 0$. Suppose $(x^B, y^B)$ does not solve Problem $A$. Then there is another point $(\tilde{x}, \tilde{y})$ such that $\sum_j \tilde{y}_j = 0$ (since the constraint in Problem $A$ binds at the optimum) and such that $f(\tilde{x}, \tilde{y}) > f(x^B, y^B)$. Since $g$ is strictly increasing this implies that $g\left(f(\tilde{x}, \tilde{y})\right) - \lambda \left(\sum_j \tilde{y}_j\right) > g\left(f(x^B, y^B)\right) - \lambda \left(\sum_j y^B_j\right)$ so that $(x^B, y^B)$ is not a solution to Problem $B$, which is a contradiction. □
Proposition 8. Every stable equilibrium is efficient.

Proof. The proof proceeds by establishing restrictions that any stable equilibrium must satisfy. It then shows that any allocation that satisfies these restrictions must be efficient.

Consider a coalition made of all the firms in the economy. For the equilibrium to be stable there cannot be an alternative arrangement that would yield larger aggregate profits. Otherwise, transfers could be designed to make one firm better off while keeping the other firms at the same profit level. The arrangement \( \{x_{ij}, T_{ij}\}_{i,j} \) must therefore maximize \( \sum_{j \in N} \pi_j \). Now, since a change in the arrangement leads to reoptimization by the firms, an equilibrium arrangement needs to maximize the coalition’s profits subject to \( \{p_j, c_j, l_j, \theta_j\}_j \) being determined by the individual firms. Any equilibrium allocation therefore solves

\[
\max_{\{x_{ij}, T_{ij}\}_{i,j}} \sum_{j \in N} \max_{\{p_j, c_j, l_j, \theta_j\}} \pi_j \left( p_j, c_j, l_j, \theta_j, \{x_{ij}\}_{ij} \right) \quad \text{s.t. (9) and (10)}.
\] (12)

It is, however, equivalent to let the coalition itself directly optimize over \( \{p_j, c_j, l_j, \theta_j\}_j \). To see this, notice that, conditional on the arrangement, the inner maximization problems in (12) are all independent from each other. In other words, the decisions of a firm \( i \) have no effect on the profit of a firm \( j \) as long as the contracts specified by the arrangement are fulfilled. As a result, we can write (12) as

\[
\max_{\{x_{ij}\}_{i,j},\{c_j,l_j,\theta_j\}_j} \sum_{j \in N} \pi_j \text{ subject to the constraints (9) and (10) for all firms.}
\]

By including the household’s demand curves directly in the objective function, and by using the definition of \( \pi_j \), the absence of dominating deviations therefore implies that the allocation must solve

\[
\max_{\{x_{ij}\}_{i,j},\{c_j,l_j,\theta_j\}_j} C^\frac{1}{2} P \sum_{j \in N} \beta_j^\frac{1}{\sigma} c_j^{\frac{\sigma-1}{\sigma}} - w \sum_{j \in N} (l_j + \theta_j f_j L)
\]
subject to (9) for all \( j \in N \) and where \( C \) and \( P \) are taken as given. Now, by Lemma 2 this problem is equivalent to an alternative problem in which the coalition maximizes \( \left( \sum_{j \in N} \beta_j^\frac{1}{\sigma} c_j^{\frac{\sigma-1}{\sigma}} \right)^{\sigma-1} \) subject to \( \sum j l_j + \theta_j f_j L \leq L \) in addition to the other constraints.\(^{40}\) This reformulated problem is identical to the problem \( \mathcal{P} \) of the social planner such that any stable equilibrium must be efficient. \( \square \)

D.2 Results about \( q \)

The proof of the uniqueness of \( q \) relies on the following definitions from Kennan (2001).

**Definition.** A function \( g : \mathbb{R}^n \to \mathbb{R}^n \) is radially quasiconcave (“R-concave”) if \( g(x) = 0 \) and \( x > 0 \) and \( 0 \leq \lambda \leq 1 \) implies \( g(\lambda x) \geq 0 \). If (in addition) \( 0 < \lambda < 1 \) implies \( g(\lambda x) > 0 \), then \( g \) is strictly R-concave.

\(^{40}\) The constraints (9) can be included directly in the function \( f \) in Lemma 2 by setting \( f = -\infty \) for points outside the constraint set.
**Definition.** A function \( g = (g_1, g_2, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n \) is quasi-increasing if \( y_i = x_i \) and \( y_j \geq x_j \) for all \( j \) implies \( g_j(y) \geq g_i(x) \).

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

**Lemma 3.** Denote by \( \tilde{N} \) any subset of \( N \) with \( \tilde{n} \) firms and such that \( \sum_{i \in \tilde{N}} \Omega_{ij} > 0 \) for all \( j \in \tilde{N} \). The function \( g : \mathbb{R}^\tilde{n} \to \mathbb{R}^\tilde{n} \) defined, for all \( j \in \tilde{N} \), as

\[
g_j(p) = (z_j A)^{\varepsilon_j} \left( \sum_{i \in \tilde{N}} \Omega_{ij} p_i^{\varepsilon_j} - p_j \right)
\]

is strictly \( R \)-concave.

