Econometric Analysis of Production Networks with Dominant Units*

M. Hashem Pesaran

Department of Economics & USC Dornsife INET, University of Southern California, USA and Trinity College, Cambridge, UK

Cynthia Fan Yang

Department of Economics, University of Southern California

August 4, 2017

Abstract

This paper considers production and price networks with unobserved common factors, and derives an exact expression for the rate at which aggregate fluctuations vary with the dimension of the network. It introduces the notions of strongly and weakly dominant and non-dominant units, and shows that at most a finite number of units in the network can be strongly dominant. The pervasiveness of a network is measured by the degree of dominance of the most pervasive unit in the network, and is shown to be equivalent to the inverse of the shape parameter of the power law fitted to the network outdegrees. New cross-section and panel extremum estimators for the degree of dominance of individual units in the network are proposed and their asymptotic properties investigated. Using Monte Carlo techniques, the proposed estimator is shown to have satisfactory small sample properties. An empirical application to US input-output tables spanning the period 1972 to 2007 is provided which suggests that no sector in the US economy is strongly dominant. The most dominant sector turns out to be the wholesale trade with an estimated degree of dominance ranging from 0.72 to 0.82 over the years 1972-2007.

Keywords: aggregate fluctuations, strongly and weakly dominant units, spatial models, outdegrees, degree of pervasiveness, power law, input-output tables, US economy

JEL Classifications: C12, C13, C23, C67, E32

*We acknowledge helpful comments by Daron Acemoglu, Vasco Carvalho, Ron Smith, Martin Weidner, and participants at the Max King Conference at Monash University, the 2017 Asian Meeting of the Econometric Society, the third Vienna Workshop on High Dimensional Time Series in Macroeconomics and Finance, the fourth annual conference of the International Association for Applied Econometrics (IAAE), and seminar participants at University of Southern California. We would also like to thank Ida Johnsson for help with the compilation of input-output tables. Yang gratefully acknowledges the IAAE travel grant. Corresponding address: Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089, USA. E-mail addresses: pesaran@usc.edu (M. Hashem Pesaran), yang488@usc.edu (Cynthia Fan Yang).

1 Introduction

Over the past decade, there has been renewed interest in production networks and the role that individual production units (firms/sectors) can play in propagation of shocks across the economy. This literature builds on the multisectoral model of real business cycles pioneered by Long and Plosser (1983), and draws from a variety of studies on social and economic networks, including network games, cascades, and micro foundations of macro volatility. Notable theoretical contributions in this area include Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), Horvath (1998, 2000), Gabaix (2011), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2016), and Siavash (2016). Empirical evidence for such propagation mechanism is presented in Foerster, Sarte, and Watson (2011), Acemoglu, Akcigit, and Kerr (2016), and Carvalho, Nirei, Saito, and Tahbaz-Salehi (2016). One important issue in this literature relates to conditions under which sector-specific shocks are likely to have lasting aggregate (macro) effects. Similar issues arise in financial networks where it is of interest to ascertain if an individual bank can be considered as "too big to fail". Recent reviews are provided by Carvalho (2014) and Acemoglu, Ozdaglar and Tahbaz-Salehi (2016).

In this paper we consider a production network with unobserved common technological factors, and derive an associated price network which is dual to the production network, which we use to derive an exact characterization of the effect of sector-specific shocks on aggregate output. We show that sector-specific shocks have aggregate effects if there are "dominant" sectors in the sense that their outdegrees are not bounded in the number of production units, N, in the economy. The outdegree of a sector is defined as the share of that sector's output used as intermediate inputs by all other sectors in the economy. The degree of dominance (or pervasiveness) of a sector is measured by the exponent δ that controls the rate at which the outdegree of the sector in question rises with N. This measure turns out to be the same as the exponent of cross-sectional dependence introduced in Bailey et al. (2016), for the analysis of cross-section dependence in panel data models with large cross-section and time dimensions.

Our approach differs from Acemoglu et al. (2012) in three important respects. First, we provide a more general setting that allows for unobserved common factors and derive a spatial model in sectoral prices that can be taken directly to the data. We establish a one-to-one relationship between the pervasiveness of price shocks and aggregate output shocks. Second, Acemoglu et al. (2012) express the aggregate output as a reduced form function of the sector-specific shocks, based on which they are only able to derive a lower bound to the decay rate of sector-specific shocks on aggregate outcomes. They consider the first- and second-order effects, and acknowledge that ignoring higher-order interconnections might bias the results. In contrast, the present paper provides an exact expression for the effects of sector-specific shocks on aggregate fluctuations, and shows that its rate of decay only depends on the extent to which the dominant unit (sector) is pervasive, namely the one with the largest δ, denoted by $\delta_{\max}$. We derive upper as well as lower bounds for the rate of convergence of the variability of aggregate output in terms of N, and show that these bounds converge at the same rate, and thus establish an exact rate of convergence for aggregate output variability. Finally, Acemoglu et al. (2012) do not identify the dominant unit(s). Instead, they approximate the tail distribution, for some given cut-off value, of the outdegrees by a power law distribution and provide estimates for the shape parameters. By contrast, we propose a nonparametric approach, which is applicable irrespective of whether the outdegrees are Pareto distributed, and does not require knowing the cut-off value above which the Pareto tail behavior begins. The inverse of the proposed estimator of $\delta_{\max}$ is an extremum estimator of the shape parameter of the Pareto distribution, β. It is simple to compute and is given by the average log of the largest outdegree relative to all other outdegrees, scaled by the size of the network, N. In addition, our approach also allows us to identify the most dominant units and their degrees of dominance, $\delta_{(i)}$, where $\delta_{(1)} = \delta_{\max} \geq \delta_{(2)} \geq ...$, for all $\delta_{(i)} > 1/2$.

Small sample properties of the extremum estimator are investigated by Monte Carlo techniques and are shown to be satisfactory. A comparison of the estimates of the shape parameter β based on Pareto distribution with the estimates based on the inverse of the extremum estimator of $\delta_{\max}$, shows that the latter performs much better, particularly when N is large (300+). Furthermore, the extremum estimator is shown to perform well even under a Pareto tail distribution, whereas the commonly used estimators of the shape parameter, β, display substantial biases if the true underlying distribution is non–Pareto.

Application of our estimation procedure to US input-output tables over the period 1972-2002 yields yearly estimates of $\delta_{\max}$ that lie between 0.72 and 0.82. These estimates are by and large close to the inverse of the estimates of the shape parameter β considered in Acemoglu et al. (2012) when a 20% cut-off value is used, although the log-log regression estimates of β tend to be highly sensitive to the choice of the cut-off values and the different orders of interconnections considered. To provide more reliable estimates, we also conduct panel estimation and find that the largest estimate of $\delta_{\max}$ is about 0.76 for the sub-sample covering 1972-1992 and 0.72 for the sub-sample covering 1997-2007. Quite remarkably, we find that estimates of $\delta_{\max}$ and the identity of the dominant sector are rather stable throughout the period from 1972 to 2007, with the wholesale trade sector identified as the most dominant sector for all years except for the year 2002 when the wholesale trade is estimated to be the second most dominant sector. Our estimates also suggest that no sector in the US economy is strongly dominant, which requires the value of $\delta_{\max}$ to be close to unity, whilst the largest estimate we obtain is around 0.8. Overall, our analyses support the view that sector-specific shocks have some macro effects, but we do not find such effects to be sufficiently strong.

The rest of the paper is organized as follows. Section 2 presents the production network. Section 3 derives the associated price network. Section 4 introduces the concepts of strongly and weakly dominant, and non-dominant units, and network pervasiveness. Section 5 derives exact conditions under which micro (sectoral) shocks can lead to aggregate fluctuations. Section 6 shows the relation between the degree of network pervasiveness and the shape parameter of the power law distribution. Section 7 introduces the extremum estimator, derives its asymptotic distribution, and shows its robustness to the choice of the underlying distribution. Section 8 provides evidence on the small sample properties of the alternative estimators of $\delta_{\max}$ using a number of Monte Carlo experiments. Section 9 presents the empirical application, and Section 10 concludes. Some of the mathematical details and a brief discussion of data sources are provided in the Appendix. An extension of the extremum estimator to large N and T panels, and additional Monte Carlo results are provided in an Online Supplement.

Notations: The total number of cross section units (sectors) in the economy is denoted by N, which is then decomposed into m dominant units and n non-dominant units. The number of dominant units is also decomposed into strongly dominant units and weakly dominant units. (See Definition 1). If $\{f_N\}_{N=1}^{\infty}$ is any real sequence and $\{g_N\}_{N=1}^{\infty}$ is a sequence of positive real numbers, then $f_N = O(g_N)$ if there exists a positive finite constant K such that $|f_N|/g_N \leq K$ for all N. $f_N = o(g_N)$ if $f_N/g_N \to 0$, as $N \to \infty$. If $\{f_N\}_{N=1}^{\infty}$ and $\{g_N\}_{N=1}^{\infty}$ are both positive sequences of real numbers, then $f_N = \ominus(g_N)$ if there exists $N_0 \geq 1$ and positive finite constants K_0 and K_1, such that $\inf_{N \geq N_0}(f_N/g_N) \geq K_0$, and $\sup_{N \geq N_0}(f_N/g_N) \leq K_1$. $\varrho(\mathbf{A})$ is the spectral radius of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$, defined as $\varrho(\mathbf{A}) = \max\{|\lambda_i|, i = 1, 2, ..., N\}$, where λ_i is an eigenvalue of $\mathbf{A}$ and $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq \cdots \geq |\lambda_N(\mathbf{A})|$. $\|\mathbf{A}\|_\infty = \max\limits_{1 \leq i \leq N} \sum_{j=1}^{N} |a_{ij}|$ and $\|\mathbf{A}\|_1 = \max\limits_{1 \leq j \leq N} \sum_{i=1}^{N} |a_{ij}|$ are the maximum row sum norm and the maximum column sum norm of matrix $\mathbf{A}$, respectively. K is used for a generic finite positive constant not depending on N. δ_i denotes the degree of dominance (or pervasiveness) of unit i in a network, where $i = 1, 2, ..., N$, and $0 \leq \delta_i \leq 1$.

2 Production network

To show how the two strands of literature on production networks and cross-sectional dependence are related, we begin with a panel version of the input-output model developed in Acemoglu et al. (2012). Our goal is to provide an exact characterization of the effect of unit-specific shocks on aggregate output. We assume that production of sector i at time t, q_{it}, is determined by the following Cobb-Douglas production function subject to constant returns to scale

$$q_{it} = e^{\alpha u_{it}} l_{it}^{\alpha} \prod_{j=1}^{N} q_{ij,t}^{\rho w_{ij}}, \quad \text{for} \;\; i = 1, 2, \ldots, N; \; t = 1, 2, ..., T, \tag{1}$$

where the productivity shock, u_{it}, is given by

$$u_{it} = \varepsilon_{it} + \gamma_i f_t,$$

and is composed of a sector-specific shock, ε_{it}, and a common technological factor, f_t. The factor loading, γ_i, measures the importance of technological change on sector i. Following Bailey et al. (2016), we denote the cross-section exponent of the factor loadings by δ_γ, defined by the value of δ_γ that ensures

$$\lim_{N \to \infty} N^{-\delta_\gamma} \sum_{i=1}^{N} |\gamma_i| = c_\gamma > 0, \tag{2}$$

where c_γ is a finite constant, bounded in N. The standard factor model sets $\delta_\gamma = 1$, and treats the common factor as 'strong' or 'pervasive', in the sense that changes in f_t affects all sectors of the economy. But in what follows we shall also consider cases where $\delta_\gamma < 1$. In the case where the factor loadings are random we shall assume that $E(\gamma_i) = \gamma \neq 0$, and $Var(\gamma_i) = \sigma_\gamma^2 > 0$. The analysis can be easily extended to allow for multiple factors without additional complexity.

In line with Acemoglu et al. (2012), we shall assume that the sector-specific shocks are cross-sectionally independent with zero means and finite variances, $Var(\varepsilon_{it}) = \sigma_i^2$, such that $0 < \underline{\sigma}^2 < \sigma_i^2 < \bar{\sigma}^2 < K < \infty$. The independence assumption is not necessary and can be relaxed by assuming that the errors are cross-sectionally weakly dependent. We also assume that ε_{it} are serially uncorrelated, although this is not essential for our main theoretical results, and without loss of generality we assume that common and sector-specific shocks are uncorrelated.

Returning to the other factors in the production function, l_{it} denotes the labor input, $q_{ij,t}$ is the amount of output of sector j used in production of sector i, α denotes the share of labor, $\rho = 1 - \alpha$, and ρw_{ij} is the share of j^{th} output in the i^{th} sector. The amount of final goods, c_{it}, are defined by

$$c_{it} = q_{it} - \sum_{j=1}^{N} q_{ji,t}, \quad i = 1, 2, \ldots, N, \tag{3}$$

which are consumed by a representative household with the Cobb-Douglas preferences

$$u\left(c_{1t,}c_{2t}, \ldots, c_{Nt}\right) = A \prod_{i=1}^{N} c_{it}^{1/N}, \ A > 0. \tag{4}$$

We further assume that the aggregate labor supply, l_t, is given exogenously and labor markets clear

$$l_t = \sum_{i=1}^{N} l_{it}. \tag{5}$$

Let $P_{1t}, P_{2t}, \ldots, P_{Nt}$ be the sectoral equilibrium prices, $Wage_t$ the equilibrium wage rate, and denote their logarithms by $p_{it} = \log\left(P_{it}\right)$, $\omega_t = \log\left(Wage_t\right)$. Then it can be shown that in the competitive equilibrium the logarithm of real wage, which is taken as a measure of GDP in the literature, is given by

$$\omega_t - \overline{p}_t = \mu + \left(\boldsymbol{\upsilon}_N' \boldsymbol{\gamma}\right) f_t + \boldsymbol{\upsilon}_N' \boldsymbol{\varepsilon}_t, \tag{6}$$

where $\overline{p}_t$ is the aggregate log price index defined by,

$$\overline{p}_t = N^{-1} \sum_{i=1}^{N} p_{it} = N^{-1} \boldsymbol{\tau}_N' \mathbf{p}_t, \tag{7}$$

$\mathbf{p}_t = (p_{1t}, p_{2t,} \ldots, p_{Nt})'$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$, and

$$\boldsymbol{\upsilon}_N = (\upsilon_1, \upsilon_2, \ldots, \upsilon_N)' = \frac{\alpha}{N} \left(\mathbf{I}_N - \rho \mathbf{W}'\right)^{-1} \boldsymbol{\tau}_N, \tag{8}$$

where $\mathbf{W}$ is the $N \times N$ matrix $\mathbf{W} = (w_{ij})$, and $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. μ is a constant independent of f_t and $\boldsymbol{\varepsilon}_t$, which is given by

$$\mu = \alpha^{-1}\left[\alpha \log(\alpha) + \rho \log(\rho) + \rho \sum_{i=1}^{N}\sum_{j=1}^{N} \upsilon_i w_{ij} \log(w_{ij})\right].$$

The (log) real-wage equation, (6), generalizes equation (3) in Acemoglu et al. (2012) by allowing for time variations in prices and technologies. By normalizing $\overline{p}_t$ such that $\overline{p}_t = -\mu$ and ignoring the common factor, Acemoglu et al. (2012) concentrate on $\omega_t = \boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t$, as a measure of aggregate output, and refer to $\boldsymbol{\upsilon}_N$ as the 'influence vector' (their equation (4)), and show that $\upsilon_i \geq 0$, and $\boldsymbol{\tau}'_N \boldsymbol{\upsilon}_N = 1$.[1] They measure aggregate volatility by the standard deviation of aggregate output, namely $\left[Var\left(\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t\right)\right]^{1/2}$, and focus on the asymptotic properties of $\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t$, as $N \to \infty$. Since $Var\left(\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t\right) = \boldsymbol{\upsilon}'_N Var\left(\boldsymbol{\varepsilon}_t\right) \boldsymbol{\upsilon}_N$, it follows that

$$\bar{\sigma}^2\left(\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N\right) \geq Var\left(\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t\right) \geq \underline{\sigma}^2\left(\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N\right),$$

and hence the asymptotic properties of $Var\left(\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t\right)$ is governed by that of $\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N$. Acemoglu et al. (2012, p.2009) derive a lower bound for $\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N$ and show that[2]

$$\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N \geq k_{\alpha,0} N^{-1} + k_{\alpha,1} N^{-2} \sum_{j=1}^{N} d_j^2, \tag{9}$$

where $k_{\alpha,0}$ and $k_{\alpha,1}$ are finite constants that depend on α, and d_j is defined by $d_j = \sum_{i=1}^{N} w_{ij}$, which is the j^{th} column sum of $\mathbf{W}$. Therefore, to analyze the limiting behavior of $\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N$, it is sufficient to consider the limiting behavior of $N^{-2} \sum_{j=1}^{N} d_j^2$. This analysis is carried out in some detail by Acemoglu et al. (2012). But as we shall see, it is also important to consider the limiting behavior of individual column sums of $\mathbf{W}$, and in particular to identify the ones that rise with N, as distinguished from those that are bounded in N.

3 Price network

Instead of analyzing the aggregate output directly in terms of the sector-specific shocks, we derive a price network which is dual to the production network discussed in Section 2. By doing so, we are able to obtain an exact expression for the decay rate of aggregate volatility, rather than just a lower bound. Given sector prices, $P_{1t}, P_{2t}, \ldots, P_{Nt}$, and the wage rate, $Wage_t$, solving sector i's problem leads to

$$q_{ij,t} = \frac{\rho w_{ij} P_{it} q_{it}}{P_{jt}}, \tag{10}$$

and

$$l_{it} = \frac{\alpha P_{it} q_{it}}{Wage_t}. \tag{11}$$

[1] See Appendix A of Acemoglu et al. (2012).

[2] These authors also consider higher-order interconnection terms which they include on the right-hand-side of $\boldsymbol{\upsilon}'_N \boldsymbol{\upsilon}_N$, but these terms are dominated by $N^{-2} \sum_{j=1}^{N} d_j^2$.

Substituting the above results in (1) and simplifying yields

$$p_{it} = \rho \sum_{j=1}^{N} w_{ij} p_{jt} + \alpha \omega_t - b_i - \alpha \left(\gamma_i f_t + \varepsilon_{it} \right), \tag{12}$$

where the price-specific intercepts, b_i, depend only on α, ρ and the weight matrix $\mathbf{W}$,

$$b_i = \alpha \log(\alpha) + \rho \log(\rho) + \rho \sum_{j=1}^{N} w_{ij} \log(w_{ij}), \tag{13}$$

for $i = 1, 2, \ldots, N$. In cases where $w_{ij} = 0$, we set $w_{ij} \log(w_{ij}) = 0$ as well. In matrix notation the 'price network', (12), can be written as

$$\mathbf{p}_t = \rho \mathbf{W} \mathbf{p}_t + \alpha \omega_t \boldsymbol{\tau}_N - \left(\mathbf{b} + \alpha \boldsymbol{\gamma} f_t + \alpha \boldsymbol{\varepsilon}_t \right), \tag{14}$$

where $\mathbf{b} = (b_1, b_2, \ldots, b_N)'$.

A dual to the price equation in (12) can also be obtained using (10) in (3) to obtain

$$S_{it} = \rho \sum_{j=1}^{N} w_{ji} S_{jt} + C_{it}, \tag{15}$$

where $C_{it} = P_{it} c_{it}$, and $S_{it} = P_{it} q_{it}$ is the sales of sector i. The sales equation, (15), can also be written as

$$\mathbf{S}_t = \rho \mathbf{W}' \mathbf{S}_t + \mathbf{C}_t, \tag{16}$$

where $\mathbf{S}_t = (S_{1t}, S_{2t}, \ldots, S_{Nt})'$ and $\mathbf{C}_t = (C_{1t}, C_{2t}, \ldots, C_{Nt})'$. Note that $\mathbf{W}$ enters as its transpose, $\mathbf{W}'$, in (16) as compared to the price equations in (14).

Aggregating (11) over i, we have

$$Wage_t \sum_{i=1}^{N} l_{it} = \alpha \sum_{i=1}^{N} P_{it} q_{it},$$

or

$$l_t \, Wage_t = \alpha \sum_{i=1}^{N} S_{it} = \alpha \boldsymbol{\tau}_N' \mathbf{S}_t. \tag{17}$$

Also using (16)

$$\mathbf{S}_t = \left(\mathbf{I}_N - \rho \mathbf{W}' \right)^{-1} \mathbf{C}_t, \tag{18}$$

where $\left(\mathbf{I}_N - \rho \mathbf{W}' \right)^{-1}$ is known as the Leontief inverse.[3] Using (18) in (17) now yields the following expression for the total wage bill,

$$l_t \, Wage_t = \alpha \boldsymbol{\tau}_N' \left(\mathbf{I}_N - \rho \mathbf{W}' \right)^{-1} \mathbf{C}_t. \tag{19}$$

[3] A proof that the Leontief matrix is invertible even in the presence of dominant units is provided in Lemma 1 of Appendix A.

Similarly, solving (14) for the log-price vector, $\mathbf{p}_t$, and applying Lemma 1 in Appendix A we have

$$\mathbf{p}_t = \alpha\omega_t \left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\tau}_N - \alpha \left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\xi}_t, \tag{20}$$

where $\boldsymbol{\xi}_t = \alpha^{-1}\mathbf{b} + \boldsymbol{\gamma} f_t + \boldsymbol{\varepsilon}_t$. Then the aggregate log price index, $\overline{p}_t$, defined in (7), is given by

$$\overline{p}_t = \left[\frac{\alpha}{N}\boldsymbol{\tau}'_N \left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\tau}_N\right]\omega_t - \frac{\alpha}{N}\boldsymbol{\tau}'_N \left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\xi}_t. \tag{21}$$

But since $w_{ij} \geq 0$, $\mathbf{W}\boldsymbol{\tau}_N = \boldsymbol{\tau}_N$, and $0 < \alpha < 1$, then $\left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\tau}_N = \boldsymbol{\tau}_N/\alpha$, and hence (21) can also be written as

$$\omega_t - \overline{p}_t = \boldsymbol{\upsilon}'_N \boldsymbol{\xi}_t, \tag{22}$$

where $\boldsymbol{\upsilon}_N$ is influence vector given by (8). Now let

$$\mathbf{x}_t = \mathbf{p}_t - \omega_t \boldsymbol{\tau}_N, \tag{23}$$

and rewrite (14) in terms of log price-wage ratios, $\mathbf{x}_t$, as

$$\mathbf{x}_t = \rho\mathbf{W}\mathbf{x}_t - \mathbf{b} - \alpha\left(\boldsymbol{\gamma} f_t + \boldsymbol{\varepsilon}_t\right). \tag{24}$$

Equation (24) represents a first-order spatial autoregressive (SAR(1)) model with an unobserved common factor.

Consider now the following simple average over the units, x_{it}, for $i = 1, 2, ..., N$, in the above network

$$\overline{x}_{N,t} = \frac{1}{N}\boldsymbol{\tau}'_N \mathbf{x}_t = -\left(\omega_t - \bar{p}_t\right),$$

which is the negative of the aggregate output measure, defined by (6). Also, using (22) we have

$$\omega_t - \bar{p}_t = -\overline{x}_{N,t} = \alpha^{-1}\left(\boldsymbol{\upsilon}'_N \mathbf{b}\right) + \left(\boldsymbol{\upsilon}'_N \boldsymbol{\gamma}\right) f_t + \boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t, \tag{25}$$

which fully specifies the dependence of aggregate output on both common and sector-specific shocks.

Note that equations (19) and (25) are dual of each other. (19) gives the total wage bill in terms of a weighted sum of consumption expenditures, with the weights given by $\alpha\left(\mathbf{I}_N - \rho\mathbf{W}\right)^{-1} \boldsymbol{\tau}_N$. Whilst (25) gives the log of the real wage rate in terms of the common and aggregate sectoral shocks, namely f_t and $\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t$. Recall that common and sectoral shocks are assumed to be uncorrelated. The key issue is how much of the cyclical fluctuations in (log) real wages, $Var\left(\overline{x}_{N,t}\right)$, is due to common shocks $\left(\boldsymbol{\upsilon}'_N \boldsymbol{\gamma}\right)^2 Var\left(f_t\right)$, and how much is due to sectoral shocks, $Var\left(\boldsymbol{\upsilon}'_N \boldsymbol{\varepsilon}_t\right)$.

There are two advantages in directly focusing on the price network model, (24). First, it allows us to relate the production network to the literature on spatial econometrics that should facilitate the econometric analysis of production networks, and allows us to address more easily the issues of identification and estimation of the structural parameters α, $\boldsymbol{\gamma}$ and σ_i^2, for $i = 1, 2, ..., N$.[4] The direct use of the SAR model, (24), also enables us

[4]For example, see the recent contributions of Bai and Li (2013) and Yang (2016) on estimation of SAR models with unobserved common factors.

to provide exact bounds on $Var(\bar{x}_{N,t}) = Var(\omega_t - \bar{p}_t)$ rather than the lower bounds obtained by Acemoglu et al. (2012). Instead, by considering the price network explicitly we are able to show that at most only a few sectors can have significant aggregate effects, and these sectors are those with outdegrees that rise with N. The rate at which the outdegrees rise with N could very well differ across sectors and it is important that such sectors are identified and their empirical contribution to aggregate fluctuations evaluated.

4 Degrees of dominance of units in a network and network pervasiveness

Consider a network represented by the $N \times N$ adjacency matrix $\mathbf{W} = (w_{ij})$, where $w_{ij} \geq 0$ for all i and j, and $\mathbf{W}$ is row-normalized such that $\sum_{j=1}^{N} w_{ij} = 1$, for all i. Denote the j^{th} column of $\mathbf{W}$ by $\mathbf{w}_{.j}$ and the associated column sum by $d_j = \boldsymbol{\tau}'_N \mathbf{w}_{.j}$, the outdegree of unit j. The outdegree is one of many network centrality measures considered in the literature. The most widely used centrality measures is degree centrality, which refers to the number of ties a node has, and in a directed network can be classified into indegree and outdegree. The indegree counts the number of ties a node receives, and the outdegree counts the number of ties a node directs to others. In this paper, we are focusing on how the weighted outdegree vary with N and normalize the weighted indegree (row sums of $\mathbf{W}$) to one, because we are interested in studying the influence of a unit to other downstream units. Other centrality measures, including closeness, betweenness, and eigenvector centralities, are not relevant for our purpose, since we aim to characterize the effects of idiosyncratic shocks to a unit on some aggregate measure of the network, rather than the pattern of interdependencies of the network. To this end, we introduce the notions of strongly and weakly dominant units in the following definition. We consider units with nonzero outdegrees and assume throughout that $d_j > 0$, for all j.

Definition 1. (δ-dominance) We shall refer to unit j of the row-standardized network $\mathbf{W} = (w_{ij} \geq 0)$ as δ_j-*dominant* if its (weighted) outdegree, $d_j = \sum_{i=1}^{N} w_{ij} > 0$, is of order N^{δ_j}, where δ_j is a fixed constant in the range $0 \leq \delta_j \leq 1$. More specifically,

$$d_j = \kappa_j N^{\delta_j}, \text{ for } j = 1, 2, ..., N, \tag{26}$$

where κ_j is a fixed positive constant which does not depend on N. The unit j is said to be *strongly dominant* if $\delta_j = 1$, *weakly dominant* if $0 < \delta_j < 1$, and *non-dominant* if $\delta_j = 0$. We refer to δ_j as the degree of dominance of unit j in the network.

Remark 1. *Since $d_j > 0$ for all j, we must have $\kappa_j > 0$ and $\kappa_{\min} = \min(\kappa_1, \kappa_2, ..., \kappa_N) > 0$. It is also worth noting that δ_j is identified by requiring that κ_j and δ_j do not depend on N.*

In the standard case where the column sum of $\mathbf{W}$ is bounded in N we must have $\delta_j = 0$ for all j, that is, all units are non-dominant. $\mathbf{W}$ will have an unbounded column sum if $\delta_j > 0$ for at least one j. But due to the bounded nature of the rows of $\mathbf{W}$, not all columns of $\mathbf{W}$ can be δ-dominant with $\delta_j > 0$ for all j. To see this, let $\mathbf{d} = (d_1, d_2, ..., d_N)' = \mathbf{W}'\boldsymbol{\tau}_N$, and note that

$$\boldsymbol{\tau}'_N \mathbf{d} = \boldsymbol{\tau}'_N \mathbf{W}' \boldsymbol{\tau}_N = N. \tag{27}$$

Hence, there must exist $0 < \kappa_j < K < \infty$ for $j = 1, 2, ..., N$, such that

$$\sum_{j=1}^{N} \kappa_j N^{\delta_j} = N, \tag{28}$$

for a fixed N and as $N \to \infty$. Let $\delta_{\min} = \min(\delta_1, \delta_2, ..., \delta_N)$, and note that

$$N = \sum_{j=1}^{N} \kappa_j N^{\delta_j} \geq N \kappa_{\min} N^{\delta_{\min}},$$

which in turn implies

$$\kappa_{\min} N^{\delta_{\min}} \leq 1. \tag{29}$$

Since by assumption $\kappa_{\min} > 0$ and $\delta_{\min} \geq 0$, it is clear that (29) cannot be satisfied for all values of N unless $\delta_{\min} = 0$, which establishes that not all units in a given network can be dominant. This result is summarized in the following proposition.

Proposition 1. *Consider the network represented by $\mathbf{W} = (w_{ij} \geq 0)$, and assume that $\mathbf{W}$ is row-standardized. Suppose that the outdegrees of the network, $d_j = \sum_{i=1}^{N} w_{ij}$, are non-zero ($d_j > 0$) and follow the power function, (26), with δ_j being the degree of dominance of unit j in the network. Then not all units of the network can be δ-dominant, with $\delta_j > 0$ for all j.*

Let $S_N = N^{-1} \sum_{j=1}^{N} \kappa_j N^{\delta_j}$, and note that since $\kappa_j > 0$ for all j, $\kappa_{\min} = \min_j(\kappa_j) > 0$, and hence

$$S_N = N^{-1} \sum_{j=1}^{N} \kappa_j e^{\delta_j \ln N} \geq \kappa_{\min} N^{-1} \sum_{j=1}^{N} e^{\delta_j \ln N}. \tag{30}$$

Now using a Taylor series expansion of $e^{\delta_j \ln N}$, we obtain

$$\begin{aligned} \sum_{j=1}^{N} e^{\delta_j \ln N} &= \sum_{j=1}^{N} \left[1 + \sum_{s=1}^{\infty} \frac{\delta_j^s (\ln N)^s}{s!} \right] \\ &= N + \sum_{s=1}^{\infty} \frac{(\ln N)^s}{s!} \left(\sum_{j=1}^{N} \delta_j^s \right), \end{aligned} \tag{31}$$

which if substituted in (30) yields

$$S_N \geq \kappa_{\min} \left[1 + \sum_{s=1}^{\infty} \frac{\left(\sum_{j=1}^{N} \delta_j^s \right) (\ln N)^s}{s! N} \right]. \tag{32}$$

Since $S_N = 1$, and all the summands over s in (32) are nonnegative as $\delta_j \geq 0$ and $\ln N > 0$, it is necessary that

$$\frac{\left(\sum_{j=1}^{N} \delta_j^s\right) (\ln N)^s}{s!N} \to 0, \text{ as } N \to \infty, \text{ for all } s = 1, 2, 3, ... \tag{33}$$

Also note that for any finite s, $(\ln N)^s / (s!N) \to 0$, as $N \to \infty$, and since

$$\sum_{s=1}^{\infty} \frac{(\ln N)^s}{s!N} = \frac{N-1}{N} \to 1, \text{ as } N \to \infty,$$

then it must be that $(\ln N)^s / (s!N) \to 0$, as $N \to \infty$, for all s, including $s \to \infty$. Furthermore, since $0 \leq \delta_j \leq 1$ then

$$\sum_{j=1}^{N} \delta_j^s \leq \sum_{j=1}^{N} \delta_j, \text{ for } s \geq 1, \tag{34}$$

and

$$\frac{\left(\sum_{j=1}^{N} \delta_j^s\right) (\ln N)^s}{s!N} \leq \left(\sum_{j=1}^{N} \delta_j\right) \frac{(\ln N)^s}{s!N}.$$

Hence, for conditions in (33) to be met it is sufficient that $\{\delta_j\}$ is summable, namely

$$\sum_{j=1}^{N} \delta_j < K < \infty. \tag{35}$$

As we shall see, this condition plays an important role in the proof of consistency of the extremum estimator proposed in the sub-section 7.2 below.