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

\[
g_j(\lambda p^*) = \lambda^{\alpha_j} (z_j A)^{\varepsilon_j} \left( \sum_{i \in \tilde{N}} \Omega_{ij} (p_i^*)^{\varepsilon_j} \right)^{\frac{\alpha_j \varepsilon_j}{\varepsilon_j - 1}} - \lambda p_j^*
\]

\[
\geq \lambda (z_j A)^{\varepsilon_j} \left( \sum_{i \in \tilde{N}} \Omega_{ij} (p_i^*)^{\varepsilon_j} \right)^{\frac{\alpha_j \varepsilon_j}{\varepsilon_j - 1}} - \lambda p_j^*
\]

\[
\geq \lambda g_j(p^*)
\]

\[
\geq 0
\]

where the first inequality is strict for \( 0 < \lambda < 1 \) since \( 0 < \alpha_j < 1 \) and \( \sum_{i \in \tilde{N}} \Omega_{ij} > 0 \) by assumption.

**Proposition 1.** In the efficient allocation, the productivity vector \( q \) satisfies, for all \( j \in \tilde{N} \),

\[
q_j = z_j \theta_j A \left( \sum_{i \in \tilde{N}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\frac{\alpha_j}{\varepsilon_j - 1}}. \tag{5}
\]

There is a unique \( q \) that solves (5) such that \( q_j > 0 \) if \( j \) operates and has access to at least one input.

**Proof.** The first-order conditions of \( P \) with respect to \( l_j \) and \( x_{ij} \) are

\[
w l_j = \lambda_j (1 - \alpha_j) y_j \tag{13}
\]

\[
\lambda_i x_{ij} = \lambda_j \frac{A}{\alpha_j} \left( 1 - \alpha_j \right)^{1 - \alpha_j} \alpha_j z_j \theta_j \left( \sum_{k \in \tilde{N}} \Omega_{kj}^{\frac{1}{\varepsilon_j - 1}} x_k^{\varepsilon_j - 1} \right)^{\frac{\varepsilon_j}{\varepsilon_j - 1}} l_j^{1 - \alpha_j} \Omega_{ij}^{\frac{1}{\varepsilon_j - 1}} x_{ij}^{\varepsilon_j - 1}. \tag{14}
\]

Combining these conditions with the production function yields

\[
x_{ij} \lambda_i^{\varepsilon_j} = \lambda_j^{\varepsilon_j} \alpha_j \left( A z_j \theta_j \left( \frac{\lambda_j}{w} \right)^{1 - \alpha_j} \right)^{\frac{\varepsilon_j}{\alpha_j}} \Omega_{ij} y_j. \tag{15}
\]
Plugging (13) and (15) back in the production function yields (5).

I follow Kennan (2001) to prove the uniqueness of \( q \). Consider the change of variable \( p_j = q_j^{e_j} \), and let \( \tilde{N} \) be the set of firms that operate and that have access to at least one input (see footnote 4). Denote the number of such firms by \( \tilde{n} \). Clearly, \( p_j = 0 \) for \( j \notin \tilde{N} \). We can rewrite equation (5) as the following mapping from \( \mathbb{R}^{\tilde{n}} \) to \( \mathbb{R}^{\tilde{n}} \):

\[
p_j = (z_j A)^{\epsilon_j} \left( \sum_{i \in \tilde{N}} \Omega_{ij} p_i \right)^{\epsilon_j - 1}.
\]

(16)

for all \( j \in \tilde{N} \). Let \( f_j(p) \) be the right-hand side of (16) and define \( g : \mathbb{R}^{\tilde{n}} \to \mathbb{R}^{\tilde{n}} \) as \( g(p) = f(p) - p \). Then, by Lemma 3, \( g \) is strictly \( R \)-concave. Note also that \( g \) is quasi-increasing.

Consider the mapping \( h : \mathbb{R} \to \mathbb{R}^{\tilde{n}} \) defined as \( h(s) = f(1s) \) where \( 1 \) 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 constants \( \bar{p} > p > 0 \) such that \( h(p) > p \) and \( h(\bar{p}) < \bar{p} \). Then, Theorems 3.1 and 3.2 in Kennan (2001) apply and (16) has a unique positive fixed point \( p^* \).

There is therefore a unique positive \( q^* \) that satisfies (5). It is such that \( q^*_j = \left( p_j^* \right)^{\epsilon_j} \) if \( j \) operates and has access to at least one input, and \( q^*_j = 0 \) otherwise. Note that the proof is essentially unchanged if we use the reshaped equation (7) instead of (5).

**D.3 Taking \( \theta \) as fixed**

**Lemma 1.** In the efficient allocation, aggregate output is

\[
C = Q \left( 1 - \sum_{j \in \mathcal{N}} \theta_j f_j \right) L
\]

(6)

where \( Q = \left( \sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma - 1} \right)^{1/(\sigma - 1)} \).