Suppose now that m units are strongly dominant with $\delta_j = 1$, and the rest are non-dominant with $\delta_j = 0$. Then using (32) we have

$$S_N \geq \kappa_{\min} \left[1 + m \sum_{s=1}^{\infty} \frac{(\ln N)^s}{s!N}\right] = \kappa_{\min} \left[1 + m \left(\frac{N-1}{N}\right)\right],$$

and since $S_N = 1$, it follows that m cannot rise with N, and must be a fixed integer.

In the case where m units are dominant with $\delta_j > 0$, then m must be finite if the summability condition given by (35) is to hold. For example, suppose that only m units are dominant. Then $\sum_{j=1}^{N} \delta_j \geq m\delta_{\min} > 0$, and from the summability condition (35) we have $K > \sum_{j=1}^{N} \delta_j \geq m\delta_{\min}$, from which it follows that $m \leq K/\delta_{\min}$ which is bounded in N. These findings are summarized in the next proposition.

Proposition 2. *Consider the network represented by* $\mathbf{W} = (w_{ij} \geq 0)$, *and assume that* $\mathbf{W}$ *is row-standardized, and the outdegrees of the network,* $d_j = \sum_{i=1}^{N} w_{ij}$, *are non-zero*

$(d_j > 0)$. Then the number of strongly dominant units must be fixed and cannot rise with N. Moreover, if $\{\delta_j\}$ are summable, $\sum_{j=1}^{N} \delta_j < K < \infty$, then the number of dominant units with $\delta_j \neq 0$ must be finite, where δ_j is the degree of dominance of unit j in the network.

Remark 2. *Analogous results have also been found in Chudik, Pesaran, and Tosetti (2011) regarding the possible number of strong factors, and in Chudik and Pesaran (2013) on the number of dominant units in large dimensional vector autoregressions.*

Using the concept of δ-dominance of units in a given network, we now introduce the idea of network pervasiveness which is relevant for characterization of the degree to which shocks to an individual unit diffuse across the other units in the network.

Definition 2. (Network pervasiveness) Degree of pervasiveness of a given row-standardized network, $\mathbf{W} = (w_{ij} \geq 0, \sum_{j=1}^{N} w_{ij} = 1)$, is defined by $\delta_{\max} = \max(\delta_1, \delta_2, ..., \delta_N)$, where δ_j is the degree of dominance of its j^{th} unit.

Using the above concepts we now consider the rate at which $Var(\overline{x}_{N,t})$ varies with N, and show that it is governed by the pervasiveness of the network, measured by $\delta_{\max}$, and the exponent of the common factor, δ_γ, defined by (2). For unit-specific shocks to dominate the macro or common factor shocks we need $\delta_{\max} > \delta_\gamma > 1/2$. We also show how our measure of network pervasiveness, $\delta_{\max}$, is related to β, the shape parameter of the Pareto distribution fitted to the ordered outdegrees, $d_{(i)}$, for $i = 1, 2, ..., N$.

5 Price networks with one dominant unit and aggregate fluctuations

Consider the price network (24), and assume that it contains 1 dominant unit and $n = N - 1$ non-dominant units. The analysis can be readily extended to networks with m dominant units (m fixed), but to simplify the exposition here we confine our analysis to networks with one dominant unit. (The derivations for the general case is provided in Appendix B). Without loss of generality suppose the first element of $\mathbf{x}_t$, namely x_{1t}, is the dominant unit, and write (24) in the partitioned form as (setting $w_{11} = 0$)

$$\begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} 0 & \rho\mathbf{w}_{12}' \\ \rho\mathbf{w}_{21} & \rho\mathbf{W}_{22} \end{pmatrix} \begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} + \begin{pmatrix} g_{1t} \\ \mathbf{g}_{2t} \end{pmatrix}, \tag{36}$$

where $\mathbf{x}_{2t} = (x_{2t}, x_{3t}, ..., x_{Nt})'$, $\mathbf{w}_{21} = (w_{21}, w_{31}, ..., w_{N1})'$, $\mathbf{w}_{12} = (w_{12}, w_{13}, ..., w_{1N})'$, $\mathbf{g}_{2t} = (g_{2t}, g_{3t}, ..., g_{Nt})'$, and $g_{it} = -b_i - \alpha(\gamma_i f_t + \varepsilon_{it})$, for $i = 1, 2, ..., N$. $\mathbf{W}_{22}$ is the $n \times n$ weight matrix associated with the n non-dominant units and is assumed to satisfy the condition $|\rho| \left\|\mathbf{W}_{22}\right\|_1 < 1$. Furthermore, note that since

$$\mathbf{W}\boldsymbol{\tau}_N = \begin{pmatrix} 0 & \mathbf{w}_{12}' \\ \mathbf{w}_{21} & \mathbf{W}_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\tau}_n \end{pmatrix} = \begin{pmatrix} 1 \\ \boldsymbol{\tau}_n \end{pmatrix},$$

then $\mathbf{w}_{12}'\boldsymbol{\tau}_n = 1$, and $\boldsymbol{\tau}_n - \mathbf{w}_{21} = \mathbf{W}_{22}\boldsymbol{\tau}_n$. The latter result states that the i^{th} row sum of $\mathbf{W}_{22}$ is given by $1 - w_{i1} \leq 1$, and considering that $0 \leq w_{i1} < 1$, then we must have

$\|\mathbf{W}_{22}\|_{\infty} \leq 1$, which also establishes that $\varrho(\mathbf{W}_{22}) \leq 1$, where $\varrho(\mathbf{A})$ denotes the spectral radius of $\mathbf{A}$. Under the assumption that $|\rho| < 1$, by Lemma 1 in Appendix A the system of equations (36) has a unique solution given by

$$\begin{aligned} \begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} &= \begin{pmatrix} 1 & -\rho\mathbf{w}_{12}' \\ -\rho\mathbf{w}_{21} & \mathbf{S}_{22} \end{pmatrix}^{-1} \begin{pmatrix} g_{1t} \\ \mathbf{g}_{2t} \end{pmatrix} \\ &= \mathbf{S}^{-1}(\rho)\mathbf{g}_t, \end{aligned} \tag{37}$$

where $\mathbf{S}_{22} = \mathbf{I}_n - \rho\mathbf{W}_{22}$. In addition, since $|\rho| \, \|\mathbf{W}_{22}\|_1 < 1$, it follows from Lemma 2 in Appendix A that $\mathbf{S}_{22}^{-1}$ has bounded row and column norms.

For future reference also note that the $(1,1)^{th}$ element of $\mathbf{S}^{-1}(\rho)$ is given by ζ_1^{-1}, where

$$\zeta_1 = 1 - \rho^2 \mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\mathbf{w}_{21} \neq 0. \tag{38}$$

Finally, to allow unit 1 to be δ-dominant we consider the following exponent formulation

$$d_1 = \sum_{i=2}^{N} w_{i1} = \kappa_1 N^{\delta_1}, \tag{39}$$

where d_1 is allowed to rise with N, with $\kappa_1 > 0$ and $0 < \delta_1 \leq 1$. Recall that κ_1 and δ_1 can not vary with N.[5]

The system of equations (36) can now be solved for $\mathbf{x}_{2t}$ in terms of x_{1t}, namely (recall that by assumption $|\rho| \, \|\mathbf{W}_{22}\|_1 < 1$)

$$\mathbf{x}_{2t} = x_{1t}\rho\left(\mathbf{S}_{22}^{-1}\mathbf{w}_{21}\right) + \mathbf{S}_{22}^{-1}\mathbf{g}_{2t}, \tag{40}$$

and[6]

$$x_{1t} = \zeta_1^{-1}\left(g_{1t} + \rho\mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}\right). \tag{41}$$

Using the above in (40), we now have

$$\mathbf{x}_{2t} = (\rho/\zeta_1)\left(g_{1t} + \rho\mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}\right)\mathbf{S}_{22}^{-1}\mathbf{w}_{21} + \mathbf{S}_{22}^{-1}\mathbf{g}_{2t}. \tag{42}$$

The first term of $\mathbf{x}_{2t}$ refers to the direct and indirect effects of the dominant unit, and the second term relates to the network dependence of the non-dominant units.

Our primary focus is the extent to which shocks to individual units affect aggregate measures over the network. A standard aggregate measure is cross-section averages of x_{it} over $i = 1, 2, ..., N$. Here we consider the simple average

$$\bar{x}_{N,t} = \frac{x_{1t} + \sum_{i=2}^{N} x_{it}}{N} = \frac{x_{1t} + \boldsymbol{\tau}_n'\mathbf{x}_{2t}}{N},$$

but our analysis equally applies to weighted averages, $x_{N,t}^* = \sum_{i=1}^{N} \varpi_i x_{it}$, so long as the weights ϖ_i are granular in the sense that $\varpi_i = O\left(N^{-1}\right)$. Using (40) and (41) we have

$$\bar{x}_{N,t} = \frac{x_{1t} + \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\left(\rho\mathbf{w}_{21}x_{1t} - \mathbf{b}_2 - \alpha\boldsymbol{\varepsilon}_{2t} - \alpha\boldsymbol{\gamma}_2 f_t\right)}{N},$$

[5] The exponent formulation of column sums in (39) will be compared and contrasted to the power law specification favored in the literature in Section 7 below.

[6] In deriving (41), it is required that $\zeta_1 \neq 0$. This condition is met since the $N \times N$ matrix on the right-hand-side of (37) is non-singular.

where $\mathbf{b}_2 = (b_2, b_3, ..., b_N)'$ and $\boldsymbol{\gamma}_2 = (\gamma_2, \gamma_3, ..., \gamma_N)'$. Hence

$$\overline{x}_{N,t} = N^{-1}\left(-a_n + \theta_n x_{1t} - \alpha\psi_n f_t - \alpha\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right), \tag{43}$$

where $a_n = \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{b}_2$, $\boldsymbol{\phi}_n' = \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}$, and

$$\theta_n = 1 + \rho\boldsymbol{\phi}_n'\mathbf{w}_{21}, \tag{44}$$

$$\psi_n = \boldsymbol{\phi}_n'\boldsymbol{\gamma}_2. \tag{45}$$

The first term of (43), $N^{-1}a_n$, is bounded in N, since $\|\mathbf{W}_{22}\|_\infty \leq 1$ and $\rho\|\mathbf{W}_{22}\|_1 < 1$, and as a result $\mathbf{S}_{22}^{-1}$ will have bounded row and column norms by Lemma 2 in Appendix A. The second term captures the effect of the dominant unit. The third term is due to the common factor, f_t, and the final term represents the average effects of the micro productivity shocks. $N^{-1}\boldsymbol{\phi}_n$ is the influence vector associated with the non-dominant units. It is analogous to the influence vector defined by (8) which applies to all units.

Starting with the final term of (43), we first note that

$$\underline{\sigma}^2 N^{-2}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n \leq Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \leq \bar{\sigma}^2 N^{-2}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n, \tag{46}$$

where $\boldsymbol{\phi}_n' = (\phi_2, \phi_3, ..., \phi_N)$ is an $n \times 1$ vector of column sums of $\mathbf{S}_{22}^{-1}$ and has bounded elements. Furthermore, since

$$\boldsymbol{\phi}_n' = \boldsymbol{\tau}_n' + \rho\boldsymbol{\tau}_n'\mathbf{W}_{22} + \rho^2\boldsymbol{\tau}_n'\mathbf{W}_{22}^2 + ...,$$

$\rho > 0$ and $w_{ij} \geq 0$, then $\phi_{\min} = \min(\phi_2, \phi_3, ..., \phi_N) > 1$, and $\phi_{\max} = \max(\phi_2, \phi_3, ..., \phi_N) < K < \infty$. Hence,

$$1 < \phi_{\min}^2 \leq N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n \leq \phi_{\max}^2 < K < \infty,$$

and $N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n$ is bounded from below and above by finite non-zero constants. Using this result in (46) we also have

$$\underline{\sigma}^2 < NVar\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) < \bar{\sigma}^2\phi_{\max}^2 < \infty,$$

which establishes that

$$Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = \ominus\left(N^{-1}\right), \tag{47}$$

where $\ominus\left(N^{-1}\right)$ denotes the convergence rate of $Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right)$ in terms of N, and should be distinguished from the $O\left(N^{-1}\right)$ notation, which provides only an upper bound on $Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right)$.

Next, using (41) we have

$$Cov\left(x_{1t}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = -\alpha\rho\zeta_1^{-1}N^{-1}\mathbf{w}_{12}'\mathbf{H}_{22}\boldsymbol{\tau}_n, \tag{48}$$

and

$$Cov\left(x_{1t}, f_t\right) = -\alpha\left(\zeta_1^{-1}\gamma_1 + \rho\zeta_1^{-1}h_2\right)Var\left(f_t\right), \tag{49}$$

where $\mathbf{H}_{22} = \mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}'^{-1}$, $\mathbf{V}_{22,\varepsilon} = diag\left(\sigma_2^2, \sigma_3^2, ..., \sigma_N^2\right)$, and $h_2 = \mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2$. It then follows that overall (recalling that f_t and ε_{it} are independently distributed), we have

$$\begin{aligned} Var\left(\overline{x}_{N,t}\right) &= N^{-2}\theta_n^2 Var\left(x_{1t}\right) - 2\alpha N^{-2}\theta_n Cov\left(x_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) + \alpha^2 N^{-2}Var\left(\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \\ &\quad + \alpha^2 N^{-2}\chi_n Var\left(f_t\right), \end{aligned} \tag{50}$$

where

$$\chi_n = \psi_n^2 + 2\psi_n \theta_n \zeta_1^{-1} \gamma_1 + 2\rho \psi_n \theta_n \zeta_1^{-1} h_2.$$

Also, using (41) we have

$$Var\left(x_{1t}\right) = \zeta_1^{-2} \alpha^2 \left[\left(\gamma_1^2 + \rho^2 h_2^2 \right) Var(f_t) + \sigma_1^2 \right] + \zeta_1^{-2} \rho^2 \alpha^2 \mathbf{w}_{12}' \mathbf{H}_{22} \mathbf{w}_{21}, \tag{51}$$

which is easily seen to be bounded in N.