*Proof.* The first-order condition of \( P \) with respect to \( c_j \) is

\[
c_j = \beta_j \left( \frac{q_j}{w} \right)^\sigma C.
\]

(17)

Raising both sides to the power \( \frac{\sigma - 1}{\sigma} \) and summing across \( j \)'s yields \( w = Q \). Using the first-order conditions (13), (15) and (17) into the resource constraints (3), we find

\[
0 \geq \beta_j \left( \frac{q_j}{Q} \right)^{\sigma - 1} C + \sum_{k \in \mathcal{N}} \frac{\Omega_{jk} q_j^{\epsilon_k - 1}}{\sum_{i \in \mathcal{N}} \Omega_{ik} q_i^{\epsilon_k - 1}} \frac{l_k}{1 - \alpha_k} - \frac{l_j}{1 - \alpha_j}
\]

(18)

for all \( j \in \mathcal{N} \). Summing across \( j \)'s and simplifying yields (6). Note that once \( q \) is known we can find \( l \) by inverting (18). We can then find \( y \) and \( x \) using the first-order conditions (13) and (15).
D.4 Reshaping the planner’s problem

Let \( V_R : [0,1]^n \rightarrow \mathbb{R} \) be the objective function of \( R \) defined as

\[
V_R(\theta) = \left( \sum_{j \in \mathcal{N}} (q_j(\theta))^{\sigma-1} \right)^{\sigma-1} \left( 1 - \sum_{j \in \mathcal{N}} f_j \theta_j \right)^L
\]  
(19)

where \( q_j \) is implicitly defined by (7). Similarly, let \( V_P : \{0,1\}^n \rightarrow \mathbb{R} \) be the objective function of \( P \).

Preliminary result

The proof of Proposition 3 relies on the following lemma.

Lemma 4. Let \( F = A - fB \) where \( f > 0 \), \( A \) is the all-one \( n \times n \) matrix and \( B \) is an \( n \times n \) matrix.

If \( B \) is negative definite on the subspace \( S : \sum_{i=1}^n x_i = 0 \) then \( F \) is positive definite for \( f > 0 \) small enough.

Proof. The negative definiteness of \( B \) on \( S \) implies that \( x'Bx \leq -d \|x\|^2 \) for \( x \in S \) and some \( d > 0 \). We can write any vector \( z \) as \( z = x + y \) where \( x \in S \) and \( y \perp S \). Then,

\[
z'(A - fB)z = n\|y\|^2 - fx'Bx - fy'By - 2fy'Bx \\
geq (n - 1/2)\|y\|^2 + df\|x\|^2 - 2f\|B\||x||y|
\]

for \( f \) small enough. For \( f \) small enough, this last expression is strictly convex in \((\|x\|,\|y\|)\) with a minimum of 0 at \((0,0)\) or, equivalently, at \( z = 0 \). Since \( z'(A - fB)z > 0 \) for any \( z \neq 0 \), it follows that \( F \) is positive definite. \( \square \)

Proofs of concavity

Proposition 2. Let \( \varepsilon_j = \varepsilon \) and \( \alpha_j = \alpha \). If \( \Omega_{ij} = d_ie_j \) for some vectors \( d \) and \( e \) then the Karush-Kuhn-Tucker conditions are necessary and sufficient to characterize a solution to \( R \).

Proof. Raise both sides of (7) to the power \( \varepsilon - 1 \), multiply by \( d_j \theta_j^b \) and sum across \( j \)'s to find

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

so that, once combined with (7), we find

\[
q_j(\theta) = Az_j \theta_j^a d_j^{\frac{\alpha}{1-a}} \left( \sum_{i \in \mathcal{N}} d_i e_i^a (Az_i)^{\varepsilon-1} \theta_i^{a(\varepsilon-1)+b} \right)^{\frac{\alpha}{1-a}}.
\]

32
Computing the log of $Q$, we get
\[
\log (Q) = \frac{1}{\sigma - 1} \log \left( \left( \sum_{j \in N} \beta_j \left( z_j \theta_j^a A_d_j^{\sigma-1} \right) \right) \left( \sum_{i \in N} d_i e_i^a (Az_i)^{\sigma-1} \theta_i^{a(\varepsilon-1)+b} \right)^{\frac{\alpha-1}{\alpha}} \right)
= \frac{1}{\sigma - 1} \log \left( \sum_{j \in N} \beta_j \left( z_j \theta_j^a A_d_j^{\sigma-1} \right) \right) + \frac{1}{\varepsilon - 1} \log \left( \sum_{i \in N} d_i e_i^a (Az_i)^{\sigma-1} \theta_i^{a(\varepsilon-1)+b} \right)
\]

If $0 < a \leq (\sigma - 1)^{-1}$ and $-a (\varepsilon - 1) \leq b \leq 1 - a (\varepsilon - 1)$ (and in particular if ($\star$) holds) the exponents on $\theta$ are all between 0 and 1 so that the summations in $\log (Q)$ are concave functions of $\theta$. The log of a concave function is concave so $\log (Q)$ is also concave. Moving towards the full objective function, the term $1 - \sum_{j \in N} \theta_j f_j$ is concave and so is $\log V_R$. 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 $\log (V_R (\theta))$ on the set $\theta \in [0,1]^n$. Since log is an increasing transformation, a solution to this problem also solves $\mathcal{R}$.