A number of results can now be obtained from (50). First, without a common factor and a dominant unit, $Var\left(\overline{x}_{N,t}\right) = \ominus(N^{-1})$, and the effects of idiosyncratic shocks on $\overline{x}_{N,t}$ will vanish at the rate of $N^{-1/2}$, as $N \to \infty$. This rate matches the decay rate of shocks in models without a network structure, namely even if we set $\mathbf{W} = \mathbf{0}$. Therefore, for micro shocks to have macroeconomic implications there must be at least one dominant unit in the network. To see this consider now the case where there is no common factor but the network includes a dominant unit. Then using (47) and (50) we have

$$Var\left(\overline{x}_{N,t}\right) = N^{-2} \theta_n^2 Var\left(x_{1t}\right) - 2\alpha N^{-2} \theta_n Cov\left(x_{1t}, \boldsymbol{\phi}_n' \boldsymbol{\varepsilon}_{2t}\right) + O\left(N^{-1}\right). \tag{52}$$

Recall that $Var(x_{1t})$ is bounded in N, and consider the limiting properties of $N^{-1}\theta_n$ defined by (44). Since

$$N^{-1} + \phi_{\min} \rho N^{-1} d_1 \leq N^{-1} \theta_n \leq N^{-1} + \phi_{\max} \rho N^{-1} d_1, \tag{53}$$

where $1 \leq \phi_{\min} \leq \phi_{\max} < K$, then the asymptotic behavior of $N^{-1}\theta_n$ depends on the way the outdegree of the dominant unit, namely d_1, varies with N. Using the exponent specification given by (26), $d_1 = \kappa_1 N^{\delta_1}$, it follows that

$$N^{-1} + \phi_{\min} \rho \kappa_1 N^{\delta_1 - 1} \leq N^{-1} \theta_n \leq N^{-1} + \phi_{\max} \rho \kappa_1 N^{\delta_1 - 1}, \tag{54}$$

which leads to

$$N^{-1} \theta_n = \ominus(N^{\delta_1 - 1}), \ 0 < \delta_1 \leq 1. \tag{55}$$

Consider now the second term of (52), and note from (48) that

$$\left| Cov\left(x_{1t}, \mathbf{v}_n' \boldsymbol{\varepsilon}_{2t}\right) \right| \leq \left| \frac{(1-\rho)\rho}{\zeta_1} \right| N^{-1} \left\| \mathbf{w}_{12}' \right\|_\infty \left\| \mathbf{S}_{22}^{-1} \right\|_\infty \left\| \mathbf{V}_{22,\varepsilon} \right\|_\infty \left\| \boldsymbol{\phi}_n \right\|_\infty = O\left(N^{-1}\right),$$

since $\left\| \mathbf{w}_{12}' \right\|_\infty = \left\| \mathbf{w}_{12} \right\|_1 = \sum_{i=2}^N w_{1i} = 1$, $\left\| \mathbf{S}_{22}^{-1} \right\|_\infty < K$, $\left\| \mathbf{V}_{22,\varepsilon} \right\|_\infty = \bar{\sigma}^2 < K$, and $\left\| \boldsymbol{\phi}_n \right\|_\infty = \phi_{\max} < K$. Using the above results in (52) we have

$$Var\left(\overline{x}_{N,t}\right) = \ominus(N^{2\delta_1 - 2}) + O(N^{\delta_1 - 2}) + O\left(N^{-1}\right),$$

which simplifies to (since $\delta_1 \leq 1$)

$$Var\left(\overline{x}_{N,t}\right) = \ominus(N^{2\delta_1 - 2}) + O\left(N^{-1}\right), \tag{56}$$

and hence

$$Var\left(\overline{x}_{N,t}\right) = \ominus(N^{2\delta_1 - 2}), \text{ if } \delta_1 > 1/2. \tag{57}$$

This is the main result for the analysis of macro economic implications of micro shocks, and is more general than the one established by Acemoglu et al. (2012) who only provide a lower bound on the rate at which aggregate volatility changes with N.

It is also instructive to relate $N^{-1}\theta_n$ to the first- and higher-order network connections discussed in Acemoglu et al. (2012). Expanding the terms of the inverse $\mathbf{S}_{22}^{-1}$, $N^{-1}\theta_n$ can also be written as

$$N^{-1}\theta_n = N^{-1}\left(1 + \rho\boldsymbol{\tau}_n'\mathbf{w}_{21} + \rho^2\boldsymbol{\tau}_n'\mathbf{W}_{22}\mathbf{w}_{21} + \rho^3\boldsymbol{\tau}_n'\mathbf{W}_{22}^2\mathbf{w}_{21} + ...\right),$$

where $N^{-1}\rho\boldsymbol{\tau}_n'\mathbf{w}_{21} = \rho N^{-1}d_1$ represents the effects of the first-order network connections on θ_N, $N^{-1}\rho^2\boldsymbol{\tau}_n'\mathbf{W}_{22}\mathbf{w}_{21}$, the effects of the second-order network connections and so on. But in view of (53) and (55) all these higher order interconnections (individually and together) at most behave as $\ominus(N^{\delta_1-1})$.

Therefore, the rate at which micro shocks influence the macro economy depends on δ_1, which measures the strength of the dominant unit. But it should be noted from (56) that to ensure a non-vanishing variance, $Var\left(\overline{x}_{N,t}\right) > 0$, as $N \to \infty$, we need a value of $\delta_1 = 1$. When $1/2 < \delta_1 < 1$, the network accentuates the diffusion of the idiosyncratic shocks across the network but does not lead to lasting impacts. No network effects of unit-specific shocks can be identified when $\delta_1 \leq 1/2$. Hence, for the dominant unit to have any impact over and above the standard rates of diversification of micro shocks on $\overline{x}_{N,t}$, we need $\delta_1 > 1/2$.

Remark 3. *The finding that δ_1 cannot be distinguished from zero if $\delta_1 < 1/2$ is also related to the study by Bailey et al. (2016), who show that the exponent of cross-sectional dependence, α, can only be identified and consistently estimated for values of $\alpha > 1/2$.*

Suppose now that the network is subject to common shocks without a dominant unit. In this case we have

$$Var\left(\overline{x}_{N,t}\right) = \alpha^2 N^{-2}\psi_n^2 Var\left(f_t\right) + \ominus\left(N^{-1}\right),$$

and the rate of convergence of $\overline{x}_{N,t}$ is determined by the strength of the factor as given by $N^{-2}\psi_n^2$. Using (45) we have (recall that $\varrho\left(\mathbf{W}_{22}\right) \leq 1$),

$$\begin{aligned} N^{-1}\psi_n &= N^{-1}\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2 \\ &= N^{-1}\left(\boldsymbol{\tau}_n'\boldsymbol{\gamma}_2 + \rho\boldsymbol{\tau}_n'\mathbf{W}_{22}\boldsymbol{\gamma}_2 + \rho^2\boldsymbol{\tau}_n'\mathbf{W}_{22}^2\boldsymbol{\gamma}_2 + ...\right). \end{aligned}$$

By a similar line of reasoning as before, it is then easily seen that $N^{-1}\psi_n = \ominus\left(N^{\delta_\gamma - 1}\right)$, where δ_γ (already defined by (2)) is the cross-section exponent of the factor loadings, γ_i, and measures the degree to which the common factor is pervasive in its effects on sector-specific productivity.

Finally, suppose that the production network is subject to a common factor as well as containing a dominant unit. Then for $\delta_1 > 1/2$ and $\delta_\gamma > 1/2$ we have[7]

$$Var\left(\overline{x}_{N,t}\right) = \ominus\left(N^{2\delta_1-2}\right) + \ominus\left(N^{2\delta_\gamma-2}\right) + \ominus\left(N^{-1}\right). \tag{58}$$

[7] Note that for values of δ_1 and $\delta_\gamma < 1/2$ the third term in (58) will dominate the network and the common factor effects.

It is clear that the relative importance of the dominant unit and the common factor depends on the relative magnitudes of δ_1 and δ_γ. We need estimates of these exponents for a further understanding of the relative importance of macro and micro shocks in business cycle analysis.

Allowing for multiple factors and multiple dominant units does not alter the main results, and the general expression in (58) will continue to apply, with the difference that δ_γ and δ_1 in such a general setting will refer to the maximum of the exponents of the factors and $\delta_{\max} = \max(\delta_1, \delta_2, ..., \delta_N)$.

6 $\delta_{\max}$ and β measures of network pervasiveness

The degree of network pervasiveness, $\delta_{\max}$, defined in Definition 2, is related to β, the shape parameter of the power law assumed by Acemoglu et al. (2012, Definition 2) for the outdegree sequence, $\{d_1, d_2, ..., d_N\}$. To see this we first use the specification of the outdegrees given by (26) in (9) to obtain

$$\boldsymbol{v}_N'\boldsymbol{v}_N \geq k_{\alpha,0}N^{-1} + k_{\alpha,1}N^{-2}\sum_{j=1}^{m}\kappa_j^2 N^{2\delta_j} + k_{\alpha,1}\frac{N-m}{N^2}\left(\sum_{j=m+1}^{N}\kappa_j^2/\left(N-m\right)\right),$$

where $(N-m)^{-1}\sum_{j=m+1}^{N}\kappa_j^2 = O(1)$. Also, recall that m must be finite if $\{\delta_j\}$ is to be summable (Proposition 2), and therefore

$$N^{-2}\sum_{j=1}^{m}\kappa_j^2 N^{2\delta_j} \leq m\kappa_{\max}^2 N^{2(\delta_{\max}-1)},$$

where as before $\kappa_{\max} = \max\left(\kappa_1, \kappa_2, ..., \kappa_N\right)$. Then the limiting behavior of $\boldsymbol{v}_N'\boldsymbol{v}_N$ will be determined by $N^{2(\delta_{\max}-1)}$, namely the cross section exponent of the strongest of the dominant units, $\delta_{\max}$. For the production network to affect macro economic fluctuations we need $\delta_{\max} > 1/2$. But to make sure that $\boldsymbol{v}_N'\boldsymbol{v}_N$ does not converge to zero as $N \to \infty$, as noted earlier a much stronger condition, namely $\delta_{\max} = 1$, is required. It is clear that the same result follows if $\mathbf{W}$ has one strongly dominant unit with $\delta = 1$. The remaining dominant units either behave similarly, or will be dominated by the leading dominant unit, with $\delta_{\max}$.

Consider now Corollary 1 of Acemoglu et al. (2012), where it is established that aggregate volatility behaves asymptotically as $N^{-2(\beta-1)/\beta-2\epsilon_\beta}$, for some small $\epsilon_\beta > 0$ and $\beta \in (1,2)$. Matching this rate of expansion with $N^{2(\delta_{\max}-1)}$, we have

$$2\left(\delta_{\max}-1\right) \geq -2(\beta-1)/\beta - 2\epsilon_\beta,$$

or $\delta_{\max} \geq 1/\beta - \epsilon_\beta$. Therefore, $\delta_{\max}$ can be viewed as measuring the inverse of β, a result that we formally establish below.

It is also easily seen that $\delta_{\max} \geq 1/\varsigma - \epsilon_\varsigma$, where $\epsilon_\varsigma > 0$, and ς is the shape parameter of the power law assumed by Acemoglu et al. (2012, Corollary 2) for the second-order degree sequences (based on the second-order interactions in the network). Other shape parameters are also considered by Acemoglu et al. (2012) for higher-order degree sequences. But clearly these shape parameters are closely related, and are not needed since as shown above it is possible to derive an exact rate for the asymptotic behavior of $\boldsymbol{v}_N'\boldsymbol{v}_N$.

7 Estimation and inference

In this section we consider the problem of estimating the degree of dominance of units in a given network. We consider the power law approach employed in the literature as well as a new method that we propose when the outdegrees, $\{d_1, d_2, ..., d_N\}$, follow the exponent specification defined by (39). It is unclear if a power law specification for the outdegrees (above a given cut-off value) is necessarily to be preferred to a specification which relates the outdegrees directly to the size of the network, N, without the need to specify a cut-off value. The exponent specification of outdegrees has the added advantage that it also allows identification of more than one dominant units in the network.

7.1 Power law estimators

Suppose that we have observations on the outdegrees, d_{it}, for $i = 1, 2, ..., N$; and $t = 1, 2, ...T$. The power law estimate of $\delta_{\max}$ is given by $1/\hat{\beta}$, where $\hat{\beta}$ is an estimator of the shape parameter of the power law distribution fitted to the outdegrees that lie above a given minimum cut-off value, $d_{\min}$. (See Section 6). A random variable D is said to follow a power law distribution if its complementary cumulative density function (CCDF) satisfies

$$\Pr(D \geq d) \propto d^{-\beta}, \tag{59}$$

where $\beta > 0$ is a constant known as the shape parameter of the power law, and $\propto$ denotes asymptotic equivalence.[8] As the name suggests, the tail of the power law distribution decays asymptotically at the power of β. It is readily seen that the probability density function of D follows $f_D(d) \propto d^{-(\beta+1)}$.

A popular specific case of power laws is the Pareto distribution. Its CCDF is given by

$$\Pr(D \geq d) = (d/d_{\min})^{-\beta}, \; d \geq d_{\min}, \tag{60}$$

for some shape parameter $\beta > 0$ and lower bound $d_{\min} > 0$. The Pareto distribution has been widely used to study the heavy-tailed phenomena in many fields including economics, finance, geology, physics, just to name a few. Since our focus is on the estimation of the shape parameter β, in what follows we briefly describe three approaches that are frequently used in the literature.[9]

The first is to run the following log-log regression (also known as Zipf regression),

$$\ln i = a - \beta \ln d_{(i)}, \; i = 1, 2, ..., N_{\min}, \tag{61}$$

where a is a constant, i is the rank of the unit i in the sequence $\{d_{(i)}\}$, and $d_{\max} = d_{(1)} \geq d_{(2)} \geq ... \geq d_{(N_{\min})}$, are the largest ordered outdegrees such that $d_{(N_{\min})} \geq d_{\min}$, and $N_{\min}$ is the number of cut-off observations used in the regression. A bias-corrected version of the log-log estimator of β, is proposed by Gabaix and Ibragimov (2011) who suggest shifting

[8] More generally, power law distributions take the form $\Pr(D \geq d) \propto L(d)d^{-\beta}$, where $L(d)$ is some slowly varying function, which satisfies $\lim_{d\to\infty} L(rd)/L(d) = 1$, for any $r > 0$.

[9] More sophisticated techniques are reviewed in, among others, Beirlant et al. (2006), and are beyond the scope of this paper.

the rank i by 1/2 and estimating β by Ordinary Least Squares (OLS) using the following regression

$$\ln(i - 1/2) = a - \beta \ln d_{(i)},\ i = 1, 2, ..., N_{\min}. \tag{62}$$

In what follows we consider this log-log estimator and refer to it as the Gabaix-Ibragimov (GI) estimator, which we denote by $\hat{\beta}_{GI}$. The standard error of $\hat{\beta}_{GI}$ is estimated by $\hat{\sigma}\left(\hat{\beta}_{GI}\right) = \sqrt{2/N_{\min}}\hat{\beta}_{GI}$.

Another often-used estimator of β is the maximum likelihood estimator (MLE), denoted by $\hat{\beta}_{MLE}$, which is also the well-known Hill estimator (Hill, 1975). It can be easily verified that[10]

$$\hat{\beta}_{MLE} = \frac{N_{\min}}{\sum_{i=1}^{N_{\min}} \ln d_{(i)} - N_{\min} \ln d_{(N_{\min})}}, \tag{63}$$

and its standard error is given by $\hat{\sigma}\left(\hat{\beta}_{MLE}\right) = \hat{\beta}_{MLE}/\sqrt{N_{\min}}$. The ML estimator is most efficient if $d_{\min}$ is known and the underlying distribution above the cut-off point is Pareto.

Finally, some researchers, notably Clauset, Shalzi and Newman (2009, CSN), propose joint estimation of β and $d_{\min}$ and recommend estimating $d_{\min}$ by minimizing the Kolmogorov-Smirnov or KS statistics, which is the maximum distance between the empirical cumulative distribution function (CDF) of the sample, $S(d)$, and the CDF of the reference distribution, $F(d)$, namely,

$$\mathcal{T}_{KS} = \max_{d \geq d_{\min}} |S(d) - F(d)|.$$

Here $F(d)$ is the CDF of the Pareto distribution that best fits the data for $d \geq d_{\min}$. The MLE in (63) is then computed using the estimated value of $d_{\min}$. Hereafter, we refer to this estimator as the feasible maximum likelihood estimator and denote it by $\hat{\beta}_{CSN}$.[11]

In the subsequent analysis, we examine how the inverse of β, which is estimated by the three procedures discussed above, behave as an estimator of $\delta_{\max}$, and how these estimators compare to the extremum estimators that we now consider.