Proof. To simplify the notation, define $p_j = q_j^{\sigma-1}$ and let
\[ g^j = \frac{p_j}{z_j^{\sigma-1} (\sum_i \Omega_{ij} p_i)^{\alpha_j}}. \]
$\mathcal{R}$ can then be written as
\[
\min_{p \in P} -\frac{1}{\sigma - 1} \log \left( \sum_{j \in N} \beta_j p_j \right) - \log \left( 1 - \sum_{j \in N} f_j g^j (p) \right)
\]
where $P = \left\{ p \in \mathbb{R}_{\geq 0}^n : p_j \leq z_j^{\sigma-1} (\sum_{i \in N} \Omega_{ij} p_i)^{\alpha_j}, \forall j \right\}$.

Denote the objective function by $\Lambda$. Its Hessian matrix has typical element
\[
\frac{\partial^2 \Lambda}{\partial p_k \partial p_l} = \frac{1}{\sigma - 1} \beta_k \beta_l \left( \sum_{j \in N} \beta_j p_j \right)^{-2} \sum_{j \in N} f_j g_k^j (p) \sum_{j \in N} f_j g_l^j (p) + \frac{\left( \sum_{j \in N} f_j g_k^j (p) \right) \left( \sum_{j \in N} f_j g_l^j (p) \right)}{(1 - \sum_{j \in N} f_j g^j (p))^2},
\]
and define $A$, $B$ and $C$ as the matrices with typical elements $A_{kl}$, $B_{kl}$ and $C_{kl}$.

We will show that in the limit as $\Omega \to \bar{\Omega}$ and $\beta_j \to \bar{\beta}$ for all $j$ the Hessian is positive definite when the largest fixed cost $\max_j f_j$ is small enough. To do so, we will rely on Lemma 4 above. For that
purpose, notice that, in the limit, $A$ is a positive multiple of the all-one matrix.

Pick $\bar{f} > 0$ and $\bar{f}_j \in [0, 1]^n$ so that $f_j = \bar{f} \bar{f}_j$ for all $j$. Taking the derivatives of $g$, we find

$$B_{kl} = \frac{1}{L - \sum_j f_j g^2(p)} \left( \sum_{j \in \mathcal{N}} \bar{f}_j \frac{\alpha_j (\alpha_j + 1) p_j \Omega_{kj} \Omega_{ij}}{z_j^{\sigma+1} (\sum_i \Omega_{ij} p_i)^{\alpha_j+2}} B^1_{kl} - \bar{f}_k \frac{\alpha_k \bar{f} \Omega_{lk}}{z_k^{\sigma-1} (\sum_i \Omega_{ik} p_i)^{\alpha_k+1}} + \bar{f}_l \frac{\alpha_l \bar{f} \Omega_{lk}}{z_l^{\sigma-1} (\sum_i \Omega_{il} p_i)^{\alpha_l+1}} \right).$$

(21)

$B^1$ is a Gramian matrix where $B^1_{kl}$ is the scalar product of a pair of vectors $v_k$ and $v_l$ defined as

$$v_m = \left[ \frac{\bar{f}_m \alpha_m (\alpha_m+1) p_m}{z_m^{\sigma-1} (\sum_i \Omega_{im} p_i)^{\alpha_m+2}} \Omega_{m1} \cdots \sqrt{\frac{\bar{f}_l \alpha_l (\alpha_l+1) p_l}{z_l^{\sigma-1} (\sum_i \Omega_{il} p_i)^{\alpha_l+2}}} \Omega_{mj} \cdots \sqrt{\frac{\bar{f}_n \alpha_n (\alpha_n+1) p_n}{z_n^{\sigma-1} (\sum_i \Omega_{in} p_i)^{\alpha_n+2}}} \Omega_{mn} \right].$$

Since Gramian matrices are positive semi-definite, so is $B^1$. For $B^2$, we will show that, in the limit, it is negative definite on the subspace $S : \sum_{i=1}^n x_i = 0$. Define $b$ as a vector with typical element $b_j = \bar{f}_j \alpha_j \left[ z_j^{\sigma-1} (\sum_i \Omega_{ij} p_i)^{\alpha_j+1} \right]^{-1}$. We can write $B^2 = \Omega \text{diag}(b) + (\Omega \text{diag}(b))'$. Take any vector $x \in S$, then in the limit,

$$x' B^2 x = x' \left[ \omega (\mathbb{1}_n - I_n) \text{diag}(b) + (\omega (\mathbb{1}_n - I_n) \text{diag}(b))' \right] x$$

$$= x' [-2 \omega I_n \text{diag}(b)] x < 0$$

for any $x \neq 0$. The matrix $B^2$ is therefore negative definite on $S$. Using Lemma 4, $A - \bar{f} B^2$ is therefore positive definite for $\bar{f} > 0$ small enough. Finally, the matrix $C$ in (20), is also a Gramian matrix and its contribution to the Hessian is thus positive semi-definite.

Putting the pieces together, we have shown that the Hessian of the objective function $\Lambda$ is positive definite for $\max_j f_j$ small enough when $\Omega = \bar{\Omega}$ and $\beta = \bar{\beta}$. Now, each element of the Hessian is also a continuous function of $(\Omega, \beta)$ in a neighborhood of $(\bar{\Omega}, \bar{\beta})$. Since the eigenvalues are continuous functions of the elements of a matrix, they are also continuous functions of $\Omega$ and $\beta$. There is therefore a ball $B = \{ (\Omega, \beta) : \|(\Omega, \beta) - (\bar{\Omega}, \bar{\beta})\| < \delta \}$ for $\delta > 0$ such that the Hessian is also positive definite for $(\Omega, \beta) \in B$.

---

41 Elements of the Hessian become infinite at the boundary of $P$ where $\sum_i \Omega_{ij} p_i = 0$ for some $j$. However, these points are not relevant to the planner under our assumptions. When the fixed costs are small enough and when $\Omega$ is close enough to $\bar{\Omega}$, each firm is connected to at least one producing firm at the optimum. Therefore, we can exclude these points easily adding the constraints $\sum_i \Omega_{ij} p_i \geq D$ for some $D > 0$ small. These constraints will never bind and the solution to the planner’s problem is therefore unchanged.

42 Since all norms are equivalent in a finite dimensional space, there is no need to specify one here.
Proof of the equivalence of the solutions

**Proposition 4.** If a solution $\theta^*$ to $\mathcal{R}$ is such that $\theta^* \in \{0, 1\}^n$ then $\theta^*$ also solves $\mathcal{P}$.

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

**Proposition 5.** The (net) marginal benefit of increasing $\theta_j$ only depends on $\theta_j$ through aggregates.

*Proof.* Rewrite $\mathcal{R}$ as

$$\max_{\theta \in [0, 1]^n} \left( \sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma-1} \right)^{\frac{1}{\sigma-1}} \left( 1 - \sum_{j \in \mathcal{N}} f_j \theta_j \right) L$$

subject to $q_j \leq A_z \theta_j^\alpha_j B_j^\beta_j$ for all $j \in \mathcal{N}$ and where $B_j = \left( \sum_{i \in \mathcal{N}} \theta_i^\beta_i \Omega_{ij} \xi_i^{\sigma_j-1} \right)^{\frac{1}{\sigma_j-1}}$. This problem is equivalent to $\mathcal{R}$ since the inequality constraints always bind at the optimum. The first-order conditions with respect to $\theta_k$ and $q_k$ are

$$f_k Q + \mu_k - \bar{\mu}_k - \zeta_k \frac{\partial q_k}{\partial \theta_k} + \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \frac{\partial B_j}{\partial \theta_k}$$

$$\frac{\partial Q}{\partial q_k} \left( 1 - \sum_{j \in \mathcal{N}} f_j \theta_j \right) L = \zeta_k - \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \frac{\partial B_j}{\partial q_k}$$

where $\zeta_j$ is the Lagrange multiplier on the $j$-th inequality constraints, and $\bar{\mu}_j$ and $\mu_j$ are the Lagrange multipliers on the constraint $\theta_j \leq 1$ and $\theta_j \geq 0$. Combining these first-order condition yields

$$f_k Q + \mu_k - \bar{\mu}_k = \left( 1 - \sum_{j \in \mathcal{N}} f_j \theta_j \right) L \frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} + \sum_{j \in \mathcal{N}} \zeta_j \frac{\partial q_j}{\partial B_j} \left( \frac{\partial B_j}{\partial q_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial B_j}{\partial \theta_k} \right).$$

For the result to hold, we need $\frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k}$ and $\frac{\partial B_j}{\partial q_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial B_j}{\partial \theta_k}$ to depend on $\theta_k$ only through aggregates. We can write

$$\frac{\partial Q}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} = \left( \sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma_j-1} \right)^{\frac{1}{\sigma_j-1}} \beta_k \left( A_{zk} \theta_k^\alpha_k B_k^\beta_k \right)^{\sigma_j-2} \times a_k A_{zk} \theta_k^\alpha_k B_k^\beta_k.$$

---

43 The partial derivatives of $q_j$ are to be understood for the binding inequality constraint, i.e. $\frac{\partial q_j}{\partial \theta_j} = A_z \theta_j^{\sigma_j-1} AB_j^{\alpha_j}$ and $\frac{\partial q_j}{\partial B_j} = A_z \theta_j^{\alpha_j} A_{\alpha_j} B_j^{\beta_j-1}$. 