7.2 Extremum estimators

We now propose an extremum estimator of $\delta_{\max}$ which could also be used to estimate β, and will be shown to be more generally applicable as compared to the power law estimators of β. Our estimator is motivated by the exponent specification of outdegrees given by (26).

7.2.1 Pure cross sections

In line with the literature on estimation of β, we begin with the case where only a single set of observations on the outdegrees, $\{d_i\}$, is available, but instead of the power law

[10] See, for example, Appendix B of Newman (2005).

[11] The code implementing this method can be downloaded from http://tuvalu.santafe.edu/~aaronc/powerlaws/.

specification we assume that d_i, $i = 1, 2, ..., N$, are generated according to the following exponent specification:

$$d_i = \kappa N^{\delta_i} \exp(\upsilon_i), \ i = 1, 2, ..., N, \tag{64}$$

where $\kappa > 0$, $\upsilon_i \sim IID(0, \sigma_\upsilon^2)$, and υ_i is distributed independently of δ_i. Furthermore, given the constraint (27), we must have

$$\kappa \sum_{i=1}^{N} N^{\delta_i} \exp(\upsilon_i) = N. \tag{65}$$

Taking the log transformation of (64) we have

$$\ln d_i = \ln \kappa + \delta_i \ln N + \upsilon_i, \ i = 1, 2, ..., N.$$

Averaging across i now yields

$$N^{-1} \sum_{i=1}^{N} \ln d_i = \ln \kappa + \left(N^{-1} \sum_{i=1}^{N} \delta_i \right) \ln N + N^{-1} \sum_{i=1}^{N} \upsilon_i. \tag{66}$$

But from the summability condition, (35), it follows that

$$\left(N^{-1} \sum_{i=1}^{N} \delta_i \right) \ln N \leq K \left(\frac{\ln N}{N} \right),$$

and hence

$$\left(N^{-1} \sum_{i=1}^{N} \delta_i \right) \ln N \to 0, \text{ as } N \to \infty. \tag{67}$$

Considering that by assumption υ_i are IID over i, then the last term of (66) also tends to zero, and $\ln \kappa$ can be estimated by

$$\widehat{\ln \kappa} = N^{-1} \sum_{i=1}^{N} \ln d_i. \tag{68}$$

An extremum estimator of $\delta_{\max}$ now emerges as

$$\hat{\delta}_{\max} = \frac{\ln d_{\max} - N^{-1} \sum_{i=1}^{N} \ln d_i}{\ln N} = \frac{N^{-1} \sum_{i=1}^{N} \ln (d_{\max}/d_i)}{\ln N}, \tag{69}$$

where $d_{\max}$ is the largest value of $d_i > 0$. As we shall see $\hat{\delta}_{\max}$ is a consistent estimator of the degree of pervasiveness of the most dominant unit in the network under fairly general specifications of the outdegrees, including the exponent specification given by (64).

The exponent specification has the advantage that it is closely related to (26) in that $\kappa_j = \kappa \exp(\upsilon_i) > 0$, and is in line with the production network model derived from a set of underlying economic relations. Nonetheless, in practice it is difficult to know if the true data generating process follows the exponent or a power law specification. But it turns out that $1/\hat{\delta}_{\max}$ is a consistent estimator of β, the shape parameter of the Pareto distribution, even under the Pareto distribution.

To see this, suppose that the observations on the outdegrees, d_i, for $i = 1, 2, ..., N$, are independent draws from the following mixed-Pareto distribution

$$\begin{aligned} f(d_i) &\propto d_i^{-1-\beta}, \text{ if } d_i \geq d_{\min}, \\ &\propto \psi(d_i), \text{ if } d_i < d_{\min}, \end{aligned} \tag{70}$$

where $d_i > 0$ follows a Pareto distribution with the shape parameter β for values of d_i above $d_{\min}$, and an arbitrary non-Pareto distribution, $\psi(d_{it})$, for values of d_i below $d_{\min}$. The constants of the proportionality for both branches of the distribution are set to ensure that $\int_0^\infty f(x)dx = 1$, and that a given non-zero proportion of the observations fall above $d_{\min}$.

Using (69), the extremum estimator, $\hat{\delta}_{\max}$, can be rewritten as

$$\hat{\delta}_{\max} = \frac{z_{\max} - N^{-1} \sum_{i=1}^{N} z_i}{\ln N}, \tag{71}$$

where $z_i = \ln(d_i/d_{\min})$, for all i, and $z_{(i)} = \ln(d_{(i)}/d_{\min})$, with $d_{(i)}$ being the i^{th} largest value of d_i as before. Since $d_{\min}$ is a given constant and by assumption d_i are independently distributed, it then follows that for $z_i \geq 0$, z_i are independent draws from an exponential distribution with parameter β, namely

$$f_Z(z) = \beta e^{-\beta z}, \text{ for } z \geq 0,$$

with $E\left(z \,|z \geq 0\right) = 1/\beta$, and $Var\left(z \,|z \geq 0\right) = 1/\beta^2$, for $\beta > 0$.[12] We also assume that $E\left(z_i \,|z_i < 0\right)$ exists, which is a mild moment condition imposed on $\psi(d_i)$ for $\ln\left(d_i/d_{\min}\right) < 0$. The following proposition summarizes the consistency property of $\hat{\delta}_{\max}$ as an estimator of $1/\beta$.

Proposition 3. *Suppose that d_i, for $i = 1, 2, ..., N$, are independent draws from the Pareto tail distribution given by (70) with the shape parameter $\beta > 0$, and assume that $z_i = \ln(d_i/d_{\min})$ has finite second order moments for all values of $z_i < 0$. It then follows that*

$$\lim_{N\to\infty} E\left(\hat{\delta}_{\max}\right) = 1/\beta, \text{ and } Var\left(\hat{\delta}_{\max}\right) = O\left[\frac{1}{(\ln N)^2}\right], \tag{72}$$

where $\hat{\delta}_{\max}$ is defined by (71).

A proof is provided in Appendix C.

Remark 4. *The convergence of $\hat{\delta}_{\max}$ to $\delta = 1/\beta$, is at the rate of $1/\ln N$ which is rather slow. But it is obtained without making any assumptions about $d_{\min}$ and/or the shape of $\psi(d)$, the non-Pareto part of the distribution.*

[12] It is worth noting that z has moments even if $\beta \leq 1$, although the Pareto distribution has moments only for $\beta > 1$.

7.2.2 Short T panels

Suppose now that observations on outdegrees, d_{it}, are available across $t = 1, 2, ..., T$ and are generated as before, namely

$$d_{it} = \kappa N^{\delta_i} \exp(\upsilon_{it}),\ i = 1, 2, ..., N;\ t = 1, 2, ..., T, \tag{73}$$

where T is finite ($T > 1$) and N is large, $\kappa > 0$, and $\delta_i \geq 0$ are fixed constants. $\upsilon_{it} \sim IID(0, \sigma_\upsilon^2)$ over i and t.[13] Noting that (65) must hold and using a similar line of reasoning as in the pure cross section case we obtain the following "panel extremum estimator"

$$\hat{\delta}_{\max} = \frac{T^{-1}\sum_{t=1}^{T} \ln d_{\max,t} - (TN)^{-1} \sum_{t=1}^{T}\sum_{j=1}^{N} \ln d_{jt}}{\ln N}, \tag{74}$$

where $d_{\max,t}$ is the largest value of d_{it} for period t. Also for other units we have

$$\hat{\delta}_i = \frac{T^{-1}\sum_{t=1}^{T} \ln d_{it} - (TN)^{-1} \sum_{t=1}^{T}\sum_{j=1}^{N} \ln d_{jt}}{\ln N}, \tag{75}$$

where $\hat{\delta}_i$ is the estimator of δ_i. Recall that only estimates of $\hat{\delta}_i > 1/2$ should be considered, since any unit with an estimate of δ_i below $1/2$ will not have any network effects (see Remark 3). We denote the degree of dominance of unit i by δ_i, and the associated ordered values by $\delta_{(i)}$, where $\delta_{\max} = \delta_{(1)} \geq \delta_{(2)} \geq ... \geq \delta_{(N)}$. Then the second largest $\hat{\delta}_i$, denoted by $\hat{\delta}_{(2)}$, is a consistent estimator of $\delta_{(2)}$; the third largest $\hat{\delta}_i$, denoted by $\hat{\delta}_{(3)}$, is a consistent estimator of $\delta_{(3)}$; and so on. We refer to $\hat{\delta}_{(i)}$, for $i = 1, 2, ..., m$, as the extrema estimators, with $\hat{\delta}_{(m)} > 1/2$.

To investigate the asymptotic properties of the extrema estimators, we first note that under (73), $\widehat{\ln \kappa}$ and $\hat{\delta}_i$ can be written as

$$\widehat{\ln \kappa} - \ln \kappa = \bar{\delta} \ln N + \bar{\upsilon}, \tag{76}$$

$$\hat{\delta}_i - \delta_i = \bar{\delta} + \frac{\bar{\upsilon}_i - \bar{\upsilon}}{\ln N}, \tag{77}$$

where $\widehat{\ln \kappa} = (TN)^{-1} \sum_{t=1}^{T}\sum_{i=1}^{N} \ln d_{it}$, $\bar{\delta} = N^{-1}\sum_{i=1}^{N} \delta_i$, $\bar{\upsilon}_i = T^{-1}\sum_{t=1}^{T} \upsilon_{it}$, and $\bar{\upsilon} = N^{-1}\sum_{i=1}^{N} \bar{\upsilon}_i$. It is now easily seen that

$$\begin{aligned} Cov(\hat{\delta}_i, \hat{\delta}_j) &= -\frac{1}{(\ln N)^2 N}\frac{\sigma_\upsilon^2}{T}, \text{ for all } i \neq j, \\ Var\left(\hat{\delta}_i\right) &= \frac{\sigma_\upsilon^2}{(\ln N)^2 T}\left(1 - \frac{1}{N}\right). \end{aligned} \tag{78}$$

Clearly for any given i, $Var\left(\hat{\delta}_i\right) \to 0$, for a fixed T as $N \to \infty$, and the rate of convergence of $Var\left(\hat{\delta}_i\right)$ is given by $(\ln N)^{-2}$. Also, it is already established that $\bar{\delta} \to 0$, as $N \to \infty$,

[13] This assumption can be relaxed to allow for both heteroskedasticity and serial correlation if T is sufficiently large. This extension is considered in an Online Supplement.

and as a result $\hat{\delta}_i \rightarrow_p \delta_i$, which establishes that for any finite $T \geq 1$, $\hat{\delta}_i$ is a $\ln N$ consistent estimator of δ_i. The estimators of δ for two different units (i, j), are asymptotically independent and their covariance converges to zero at the faster rate of $(\ln N)^{-2} N^{-1}$.

Then, for any given i we have (applying standard central limit theorem to $\bar{\upsilon}_i - \bar{\upsilon}$)

$$\frac{\left(\hat{\delta}_i - \delta_i - \bar{\delta}\right)}{\left[\frac{\sigma_{\upsilon}^2}{(\ln N)^2 T}\left(1 - \frac{1}{N}\right)\right]^{1/2}} \rightarrow_d N(0,1), \text{ as } N \rightarrow \infty.$$

Ignoring lower order terms in N we now obtain

$$\frac{(\ln N)\sqrt{T}\left(\hat{\delta}_i - \delta_i - \bar{\delta}\right)}{\sigma_{\upsilon}} \rightarrow_d N(0,1), \text{ as } N \rightarrow \infty.$$

To ensure that the above statistic does not depend on the nuisance parameter, $\bar{\delta}$, we need

$$\bar{\delta}(\ln N)\sqrt{T} = \left(\sum_{i=1}^{N} \delta_i\right)\frac{(\ln N)\sqrt{T}}{N} \rightarrow 0, \text{ as } N \rightarrow \infty. \tag{79}$$

But given the summability condition, (35), it is clear that this condition is met if T is fixed, as $N \rightarrow \infty$.

To obtain a consistent estimator of σ_{υ}^2, note that

$$\begin{aligned} \hat{\upsilon}_{it} &= \ln d_{it} - \widehat{\ln \kappa} - \hat{\delta}_i \ln N \\ &= -\left(\widehat{\ln \kappa} - \ln \kappa\right) - \left(\hat{\delta}_i - \delta_i\right)\ln N + \upsilon_{it}. \end{aligned} \tag{80}$$

Now using (76) and (77)

$$\begin{aligned} \hat{\upsilon}_{it} &= -2\bar{\delta}\ln N + \upsilon_{it} - \bar{\upsilon}_i, \\ &= \upsilon_{it} - \bar{\upsilon}_i + O\left(\frac{\ln N}{N}\right). \end{aligned}$$

In view of this result, σ_{υ}^2 can be consistently estimated by (for $T > 1$)

$$\hat{\sigma}_{\upsilon}^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\upsilon}_{it}^2}{N(T-1)}. \tag{81}$$

A test of the null hypothesis that $\delta_{\max} = \delta_{\max}^0$, where $\delta_{\max}^0 > 1/2$, can be based on the statistic

$$\mathfrak{D}_{\max} = \frac{(\ln N)\left(\hat{\delta}_{\max} - \delta_{\max}^0\right)}{\hat{\sigma}_{\upsilon}\left(\frac{1}{T} - \frac{1}{TN}\right)^{1/2}}. \tag{82}$$

It then follows that $\mathfrak{D}_{\max} \rightarrow_d N(0,1)$, as $N \rightarrow \infty$, for a fixed $T > 1$.

7.2.3 Unbalanced panels

In empirical applications, production networks observed at different points in time might not have the same units in common. As a result we are often faced with unbalanced panel data sets. One approach would be to employ a sufficiently high level of aggregation so that we end up with a balanced panel. But this procedure is likely to be inefficient as we end up with a smaller number of units in the network. Here we consider estimating δ_i with the unbalanced panel, without any aggregation. We suppose for each unit i we have observations on its outdegrees for at least two time periods.

We denote the unbalanced panel of observations on the outdegrees by d_{it}, for $t = T_i^0, T_i^0+1, ..., T_i^1$, $(T_i^1 \geq T_i^0)$, and $i = 1, ..., N$. Then using a similar line of reasoning as above we have

$$\hat{\delta}_i = \frac{T_i^{-1} \sum_{t=T_i^0}^{T_i^1} \ln d_{it} - N^{-1} \sum_{i=1}^{N} \left(T_i^{-1} \sum_{t=T_i^0}^{T_i^1} \ln d_{it} \right)}{\ln N}, \tag{83}$$

where $T_i = T_i^1 - T_i^0 + 1$, and

$$Var\left(\hat{\delta}_i\right) = \frac{\sigma_v^2}{(\ln N)^2} \left(\frac{1}{T_i} - \frac{1}{NT_i} \right). \tag{84}$$

The estimator of the most dominant unit is given by $\hat{\delta}_{\max}$, and as in the balanced panel case, the asymptotic distribution of $\hat{\delta}_{\max}$ is given by

$$\mathfrak{D}_{\max} = \frac{(\ln N)\left(\hat{\delta}_{\max} - \delta_{\max}^0\right)}{\hat{\sigma}_v \left(\frac{1}{T_{\max}} - \frac{1}{NT_{\max}} \right)^{1/2}}, \tag{85}$$

where $T_{\max}$ refers to the sample size of the most dominant unit, and

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N} (T_i - 1)^{-1} \sum_{t=T_i^0}^{T_i^1} \hat{v}_{it}^2}{N}. \tag{86}$$

The distribution of the most dominant unit is well defined if $T_{\max} > 1$.