35
where we see that $\theta_k$ drops out of the equation if $a_k = \frac{1}{\sigma - 1}$. Similarly, we can write
\[
\frac{\partial B_j}{\partial \theta_k} + \frac{\partial B_j}{\partial q_k} \frac{\partial q_k}{\partial \theta_k} = \frac{1}{\varepsilon_j - 1} \left( \sum_{i \in \mathcal{N}} \theta_i^{b_{ij}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{-1} \left( b_{kj} \theta_k^{b_{kj} - 1} \Omega_{kj} q_k^{\varepsilon_j - 1} + \theta_k^{b_{kj}} \Omega_{kj} (\varepsilon_j - 1) q_k^{\varepsilon_j - 2} a_k A z_k \theta_k^{a_k - 1} B^\alpha_k \right)
\]
and, by taking into account that $q_k$ depends directly on $\theta_k$, we see that $\theta_k$ drops out of the equation if we also impose $b_{kj} = 1 - \frac{\varepsilon_j - 1}{\sigma - 1}$. The first-order condition therefore only depends on $\theta_k$ through the aggregates $B_k$ and $Q$.

**D.5 Forces shaping the network**

**Proposition 6.** In a large economy, operating a firm (weakly) increases the incentives to operate its (perhaps indirect) suppliers and customers in $\Omega$.

**Proof.** Let $j$ be a newly operating firm. In a large economy, operating $j$ does not affect the Lagrange multiplier on the labor resource constraint $w$. Consider a firm $k$ downstream from $j$ in $\Omega$. From the recursivity of (5), $q_k$ weakly increases as a result of operating $j$ (there is a strict increase if $k$ is a direct neighbor). As a result, the benefit to the planner of operating $k$ also weakly increases. On the other hand, the cost of operating $k$ is still $w f_k L$ so that the net benefit of operating $k$ weakly increases.

Now consider a firm $i$ that is upstream from $j$ in $\Omega$. From the recursivity of (5) and since $j$ now operates, operating $i$ would weakly increase $q_j$ (with a strict increase if $i$ and $j$ are direct neighbor) which the planner values. The cost of operating $i$ is still $w f_i L$ so the net benefit of operating $i$ weakly increases.

**Proposition 7.** The incentives of the planner to operate a group of firms (weakly) increase with additional potential connections between them.

**Proof.** Take a set of firms $\mathcal{J}$ and suppose that there exists $i, j \in \mathcal{J}$ such that $\Omega_{ij} = 0$. Operating these firms provides a certain benefit to the planner, and this benefit is increasing in the firms’ productivities $\{q_k\}_{k \in \mathcal{J}}$. Suppose instead that $\Omega_{ij} > 0$. By equation (7), this additional potential connection weakly increases each productivity $q_j$ for $j \in \mathcal{J}$, with a strict increase for $q_j$ if $i$ and $j$ produce a positive amount, i.e. $q_i > 0$ and $q_j > 0$. The cost to the planner of operating the firms in $\mathcal{J}$ is unchanged by the inclusion of the additional potential connection $\Omega_{ij}$. As a result, the incentives of the planner to operate the firms in $\mathcal{J}$ are larger with $\Omega_{ij} > 0$.

**D.6 Aggregate fluctuations**

**Proposition 9.** If $\alpha_j = \alpha$ for all $j \in \mathcal{N}$, then the efficient network is invariant with respect to $A$.

**Proof.** Using Proposition 1 and Lemma 1, we can rewrite $P$ as maximizing (6) over the set $\theta \in \{0, 1\}^n$ where the vector $q$ solves, for each $j \in \mathcal{N}$, (5). Let’s denote this problem as $P_A$, where $A$ refers to the
aggregate productivity level. We will show that the optimal vector $\theta_A$ that solves $P_A$ also solves an alternative problem $P_{\tilde{A}}$ in which aggregate productivity is $\tilde{A}$ instead. Define $p_j = \left(\frac{\tilde{A}}{A}\right)^{1-\alpha} q_j$, then the objective function of $P_{\tilde{A}}$ can be written as $\left(\frac{\tilde{A}}{A}\right)^{1-\alpha} \left(\sum_{j \in N} \beta_j p_j^{a-1}\right)^{\frac{1}{1-\alpha}} \left(1 - f_j \sum_{j \in N} \theta_j\right) L$ and its recursive equation (5) can be written as $p_j = z_j \theta_j A \left(\sum_{i \in N} \Omega_{ij} (p_i)^{\varepsilon_j-1}\right)^{\varepsilon_j}$. Since the constant $\left(\frac{\tilde{A}}{A}\right)^{1-\alpha}$ does not affect the maximization, $P_{\tilde{A}}$ is the same problem as $P_A$, and, at the optimum, $p_{\tilde{A}} = q_A$ and $\theta_{\tilde{A}} = \theta_A$. The production network is therefore invariant to changes in aggregate productivity $A$.

Online Appendix (not for publication)

This online appendix contains the algorithms referenced in the paper as well as additional quantitative exercises.

E Algorithms

This appendix describes the various algorithms used in the paper.

E.1 Construction of the matrix $\Omega$ for the numerical tests.

This algorithm constructs the matrices $\Omega$ used in the numerical tests of Section 3.1 in the main text and of Sections B.1 and B.3 in the Appendix. Consider an economy with $n$ firms, each with $m$ incoming potential connections on average. Set $p = m/n$ and $\Omega_{ij} = 0$ for all $i, j \in N^2$.

1. For each $i, j \in N^2$ draw $\Omega_{ij} \sim$ iid Bernoulli($p$).
2. For each $i, j \in N^2$ such that $\Omega_{ij} = 1$, draw $\Omega_{ij} \sim$ iid $U[0, 1]$.

E.2 Exhaustive search

This algorithm performs an exhaustive search of the $2^n$ vectors $\theta \in \{0,1\}^n$. It is used in Section 3.1 in the main text as well as in Sections B.1, B.3 and B.4 in the Appendix.

1. Order in an arbitrary way all the possible $\theta \in \{0,1\}^n$, from $\theta^1$ to $\theta^{2^n}$.
2. For each $p \in \{1, \ldots, 2^n\}$, use equations (5) and (6) to compute the aggregate consumption associated with $\theta^p$.
3. The vector $\theta$ that provides the highest aggregate consumption corresponds to the efficient allocation.
This algorithm is guaranteed to find the global maximum of $P$ but it is infeasible for large $n$ given the speed at which the number of vectors in $\{0, 1\}^n$ grows with $n$.

**E.3 Deviation-free allocation**

This algorithm starts from an allocation $\theta^0 \in \{0, 1\}^n$ and looks for welfare-improving deviations. It is used in Sections B.2, B.3 and B.4.

1. Initialize the 0-th iteration with $\theta^0$.

2. For the $p$-th iteration, define $\tilde{\theta} = \theta^p$ and set $j = 1$.
   
   (a) If $\theta^p_j = 0$, set $\tilde{\theta}_j = 1$. If, instead, $\theta^p_j = 1$, set $\tilde{\theta}_j = 0$.
   
   (b) Using equations (5) and (6) compute the welfare associated with $\tilde{\theta}$.
   
   (c) If the welfare under $\tilde{\theta}$ is larger than the welfare under $\theta^p$ set $\theta^p = \tilde{\theta}$.
   
   (d) Set $j = j + 1$, set $\tilde{\theta} = \theta^p$ and repeat steps (a) through (d) until $j = n$.

3. Repeat step 2 above until no welfare-improving deviations are found for some $\theta^p$.

**E.4 Iterating on the first-order conditions**

A convenient way to solve the reshaped planner’s problem is to iterate on the first-order conditions of the log of the objective function of $R$ while treating (7) as an inequality constraint. In what follows $\zeta_k$ is the Lagrange multiplier on the $k$-th inequality constraint (7), and $\mu_j$ and $\bar{\mu}_j$ are the Lagrange multipliers on the constraint $\theta_j \geq 0$ and $\theta_j \leq 1$. The algorithm is as follows:

1. Initialize the 0-th iteration with $\Delta\mu_k^0 = \mu_j^0 - \bar{\mu}_j^0 = -1$ for all $k \in \mathcal{N}$.

2. For the $p$-th iteration:
   
   (a) Using the complementary slackness condition set $\theta^p_k = 1$ if $\Delta\mu_k^p \leq 0$ and $\theta^p_k = 0$ if $\Delta\mu_k^p > 0$.
   
   (b) With $\theta^p$, iterate on (7) until convergence to find the vector $q^p$.
   
   (c) Find $\frac{\zeta_k q_k}{\theta_k}$ by solving the following system of linear equations derived from the first-order conditions:

   $$\frac{\beta_k (Az_k B_k^{\alpha_k})^{\sigma - 1}}{\sum_{j \in \mathcal{N}} \beta_j q_j^{\sigma - 1}} + \sum_{j \in \mathcal{N}} (Az_k B_k^{\alpha_k})^{\varepsilon_j - 1} \Omega_{kj} \theta_j \frac{\zeta_j q_j}{\sum_{j \in \mathcal{N}} (Az_k B_k^{\alpha_k})^{\varepsilon_j - 1} \theta_j} = \frac{\zeta_k q_k}{\theta_k}$$

   for each $k$ and where $B_j = \left( \sum_{i=1}^n \theta_i^{b_{ij}} \Omega_{ij} q_i^{\varepsilon_j - 1} \right)^{\varepsilon_j - 1}$.

   (d) Compute $\Delta\mu_k$ using the following equation derived from the first-order conditions

   $$\frac{f_k}{L - \sum_{j \in \mathcal{N}} f_j \theta_j} = \frac{1}{\sigma - 1} \frac{\zeta_k q_k}{\theta_k} + \sum_{j \in \mathcal{N}} \left( 1 - \frac{\varepsilon_j - 1}{\sigma - 1} \right) \frac{\alpha_j \Omega_{kj} \theta_j}{B_j^{\varepsilon_j - 1}} (Az_k B_k^{\alpha_k})^{\varepsilon_j - 1} \frac{\zeta_j q_j}{\theta_j} + \Delta\mu_k$$
for each $k \in \mathcal{N}$ and update to $\Delta \mu_{p+1}^k = \psi \Delta \mu^k + (1 - \psi) \Delta \mu_p^k$ where $0 < \psi \leq 1$ is some parameter to control the speed of convergence.

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

In practice, it is useful to slow down the updating rule by setting $\psi = 0.9$.