7.2.4 Comparison of power law and extremum estimators

As compared to the power law estimators, the extremum estimator has several advantages. First, it does not require knowing the true value of $d_{\min}$, whereas the estimates of the shape parameter may be highly sensitive to the choice of the cut-off value. Although procedures such as the feasible MLE proposed by Clauset et al. (2009) estimate $d_{\min}$ jointly with β, such estimates assume that the true distribution below and above $d_{\min}$ are known, whilst the extremum estimator is robust to any distributional assumptions below $d_{\min}$, so long as $\ln(d_i/d_{\min})$ has second order moments. Granted that it may not be as efficient as MLE if the true distribution is indeed Pareto, one does not need to make such strong assumptions on the entire distribution. Third, the extremum type estimators can identify the dominant units besides the most dominant one, and estimate the degrees of pervasiveness of each of the of the dominant units separately.

8 Monte Carlo experiments

In this section, we investigate the small sample properties of the proposed extremum estimator for balanced panels using Monte Carlo techniques, and compare its performance with that of the power law method.[14]

We consider two types of data generating processes (DGPs) for the outdegrees (d_{it}): an exponent specification and a power law specification. The DGP for the exponent specification is given by

$$\ln d_{it} = \ln \kappa + \delta_i \ln N + \upsilon_{it}, \; i = 1, 2, ..., N; \; t = 1, 2, ..., T, \tag{87}$$

where $\upsilon_{it} \sim IIDN(0, 1)$, with κ set as

$$\kappa = \frac{exp\left(-\frac{1}{2}\right)}{N^{-1}\sum_{i=1}^{N} N^{\delta_i}} > 0, \tag{88}$$

to ensure that d_{it} add up to N across i for each t.

For the power law model we closely follow Clauset et al. (2009), and initially generate y_{it} as random draws from the following mixture distribution that obeys an exact Pareto distribution above $y_{\min,t}$ and an exponential distribution below $y_{\min,t}$:

$$f(y_{it}) = \begin{cases} C_t (y_{it}/y_{\min,t})^{-(\beta+1)}, & \text{for } y_{it} \geq y_{\min,t} \\ \\ C_t e^{-(\beta+1)(y_{it}/y_{\min,t}-1)}, & \text{for } y_{it} < y_{\min,t} \end{cases}, \tag{89}$$

for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$. To ensure that $f(y_{it})$ integrates to 1 over its full support, $y_{it} > 0$, we set C_t as

$$C_t = \left[\frac{y_{\min,t}\left(e^{\beta+1} - 1\right)}{\beta + 1} + \frac{y_{\min,t}}{\beta} \right]^{-1}. \tag{90}$$

We then set $d_{it} = y_{it}/\bar{y}_t$ and $d_{\min,t} = y_{\min,t}/\bar{y}_t$, where $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$, which ensure that the outdegrees add up to N. It is worth noting that under this DGP

$$\Pr\left(d_{it} \geq d_{\min,t}\right) = \Pr\left(y_{it} \geq y_{\min,t}\right) = \frac{1}{\beta}\left(\frac{e^{\beta+1} - 1}{\beta + 1} + \frac{1}{\beta}\right)^{-1}, \tag{91}$$

which is time-invariant and depends only on the value of β.[15] The inverse transformation sampling method is used to generate y_{it} such that its distribution satisfies (89). To this end we first generate u_{it} as $IIDU[0, 1]$, $i = 1, 2, ..., N$; $t = 1, 2, ..., T$, and set

$$u_{\min,t} = C_t \left(\frac{y_{\min,t}}{\beta + 1}\right)\left(e^{\beta+1} - 1\right), \tag{92}$$

[14] Small sample properties of the extremum estimator for unbalanced panels are investigated in the Online Supplement.

[15] When $T > 1$, we construct a panel data assuming that all units maintain their relative dominance over time, and therefore for each t we sort d_{it} in a descending order.

and then generate y_{it} as

$$y_{it} = \begin{cases} -\dfrac{y_{\min,t}}{\beta+1}\ln\left[1-\dfrac{(\beta+1)\,u_{it}}{C_t e^{\beta+1} y_{\min,t}}\right], & \text{if } u_{it} < u_{\min,t} \\ \\ \left[\dfrac{\beta\,(y_{\min,t}-u_{it})+C_t y_{\min,t}}{C_t y_{\min,t}^{\beta+1}}\right]^{-1/\beta}, & \text{if } u_{it} \geq u_{\min,t} \end{cases}. \tag{93}$$

We carry out two sets of experiments based on the above two DGPs:

Exponent DGP: Observations on d_{it} are generated according to the exponent specification, (87), with a finite number of dominant units and a large number of non-dominant units. Specifically

- A.1. One strongly dominant unit: $\delta_{\max} = \delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$.
- A.2. Two strongly dominant units: $\delta_{\max} = \delta_{(1)} = \delta_{(2)} = 1$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$.
- A.3. One strongly dominant unit and one weakly dominant unit: $\delta_{\max} = \delta_{(1)} = 1$ and $\delta_{(2)} = 0.75$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$.

We consider all combinations of $N = 100, 300, 500$, and $1,000$, and $T = 1, 2, 6, 10$, and 20, and also provide simulation results for a very large data set with $N = 450,000$, which can arise when using inter-firm level sales data.[16] We focus on the 5 largest estimates of δ, which we denote by $\hat{\delta}_{\max} = \hat{\delta}_{(1)} \geq \hat{\delta}_{(2)} \geq ... \geq \hat{\delta}_{(5)}$, computed according to (75). When $T > 1$, the variance of $\hat{\delta}_{(i)}$ is computed by (78), with σ_v^2 estimated by (81).

Pareto DGP: Observations on d_{it} are generated according to the mixture Pareto distribution, (89), described above and we consider Experiments B.1: $\beta = 1.0$, and B.2: $\beta = 1.3$. The values of $y_{\min,t}$ are set as $y_{\min,t} = y_{\min} = 15$. The sample sizes are combinations of $N = 100, 300, 500, 1,000$, and $450,000$, and $T = 1$ and 2. We assess the performance of the Gabaix-Ibragimov estimator ($\hat{\beta}_{GI}$) given by (62) for different given cut-off values, $d_{\min,t}$, the maximum likelihood estimator ($\hat{\beta}_{MLE}$) given by (63) for different $d_{\min,t}$, and the CSN estimator ($\hat{\beta}_{CSN}$) which estimates β jointly with the cut-off value.

As shown in Proposition 3, the inverse of the extremum estimator, $1/\hat{\delta}_{\max}$, is a consistent estimator of β, and analogously one can consider the inverse of β as an estimator $\delta_{\max}$.[17] It is therefore of interest to see how the extremum estimator performs under the Pareto DGP, and conversely how the power law estimators perform under the Exponent DGP. To investigate the robustness of the alternative estimators of β to the choice of the underlying distribution, we conduct two sets of misspecification experiments where we compare the small sample properties of the four estimators of β, namely $\hat{\beta}_{GI}$, $\hat{\beta}_{MLE}$,

[16] For example, Carvalho et al. (2016) use a subset of data compiled by Tokyo Shoko Research Ltd that contains information on inter-firm transactions of around one million firms across Japan. This data set is proprietary and has not been made available to us.

[17] See also the discussions in Section 6 on the relationship between $\delta_{\max}$ and β.

$\hat{\beta}_{CSN}$, and $\hat{\beta}_{\max} = 1/\hat{\delta}_{\max}$, when the underlying DGP is Pareto, and conversely when Exponent DGP is used. We consider the values of $\beta = 1.0$, and 1.3 under Pareto DGP, and $\delta_{\max} = 1$ and 1/1.3 under Exponent DGP. We focus on small values of $T = 1$ and 2, for all combinations of $N = 100, 300, 500, 1,000$, and $450,000$.

All experiments are carried out with 2,000 replications.[18]

8.1 MC results

The estimation results under Exponent DGP for Experiments A.1 to A.3 are summarized in Table 1, and focus on the extremum estimators of $\delta_{\max} = \delta_{(1)}$ and $\delta_{(2)}$ when applicable. For each experiment we report bias ($\times 100$), root mean squared error (RMSE$\times$100), as well as size ($\times 100$) and power ($\times 100$) for the estimators under consideration. We first note that the bias and RMSE of the extremum estimators decline as N and/or T rises. The bias and RMSE reduction is particularly pronounced as T is increased. This is in line with the theoretical derivations which establish that along the cross-sectional dimension the rate of convergence is of order $1/\ln(N)$, as compared to $T^{-1/2}$ along the time dimension. We also note that the empirical sizes of the tests based on $\hat{\delta}_{\max}$ and $\hat{\delta}_{(2)}$ are close to the assumed 5% nominal size in most cases. There is some over-rejections in cases where T is much larger than N, and when there are more than one dominant units. In practice, this is unlikely to be a real concern since N is typically much larger than T. Seen from this perspective, it is particularly satisfying to note that the extremum estimator has satisfactory performance even when N approaches 450,000. The slow rate of convergence along the cross section dimension is, however, important for the power of the test. For example, in the case of Experiment A.1, the power of detecting the strongly dominant unit (against the alternative that $\delta_{\max} = 0.90$) is around 9% for $N = 100$ and $T = 2$, and rises only slowly as N is increased. However, we see a significant rise in power if T is increased to 6. For $T = 6$ the power rises from 17% for $N = 100$ to 89% for $N = 450,000$, twice as much as the values obtained for $T = 2$. The power also rises as the number of strongly dominant units is raised from one to two.

We also consider the frequency with which the dominant units are jointly selected under Exponent DGP, Experiments A.1 to A.3. The results are summarized in Table 2, and show that units with δ equal to unity are almost always correctly selected, especially when $T > 2$. The selection frequency can be low in the case of Experiment A.3 where the network has two dominant units one of which is weak with $\delta = 0.75$. In such cases we need N to be quite large if T is less than 3.

Tables 3 and 4 summarize the results for the first set of misspecification experiments where the data are generated from the Pareto tail distribution given by (70). For different values of β, the extremum estimator demonstrates robustness to the model misspecification, although it converges to the true value much more slowly than the other shape estimators under Pareto type distributions. This finding is in line with our theoretical results provided in 7.2.1. The extremum estimator, $\hat{\beta}_{\max} = 1/\hat{\delta}_{\max}$, performs well particularly when $\beta = 1$, even when N is relatively small. For example, under Pareto DGP with

[18] We also investigate the small sample properties of the extremum estimator of $\delta_{\max}$ for three other sets of experiments: (a) exponentially decaying δ_i's, (b) unbalanced panels, and (c) heteroskedastic and serially correlated errors in the case of large N and large T panels.

$\beta = 1$ (Experiment B.1), for $N = 300$ and $T = 2$, $\hat{\beta}_{\max} = 1.01$ (0.05), while $\hat{\beta}_{GI} = 1.04$ (0.19) and $\hat{\beta}_{MLE} = 1.05$ (0.14), assuming a 10% cut-off value.[19] It is also worth noting that the Gabaix-Ibragimov estimator ($\hat{\beta}_{GI}$) and the ML estimator ($\hat{\beta}_{MLE}$) are quite sensitive to the choice of the cut-off values.[20] The feasible MLE, $\hat{\beta}_{CSN}$, performs better, but it is important to note that the validity of the feasible MLE procedure critically depends on how close the assumed specification of the distribution of d_{it} below $d_{\min,t}$ is to the true underlying distribution.

Consider now the case where the outdegrees are generated according to the exponent specification, (87). In this misspecified case the Pareto estimators, $\hat{\beta}_{GI}$, $\hat{\beta}_{MLE}$, and $\hat{\beta}_{CSN}$, all show significant biases (see Tables 5 and 6). For instance, when $\beta = 1$, $N = 300$ and $T = 2$, and the outdegrees are generated according to the Exponent DGP, the bias of $\hat{\beta}_{GI}$ ranges from 0.10 to 0.35, for the cut-off values 10% to 30%. Also, the bias of $\hat{\beta}_{GI}$ increases rapidly with N. The ML type estimators exhibit similar biases.

Finally, the extremum estimator continues to perform well in the case of unbalanced panels, and large N and T panels with heteroskedastic and serially correlated errors. It is also reasonably robust to alternative network structures under different specifications of the distribution of outdegrees, such as exponentially decaying δ_i's. The relevant summary tables are available in the Online Supplement.

9 Dominant units in US production networks

In this section we apply the proposed estimation strategy to identify the top five most pervasive (dominant) sectors in the US economy. We also compare our results with the estimates of β (the inverse of $\delta_{\max}$) obtained by Acemoglu et al. (2012) for the most dominant sector. We provide estimates based on the US input-output tables for single years as well as when two or more input-output tables are pooled in an unbalanced panel. Acemoglu et al. (2012) only consider the estimates of β based on single-year input-output tables.

We begin with a re-examination of the data set used by Acemoglu et al. (2012) so that we have a direct comparison of the estimates of β (or its inverse) based on the shape of the power law, and the extremum estimates which is given by $\hat{\delta}_{\max}$, and computed using (74). The Acemoglu et al. (2012) data set are based on the US input-output accounts data over the period 1972-2002 compiled by the Bureau of Economic Analysis (BEA) every five years. We first confirmed that we can replicate their estimates of β, which we denote by $\hat{\beta}_{GI}$ assuming a 20% cut-off value (the percentage above which the degree sequences are assumed to follow the Pareto distribution). The estimates $\hat{\delta}_{\max}$ and the inverse of $\hat{\beta}$ for the years 1972, 1977, 1982, 1987, 1992, 1997 and 2002 are given in Tables 7 and 8. For the inverse of $\hat{\beta}$, Tables 7 and 8 report estimates based on the first-order and second-order interconnections, respectively.[21] We estimate β by the three approaches discussed above,

[19] Figures in brackets are standard errors.

[20] Similar Monte Carlo evidence illustrating the truncation sensitivity problem is reported in Table 1-4 of Gabaix and Ibragimov (2011). An interesting theoretical discussion can be found in Eeckhout (2004).

[21] The first-order degree of sector j is just its outdegree, d_j, defined as before, while the second-order degree of sector j is defined by $d_{j,2} = \mathbf{d}'\mathbf{w}_{\cdot j}$, where $\mathbf{d} = (d_1, d_2, ..., d_N)'$ is the vector of first-order degrees

namely Gabaix-Ibragimov estimator ($\hat{\beta}_{GI}$) given by (62), the MLE ($\hat{\beta}_{MLE}$) given by (63), and the feasible MLE ($\hat{\beta}_{CSN}$). For the Gabaix-Ibragimov regression and MLE, we give estimates for the cut-off values of 10%, 20%, and 30%. For the feasible MLE, we present both the estimates of β and the estimated cut-off values.[22]

The results in Tables 7 and 8 show that the yearly estimates of $\delta_{\max}$ are clustered within the narrow range of 0.77 to 0.82, covering a relatively long period of 30 years. We can not provide standard errors for such yearly estimates, but given the small over-time variations in these estimates we can confidently conclude that there is a high degree of sectoral pervasiveness in the US economy, although these estimates do not support the presence of a strongly dominant unit which requires $\hat{\delta}_{\max}$ to be close to unity. In contrast, the estimates of $\delta_{\max}$ based on the inverse of $\hat{\beta}$ differ considerably depending on the estimation methods, the choice of the cut-off value, and whether the first- or second-order interconnections are considered. For example, for 1972, the estimates based on the power law, inverse of $\hat{\beta}_{GI}$, range from 0.694 when the cut-off value is 10% and the first-order interconnections are used, and rise to 1.035 when the second-order interconnections are used with a 30% cut-off value. The estimates of δ based on the inverses of $\hat{\beta}_{GI}$ and $\hat{\beta}_{MLE}$, rise with the choice of cut-off values and with the order of interconnections, whilst our estimator does not require making such choices. Recall that $\delta_{\max}$ provides an exact measure of the rate at which the variance of aggregate output responds to sectoral shocks, whilst β characterizes a lower bound if the first-order interconnections are used. A 20% cut-off value, which is assumed by Acemoglu et al. (2012) seems reasonable, considering the closeness between the estimates of $\hat{\delta}_{\max}$ and the inverse of $\hat{\beta}_{GI}$, and given its similarity to the estimated cut-off values by the feasible MLE. Nevertheless, the estimated cut-off value based on the first-order interconnections for the year 1992 is only 9.5%, which is markedly lower than that for the other years. Similar issues arise when the second-order interconnections are used. The differences between $\hat{\delta}_{\max}$ and inverse of $\hat{\beta}_{GI}$ also vary across the years. For example, using the second-order interconnections and a cut-off value of 20%, $\hat{\delta}_{\max}$ and inverse of $\hat{\beta}_{GI}$ are reasonably close for the years 1992, 1997 and 2002, but diverge for the earlier years of 1972, 1977 and 1982.

The data sets provided by Acemoglu et al. (2012) do not give the identities of the sectors, which is fine if one is only interested in β or $\delta_{\max}$. But, as noted earlier, our estimation approach also allows us to identify the sectors with the highest degrees of pervasiveness in the production network. With this in mind, we compiled our own $\mathbf{W}$ matrices from the input-output tables downloaded from the BEA website.[23] The top five

and $\mathbf{w}_{\cdot j}$ is the j^{th} column of $\mathbf{W}$.

[22] Acemoglu et al. (2012) estimated the shape parameter of the power law by the log-log regression and non-parametric Nadaraya-Watson regression, taking the tail to correspond to the top 20% of the samples for each year and did not try other cut-off values. They also estimated the shape parameter by the feasible maximum likelihood method proposed by Clauset et al. (2009), but did not report the estimates for each year or the estimated cut-off values.

[23] The $\mathbf{W}$ matrices for different years were computed from commodity-by-commodity direct requirements tables at the detailed level that cover around 400-500 US industries. The $(i,j)^{th}$ entry of such a table shows the expenditure on commodity j per dollar of production of commodity i. As in Acemoglu et al. (2012), the terms sector and commodity are used interchangeably to convey the same meaning. These direct requirements tables can be derived from the total requirement tables at the detailed level, which are compiled by the BEA every five years. Further details on the data description and transfor-

largest estimates of δ, denoted by $\hat{\delta}_{\max} = \hat{\delta}_{(1)} \geq \hat{\delta}_{(2)} \geq ... \geq \hat{\delta}_{(5)}$, for each of the years 1972 to 2007 are given in Table 9. The identities of the associated sectors are given in Table 10. We note that both the degrees of dominance and the identities of the pervasive sectors in the US economy are relatively stable over the years. Consistent with the results in Table 7, no sector is strongly dominant. The highest estimate of $\delta_{\max}$ is 0.82, for the year 1992, with an average estimate of around 0.78 over the sample. The wholesale trade sector turns out to be the most dominant sector for all the years with the exception of 2002. In this year the management of companies and enterprises is the most dominant sector with the wholesale trade coming second.

But it is generally difficult to distinguish between the top two or three sectors as their δ estimates are quite close to one another and we are not able to apply formal statistical tests to their differences as standard errors can not be computed using outdegrees for one single year.[24] Accordingly, to provide more reliable estimates of $\delta_{(1)}$, $\delta_{(2)}, ..., \delta_{(5)}$ and the associated sectoral identities, we also consider pooled estimates. However, there have been major changes in the BEA industry classifications over the years, with the input-output tables for the period 1972-1992 being based on the Standard Industrial Classification (SIC) system, while starting from 1997 they are based on the North American Industry Classification System (NAICS). Accordingly, we computed panel estimates of δ for the two sub-samples separately. The results are summarized in Table 11, which also gives standard errors in parentheses, computed using (84). It is interesting that despite changes to the sectoral classifications, the wholesale trade sector is identified as the most dominant sector in both sub-samples, with $\hat{\delta}_{\max} = 0.762$ (0.036) for the first sub-sample (1972-1992), and $\hat{\delta}_{\max} = 0.716$ (0.045) for the second sub-sample (1997-2007). The two panel estimates are quite close and identify wholesale trade as the most dominant sector in the US economy. Turning to the estimates of $\delta_{(2)}$, $\delta_{(3)}, ..., \delta_{(5)}$, we find that these estimates are also remarkably similar across the two sub-samples, ranging from 0.667 to 0.605 in the first sub-sample, and 0.683 to 0.595 in the second sub-sample. What has changed is the identity of the sectors across the two sub-samples. For example, the second most dominant sector has been blast furnaces and steel mills over the first sub-sample (1972-1992), whilst it is management companies and enterprises over the second sub-sample (1997-2007).

10 Concluding remarks

This paper extends the production network considered by Acemoglu et al. (2012) and derives a dual price network, which allows us to obtain exact conditions under which sectoral shocks can have aggregate effects. The paper presents a simple nonparametric estimator of the degree of pervasiveness of sectoral shocks that compares favorably with the parametric estimators based on Pareto distribution fitted to the outdegrees. The proposed extremum estimator is simple to implement and is applicable not only to the pure cross section models where the Pareto shape parameter is estimated, but also extends

mations can be found in Appendix D.

[24] Acemoglu et al. (2012) are able to compute standard errors for their estimates of β because they impose a Pareto distribution on the ordered outdegrees beyond a cut-off point, which they assume.

readily to short T as well as large T panels. The paper also develops a simple test of the degree of pervasiveness of the most dominant units in the network, which is shown to have satisfactory size and power properties when N is large, even if T is quite small. The production and price networks considered in this paper are static, but the proposed statistical framework can be extended to allow for dynamics, along similar lines as in Pesaran and Chudik (2014) who consider aggregation of large dynamic panels.

Our empirical application to US input-output tables suggests some evidence of sector-specific shock propagation, but such effects do not seem sufficiently strong and long-lasting, and are likely to be dominated by common technological effects. Similar empirical evidence are also provided by Foerster, Sarte, and Watson (2011), who incorporate sectoral linkages into multisector growth models producing an approximate factor model. Their factor analytic approach, however, cannot distinguish dominant unit(s) from common factors and therefore may underestimate the influence of input-output linkages.[25] The issue of the relative importance of internal network interactions and external common shocks for macro economic fluctuations continues to be an open empirical question.

A Appendix: Lemmas

Lemma 1. *Let* $\mathbf{A}$ *be an* $N \times N$ *matrix whose entries are non-negative and each row sums up to* 1*. Then* $\lambda_1(\mathbf{A}) = 1$*, where* $\lambda_1(\mathbf{A})$ *is the largest eigenvalue of* $\mathbf{A}$*, and* $\mathbf{I}_N - \rho\mathbf{A}$ *is invertible given that* $|\rho| < 1$*.*

Proof. Matrix $\mathbf{A}$ is as a right stochastic matrix, and $\lambda_1(\mathbf{A}) = 1$ follows. See, for example, Property 10.1.2 in Stewart (2009). Given that $|\rho| < 1$ and $\lambda_1(\mathbf{A}) = 1$, it then readily seen that all eigenvalues of $\mathbf{I}_N - \rho\mathbf{A}$ are strictly positive in absolute value, and hence invertible. □

Remark 5. *It should be noted that this lemma holds irrespective of whether* $\mathbf{A}$ *has bounded column matrix norm. Also note that* $\lambda_1(\mathbf{A}') = 1$ *and* $\mathbf{I}_N - \rho\mathbf{A}'$ *is invertible, since a matrix and its transpose always have the same set of eigenvalues.*

Lemma 2. *Let* $\mathbf{A}$ *be an* $N \times N$ *matrix and* $\mathbf{B} = \mathbf{I}_N - \rho\mathbf{A}$*. Suppose that*

$$|\rho| < \max\left(1/\left\|\mathbf{A}\right\|_\infty, 1/\left\|\mathbf{A}\right\|_1\right).$$

Then $\mathbf{B}^{-1}$ *has bounded row and column sum matrix norms.*

Proof. See Pesaran (2015, p.756). □

B Appendix: Multiple dominant units

This appendix extends the analysis of Section 5 to the scenario where there are more than one dominant unit in the network. Specifically, we assume that the first m units are dominant with degrees of dominance $\{\delta_1, \delta_2, ..., \delta_m\}$, and the rest n units are non-dominant, with $\delta_i = 0$, for

[25] The factor analysis also requires large N and T panels and is not applicable when T is small.

$i = m+1, m+2, ..., m+n$, and let $N = m+n$. Consider now the following partitioned version of model (24)

$$\begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} \rho\mathbf{W}_{11} & \rho\mathbf{W}_{12} \\ \rho\mathbf{W}_{21} & \rho\mathbf{W}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} + \begin{pmatrix} \mathbf{g}_{1t} \\ \mathbf{g}_{2t} \end{pmatrix},$$

where $\mathbf{x}_{1t} = (x_{1t}, x_{2t}, ..., x_{mt})'$, $\mathbf{x}_{2t} = (x_{m+1,t}, x_{m+2,t}, ..., x_{Nt})'$, $\mathbf{W}_{11}$ is the $m \times m$ weight matrix associated with the dominant units, $\mathbf{W}_{22}$ is the $n \times n$ weight matrix associated with the non-dominant units and assumed to satisfy $|\rho| \|\mathbf{W}_{22}\|_1 < 1$, and $\mathbf{g}_{1t} = (g_{1t}, g_{2t}, ..., g_{mt})'$, $\mathbf{g}_{2t} = (g_{m+1,t}, g_{m+2,t}, ..., g_{Nt})'$, where $g_{it} = -b_i - \alpha(\gamma_i f_t + \varepsilon_{it})$, for $i = 1, 2, ..., N$. As $\varrho(\mathbf{W}_{22}) \leq 1$ and $|\rho| < 1$, we have

$$\mathbf{x}_{2t} = \mathbf{S}_{22}^{-1}\left(\rho\mathbf{W}_{21}\mathbf{x}_{1t} + \mathbf{g}_{2t}\right), \tag{B.1}$$

where $\mathbf{S}_{22} = \mathbf{I}_n - \rho\mathbf{W}_{22}$. Substituting (B.1) into

$$\mathbf{x}_{1t} = \rho\mathbf{W}_{11}\mathbf{x}_{1t} + \rho\mathbf{W}_{12}\mathbf{x}_{2t} + \mathbf{g}_{1t},$$

and rearranging yields

$$\mathbf{x}_{1t} = \mathbf{Z}_1^{-1}\mathbf{g}_{1t} + \rho\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}, \tag{B.2}$$

where

$$\mathbf{Z}_1 = \mathbf{I}_m - \rho\mathbf{W}_{11} - \rho^2\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{W}_{21}, \tag{B.3}$$

and $\mathbf{Z}_1$ is invertible as $(\mathbf{I}_N - \rho\mathbf{W})$ is nonsingular by Lemma 1 in Appendix A.

Now consider the cross-section average of x_{it} for $i = 1, 2, ..., N$,

$$\overline{\mathbf{x}}_{Nt} = N^{-1}\left(\boldsymbol{\tau}_m'\mathbf{x}_{1t} + \boldsymbol{\tau}_n'\mathbf{x}_{2t}\right).$$

Using (B.1) in the above equation gives

$$\overline{\mathbf{x}}_{Nt} = N^{-1}\left[\left(\boldsymbol{\tau}_m' + \rho\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\right)\mathbf{x}_{1t} + \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}\right],$$

and by the definition of $\mathbf{g}_{1t}$ we obtain

$$\overline{\mathbf{x}}_{Nt} = N^{-1}\left(-a_n + \boldsymbol{\theta}_n'\mathbf{x}_{1t} - \alpha\psi_n f_t - \alpha\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right),$$

where $a_n = \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{b}_2$, $\boldsymbol{\theta}_n' = \boldsymbol{\tau}_m' + \rho\boldsymbol{\phi}_n'\mathbf{W}_{21}$, $\psi_n = \boldsymbol{\phi}_n'\boldsymbol{\gamma}_2$, $\boldsymbol{\phi}_n' = \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}$, with $\mathbf{b}_2 = (b_{m+1}, b_{m+2}, ..., b_N)'$ and $\boldsymbol{\gamma}_2 = \left(\gamma_{m+1}, \gamma_{m+2}, ..., \gamma_N\right)'$.

We will derive $Var(\overline{\mathbf{x}}_{Nt})$ and inspect its asymptotic order of magnitude as $N \to \infty$ following similar steps as in Section 5. First, as with the case of the SAR(1) model with one dominant unit, we have $1 < \phi_{\min} \leq \phi_{\max} < K < \infty$, where $\boldsymbol{\phi}_n' = (\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$, $\phi_{\min} = \min(\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$, and $\phi_{\max} = \max(\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$. Also, it readily follows that $a_n = O(1)$ and $Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = \ominus\left(N^{-1}\right)$. Considering now the terms due to the dominant units, and using (B.2) note that

$$\begin{aligned} Cov\left(\mathbf{x}_{1t}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) &= Cov\left(\mathbf{Z}_1^{-1}\mathbf{g}_{1t} + \rho\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \\ &= -N^{-1}\rho\alpha\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}^{-1\prime}\boldsymbol{\tau}_n, \end{aligned} \tag{B.4}$$

$$Cov\left(\mathbf{x}_{1t}, f_t\right) = -\alpha Var(f_t)\left(\mathbf{Z}_1^{-1}\boldsymbol{\gamma}_1 + \rho\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_1\right), \tag{B.5}$$

$$\begin{aligned} Var\left(\mathbf{x}_{1t}\right) \;=\; & \alpha^2\mathbf{Z}_1^{-1}\mathbf{V}_{11,\varepsilon}\mathbf{Z}_1'^{-1} + \alpha^2\rho^2\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}^{-1\prime}\mathbf{W}_{12}'\mathbf{Z}_1'^{-1} \\ & +\alpha^2 Var(f_t)\left(\mathbf{Z}_1^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1'\mathbf{Z}_1'^{-1} + \rho^2\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2\boldsymbol{\gamma}_2'\mathbf{S}_{22}^{-1\prime}\mathbf{W}_{12}'\mathbf{Z}_1'^{-1}\right), \end{aligned} \tag{B.6}$$

where $\mathbf{V}_{11,\varepsilon} = diag\left(\sigma_1^2, \sigma_2^2, ..., \sigma_m^2\right)$, $\mathbf{V}_{22,\varepsilon} = diag\left(\sigma_{m+1}^2, \sigma_{m+2}^2, ..., \sigma_N^2\right)$, and $\boldsymbol{\gamma}_1 = (\gamma_1, \gamma_2, ..., \gamma_m)'$.