Notice that this algorithm imposes that $\theta \in \{0, 1\}^n$ at every iteration. When the solution to $\mathcal{R}$ is not in $\{0, 1\}^n$, the algorithm does not converge and the status $\theta$ of some firms keeps alternating between 0 and 1. In practice, I stop the algorithm when the distance between $\Delta \mu_{p+1}^k$ and $\Delta \mu_p^k$ starts to increase, which usually indicates that there will be no convergence. I then look at the set of firms for which $\theta$ keeps alternating (different sign for $\Delta \mu_{p+1}^k$ and $\Delta \mu_p^k$), and then pick the best $\theta \in \{0, 1\}$ to maximize the planner’s objective function.

E.5 Construction of the matrix $\Omega$ in the calibrated economy

The matrix $\Omega$ is constructed by assuming that the number of potential incoming and outgoing connections $(x_{in}, x_{out})$, for any given firm, is drawn from a bivariate power law of the first kind $\mathcal{K}$ for which the joint density over $(x_{in}, x_{out})$ is $g(x_{in}, x_{out}) = \xi (\xi - 1) (x_{in} + x_{out} - 1)^{-((\xi + 1)}$. The full algorithm to construct the matrix is as follows:

1. Begin with $\Omega_{ij} = 0$ for all $i, j \in \mathcal{N}^2$.

2. For each firm $j \in \mathcal{N}$, draw from $\mathcal{K}$ a pair $(x_{in}^j, x_{out}^j)$ for the number of incoming and outgoing connections for $j$. Redraw until $\sum_j x_{in}^j = \sum_j x_{out}^j$ so that the total number of incoming connections is equal to the total number of outgoing connections in the economy.

3. For each $j \in \mathcal{N}$, create $x_{in}^j$ incoming stubs and $x_{out}^j$ outgoing stubs.

4. Randomly match each incoming stub to an outgoing stub. An incoming stub has the same probability of being matched with any outgoing stub. Set $\Omega_{ij} = 1$ where $i$ is the firm associated with the outgoing stub and $j$ is the firm associated with the incoming stub.

F Simulations with a large number of firms and aggregate shocks

To investigate how the model behaves under a more realistic parametrization, I simulate the calibrated economy with $n = 20,000$ firms and with aggregate shocks to total factor productivity $A$. The number of firms was chosen to roughly match the number of operating firms in the Factset data. I assume that $\log(A_t)$ follows an AR(1) process with an autocorrelation of 0.9 and a standard deviation parameter set to match empirical estimates about the impact of aggregate shocks on volatility.\footnote{Atalay (2017) generalizes an empirical strategy introduced by Foerster et al. (2011) to evaluate the impact of aggregate shocks on aggregate fluctuations. He finds that they account for 17% of volatility. I parametrize the stochastic process followed by $\log(A_t)$ to match that estimate.} Table 14 shows the correlations between aggregate output and the shape of the network.
We see that the numbers are broadly similar to those of the benchmark calibration. I also compute the difference in output volatility between the flexible and fixed networks in this setting. I find that the flexible network economy is about 13% less volatile. Finally, aggregate output is also 10% larger under the flexible network, roughly the same number as in the benchmark economy.

Table 14: Correlation between with aggregate output with \( n = 20,000 \) firms and aggregate shocks

<table>
<thead>
<tr>
<th>Network</th>
<th>Power law exponents</th>
<th>Clustering coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-degree</td>
<td>Out-degree</td>
</tr>
<tr>
<td>Model with ( n = 20,000 ) firms and aggregate shocks</td>
<td>-0.73</td>
<td>-0.73</td>
</tr>
<tr>
<td>Benchmark model</td>
<td>-0.59</td>
<td>-0.70</td>
</tr>
</tbody>
</table>

Notes: All time series are in logs. The parameters of the economy are as in the benchmark calibration except as mentioned in the text. Since these simulations are computationally intensive, I simulate four economies instead of twenty in the benchmark exercises.

G Changing \( \Omega \)

To see how the matrix \( \Omega \) affects the efficient allocation, I simulate the model under different parametrization for \( \Omega \). I still assume that \( \Omega \) is drawn from a bivariate power law of the first kind but I vary its exponent to \( \xi = 1.7 \) and \( \xi = 1.9 \). The results are presented in Table 15. We see from the table that changing \( \xi \) has a direct impact on the degree distributions and the global clustering coefficient in the efficient network. Under \( \xi = 1.7 \), the distribution of the number of potential connections features thicker tails such that \( \Omega \) offers a lot of options for the planner to create highly-connected firms and dense clusters of producers. The planner takes advantage of these possibilities to increase welfare: the mean of aggregate output is 14.7 under \( \xi = 1.7 \), but only 13.4 under \( \xi = 1.9 \).

Table 15: Impact of \( \Omega \) on the production network

<table>
<thead>
<tr>
<th>Network</th>
<th>Power law exponent</th>
<th>Clustering coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-degree</td>
<td>Out-degree</td>
</tr>
<tr>
<td>( \xi = 1.7 )</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>( \xi = 1.79 ) (benchmark)</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>( \xi = 1.9 )</td>
<td>1.12</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Notes: The parameters are the same as in the benchmark calibration except for the distribution from which \( \Omega \) is drawn (see text).
References