Consider now the individual terms of $Var(\overline{\mathbf{x}}_{Nt})$, which is given by

$$\begin{aligned} Var\left(\overline{\mathbf{x}}_{Nt}\right) &= N^{-2}\boldsymbol{\theta}_n' Var(\mathbf{x}_{1t})\boldsymbol{\theta}_n - 2\alpha N^{-2}\boldsymbol{\theta}_n' Cov\left(\mathbf{x}_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) + \alpha^2 N^{-2} Var\left(\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \\ &\quad + \alpha^2 N^{-2} Var(f_t)\left(\psi_n^2 + 2\psi_n\boldsymbol{\theta}_n'\mathbf{Z}_1^{-1}\boldsymbol{\gamma}_1 + 2\rho\psi_n\boldsymbol{\theta}_n'\mathbf{Z}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2\right). \end{aligned}$$

In the case where the network contains m dominant units but is not subject to common shocks,

$$Var\left(\overline{\mathbf{x}}_{Nt}\right) = N^{-2}\boldsymbol{\theta}_n' Var(\mathbf{x}_{1t})\boldsymbol{\theta}_n - 2\alpha N^{-2}\boldsymbol{\theta}_n' Cov\left(\mathbf{x}_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) + \ominus(N^{-1}). \tag{B.7}$$

Consider the i^{th} element of $N^{-1}\boldsymbol{\theta}_n$, denoted by $N^{-1}\theta_{i,n}$, for $i = 1, 2, ..., m$, and note that m is fixed and does not rise with N. Then by definition, $N^{-1}\theta_{i,n} = N^{-1}\left(1 + \rho\boldsymbol{\phi}_n'\mathbf{w}_{\cdot i,21}\right)$, where $\mathbf{w}_{\cdot i,21}$ is the i^{th} column of $\mathbf{W}_{21}$. Hence

$$\phi_{\min} N^{-1}\sum_{j=1}^{n} w_{ji,21} \leq N^{-1}\boldsymbol{\phi}_n'\mathbf{w}_{\cdot i,21} \leq \phi_{\max} N^{-1}\sum_{j=1}^{n} w_{ji,21},$$

and

$$N^{-1} + \phi_{\min} N^{-1}\sum_{j=1}^{n} w_{ji,21} \leq N^{-1}\theta_{i,n} \leq N^{-1} + \phi_{\max} N^{-1}\sum_{j=1}^{n} w_{ji,21}. \tag{B.8}$$

Also note that $\mathbf{w}_{\cdot i,21}'\boldsymbol{\tau}_n = \ominus(N^{\delta_i})$, which immediately follows that

$$\mathbf{w}_{\cdot i,21}'\boldsymbol{\tau}_n + \mathbf{w}_{\cdot i,11}'\boldsymbol{\tau}_m = d_i = \kappa_i N^{\delta_i},$$

with m being fixed. Therefore, by (B.8) it follows that $N^{-1}\theta_{i,n} = \ominus(N^{\delta_i - 1})$, for $i = 1, 2, ..., m$, and then $N^{-2}\boldsymbol{\theta}_n'\boldsymbol{\theta}_n = \ominus(N^{2\delta_{\max}-2})$, where $0 < \delta_{\max} = \max\left(\delta_1, \delta_2, ..., \delta_m\right) \leq 1$. Further notice that

$$N^{-2}\boldsymbol{\theta}_n'\boldsymbol{\theta}_n\lambda_m\left[Var(\mathbf{x}_{1t})\right] \leq N^{-2}\boldsymbol{\theta}_n' Var(\mathbf{x}_{1t})\boldsymbol{\theta}_n \leq N^{-2}\boldsymbol{\theta}_n'\boldsymbol{\theta}_n\lambda_1\left[Var(\mathbf{x}_{1t})\right],$$

where $\lambda_1\left[Var(\mathbf{x}_{1t})\right]$ and $\lambda_m\left[Var(\mathbf{x}_{1t})\right]$ denote the largest and smallest eigenvalue of $Var(\mathbf{x}_{1t})$, respectively, and $0 < \lambda_m\left[Var(\mathbf{x}_{1t})\right] \leq \lambda_1\left[Var(\mathbf{x}_{1t})\right] < K < \infty$. Hence we obtain

$$N^{-2}\boldsymbol{\theta}_n' Var(\mathbf{x}_{1t})\boldsymbol{\theta}_n = \ominus(N^{2\delta_{\max}-2}).$$

Turning now to the j^{th} element of the covariance term, for $j = 1, 2, ..., m$, we have

$$\left|Cov\left(x_{jt}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right)\right| \leq N^{-1}\left|\rho\alpha\right|\left\|\mathbf{Z}_{j.,1}^{-1}\right\|_\infty \left\|\mathbf{W}_{12}\right\|_\infty \left\|\mathbf{S}_{22}^{-1}\right\|_\infty \left\|\mathbf{V}_{22,\varepsilon}\right\|_\infty \left\|\boldsymbol{\phi}_n\right\|_\infty,$$

where $\mathbf{Z}_{j.,1}^{-1}$ denotes the j^{th} row of $\mathbf{Z}_1^{-1}$, and using similar line of reasoning as in the main text it is easily verified that $\left|Cov\left(x_{jt}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right)\right| = O\left(N^{-1}\right)$, and therefore

$$N^{-2}\boldsymbol{\theta}_n' Cov\left(\mathbf{x}_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = O(N^{\delta_{\max}-2}).$$

Consequently, in the absence of common shocks, we have demonstrated that

$$Var\left(\overline{\mathbf{x}}_{Nt}\right) = \ominus(N^{2\delta_{\max}-2}) + \ominus(N^{-1}),$$

which clearly shows that the rate of convergence of $\overline{\mathbf{x}}_{Nt}$ depends on the strongest dominant unit in the network.

Finally, if the network is subject to both dominant units and a common factor, it immediately follows from $N^{-1}\psi_n = \ominus(N^{\delta_\gamma - 1})$ and similar arguments as before that

$$Var\left(\overline{\mathbf{x}}_{Nt}\right) = \ominus(N^{2\delta_{\max}-2}) + \ominus(N^{2\delta_\gamma - 2}) + \ominus(N^{-1}),$$

which is a direct extension of (58) to the multiple-dominant-units network, and shows that the relative magnitudes between $\delta_{\max}$ and δ_γ determines the limiting properties of the aggregate effects.

C Appendix: Consistency of $\hat{\delta}_{\max}$

First we note that since z_i are distributed independently with finite means and variances then

$$E\left(\bar{z}_N\right) = N^{-1}\sum_{i=1}^{N} E\left(z_i\right) = \beta^{-1}\Pr\left(z \geq 0\right) + E\left(z\,|z<0\right)\left[1 - \Pr\left(z \geq 0\right)\right],$$

which is finite. Further using standard results for the moments of ordered random variables (see, for example, Section 4.6 of Arnold et al. (1992)) we have

$$E\left(z_{(i)}\right) = (1/\beta)\left(\sum_{j=1}^{N-i+1}\frac{1}{j}\right),\ Var(z_{(i)}) = (1/\beta)^2\left(\sum_{j=1}^{N-i+1}\frac{1}{j^2}\right), \text{ for } i = 1, 2, ..., N. \tag{C.1}$$

Taking expectations and variance of $\hat{\delta}_{\max}$ given by (71), and making use of the above results we now have

$$E\left(\hat{\delta}_{\max}\right) = \frac{E\left(z_{\max}\right) - E\left(\bar{z}_N\right)}{\ln N} = \frac{(1/\beta)\sum_{j=1}^{N} j^{-1} - E\left(\bar{z}_N\right)}{\ln N}, \tag{C.2}$$

$$\begin{aligned} Var\left(\hat{\delta}_{\max}\right) &= \frac{Var\left(z_{\max}\right) + N^{-2}\sum_{i=1}^{N} Var(z_i) - 2N^{-1}\sum_{i=1}^{N} Cov(z_{\max}, z_{(i)})}{\left(\ln N\right)^2} \\ &= \frac{Var\left(z_{\max}\right) + N^{-2}\sum_{i=1}^{N} Var(z_i) - 2N^{-1}\sum_{i=1}^{N} Var(z_{(i)})}{\left(\ln N\right)^2}. \end{aligned} \tag{C.3}$$

Also using well known bounds to harmonic series (see, for example, Section 3.1 and 3.2 of Bonar et al. (2006)), we have

$$\ln(N+1) < \left(\sum_{j=1}^{n}\frac{1}{j}\right) \leq 1 + \ln N,$$

and hence

$$\lim_{N\to\infty}\frac{\sum_{j=1}^{N} j^{-1}}{\ln N} = 1. \tag{C.4}$$

Using (C.1) and (C.4) in (C.2) we now have $\lim_{N\to\infty} E\left(\hat{\delta}_{\max}\right) = 1/\beta$.

Turning to the variance of $\hat{\delta}_{\max}$, we note that

$$
\begin{aligned}
Var\left(\hat{\delta}_{\max}\right) &= \frac{Var\left(z_{\max}\right) + \left(\frac{1-2N}{N^2}\right)\sum_{i=1}^{N} Var(z_{(i)})}{\left(\ln N\right)^2}, \\
\left(\ln N\right)^{-2} Var\left(z_{\max}\right) &\leq Var\left(\hat{\delta}_{\max}\right) \leq \left(\ln N\right)^{-2}\left[Var\left(z_{\max}\right) + \left(\frac{2N-1}{N^2}\right) N Var\left(z_{\max}\right)\right], \\
\left(\ln N\right)^{-2} \delta^2 \left(\sum_{j=1}^{N} \frac{1}{j^2}\right) &\leq Var\left(\hat{\delta}_{\max}\right) \leq \left(\ln N\right)^{-2} \left(\frac{3N-1}{N}\right) \delta^2 \left(\sum_{j=1}^{N} \frac{1}{j^2}\right).
\end{aligned}
$$

But $\sum_{j=1}^{N} j^{-2} \leq \pi^2/6$, and hence $Var\left(\hat{\delta}_{\max}\right) = O\left[\left(\ln N\right)^{-2}\right]$.

D Data appendix

The input-output accounts data are obtained from the Bureau of Economic Analysis (BEA) website at http://www.bea.gov/industry/io_annual.htm. The input-output tables at the finest level of disaggregation are compiled every five years, and the latest available data are for year 2007. We derive the Commodity-by-Commodity Direct Requirements (**DR**) table by applying the following formula:

$$\mathbf{DR} = \left(\mathbf{TR} - \mathbf{I}\right)\left(\mathbf{TR}\right)^{-1},$$

where **I** is an identity matrix, and **TR** denotes the Commodity-by-Commodity Total Requirements table that is available from the BEA. The input-output matrix, **W**, is set to the transpose of **DR** and row-standardized so that the intermediate input shares sum to one for each sector. The sectors without any direct requirements and those with zero outdegrees are excluded from **W**.

References

Acemoglu, D., U. Akcigit, and W. Kerr (2016). Networks and the macroeconomy: An empirical exploration. In M. Eichenbaum and J. Parker (Eds.), *NBER Macroeconomics Annual 2015*, Volume 30, Chapter 4, pp. 276 – 335. University of Chicago Press.

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016). Import competition and the great US employment sag of the 2000s. *Journal of Labor Economics 34*, S141–S198.

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*, 1977–2016.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2016). Networks, shocks, and systemic risk. In Y. Bramoulle, A. Galeotti, and B. Rogers (Eds.), *The Oxford Handbook of the Economics of Networks*, Chapter 21, pp. 569 – 609. Oxford University Press.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992). *A first course in order statistics*, Volume 54. SIAM-Society for Industrial and Applied Mathematics.

Bai, J. and K. Li (2013). Spatial panel data models with common shocks. Available at SSRN 2373628.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics 31*, 929–960.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.

Bonar, D. D., M. J. Khoury Jr, and M. Khoury (2006). *Real infinite series*. Mathematical Association of America.

Carvalho, V. M. (2014). From micro to macro via production networks. *The Journal of Economic Perspectives 28*, 23–47.

Carvalho, V. M., M. Nirei, Y. Saito, and A. Tahbaz-Salehi (2016). Supply chain disruptions: Evidence from the Great East Japan Earthquake. Available at http://www.columbia.edu/ at2761/JapanEQ.pdf.

Chudik, A. and M. H. Pesaran (2013). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews 32*, 592–649.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal 14*, C45–C90.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM review 51*, 661–703.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *The American Economic Review 94*, 1429–1451.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy 119*, 1–38.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*, 733–772.

Gabaix, X. and R. Ibragimov (2011). Rank- 1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics 29*, 24–39.

Hill, B. M. et al. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*, 1163–1174.

Horvath, M. (1998). Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks. *Review of Economic Dynamics 1*, 781–808.

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

Long, J. B. and C. I. Plosser (1983). Real business cycles. *The Journal of Political Economy 91*, 39–69.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics 46*, 323–351.

Pesaran, M. H. (2015). *Time series and panel data econometrics.* Oxford University Press.

Pesaran, M. H. and A. Chudik (2014). Aggregation in large dynamic panels. *Journal of Econometrics 178*, 273–285.

Siavash, S. S. (2016). Dominant sectors in the US: A factor model analysis of sectoral industrial production. Unpublished manuscript. Goethe University Frankfurt.

Stewart, W. J. (2009). *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling.* Princeton University Press.

Yang, C. F. (2016). Identification and estimation of spatial autoregressive models with common factors. Available at SSRN: https://ssrn.com/abstract=2856818.

Table 1: Bias, RMSE, size and power of the extremum estimator for the dominant unit or units under Exponent DGP for Experiments A.1 to A.3

	Bias(×100)					RMSE(×100)					Size(×100)					Power(×100)				
$T\backslash N$	100	300	500	1,000	450,000	100	300	500	1,000	450,000	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Experiment A.1: One strongly dominant unit, $\delta_{\max} = 1$																				
1	-0.90	-0.46	-0.31	-0.21	-0.06	20.87	17.28	15.89	14.34	7.63	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-1.29	-0.55	-0.40	-0.29	-0.10	15.31	12.40	11.37	10.24	5.44	5.75	5.25	4.90	5.05	4.95	8.90	11.55	12.75	14.85	44.45
6	-1.19	-0.49	-0.35	-0.23	-0.07	8.83	7.09	6.51	5.85	3.10	4.75	4.60	4.60	4.60	4.65	16.70	25.45	30.40	37.40	89.10
10	-1.26	-0.55	-0.39	-0.27	-0.09	6.98	5.61	5.14	4.62	2.45	5.35	5.35	5.15	5.20	5.35	24.45	39.10	46.55	56.35	98.75
20	-1.14	-0.44	-0.29	-0.19	-0.04	4.97	3.94	3.61	3.24	1.72	5.25	4.80	4.90	4.95	4.90	45.65	67.90	76.85	86.10	100.00
Experiment A.2: Two strongly dominant units, $\delta_{(1)} = \delta_{(2)} = 1$																				
$\delta_{(1)} = 1$																				
1	9.89	8.91	8.39	7.72	4.20	20.38	16.99	15.73	14.25	7.63	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	6.15	5.93	5.65	5.24	2.89	13.99	11.78	10.92	9.91	5.32	4.00	4.45	4.25	4.45	4.75	14.95	20.25	22.70	26.75	67.40
6	2.58	3.02	2.99	2.85	1.62	7.75	6.66	6.22	5.68	3.07	2.95	3.85	4.15	4.40	5.05	25.05	41.50	49.00	59.15	98.25
10	1.76	2.37	2.39	2.31	1.33	6.01	5.24	4.91	4.50	2.44	3.20	4.45	4.80	5.45	5.85	36.45	62.00	71.20	80.85	100.00
20	0.68	1.50	1.60	1.59	0.95	3.99	3.52	3.34	3.08	1.70	2.30	3.40	3.95	4.45	5.40	60.45	89.10	95.00	98.55	100.00
$\delta_{(2)} = 1$																				
1	-13.87	-10.66	-9.58	-8.48	-4.39	22.05	18.06	16.53	14.82	7.82	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-10.66	-7.64	-6.81	-5.97	-3.06	16.22	12.54	11.40	10.17	5.34	6.45	5.15	4.80	5.15	4.60	1.95	1.55	1.60	2.00	20.40
6	-7.09	-4.79	-4.18	-3.60	-1.80	10.24	7.67	6.92	6.14	3.19	8.55	6.10	5.75	5.60	5.25	3.10	7.05	9.20	14.15	78.40
10	-6.02	-3.92	-3.38	-2.88	-1.42	8.16	5.95	5.32	4.70	2.43	9.40	6.05	5.15	4.70	4.15	3.85	14.45	21.15	32.00	97.50
20	-4.79	-2.91	-2.45	-2.05	-0.98	6.19	4.32	3.82	3.34	1.71	11.35	6.85	6.05	5.00	4.70	13.90	44.05	58.05	73.80	100.00
Experiment A.3: One strongly dominant unit and one weakly dominant unit, $\delta_{(1)} = 1$, $\delta_{(2)} = 0.75$																				
$\delta_{(1)} = 1$																				
1	1.68	1.30	1.11	0.83	-0.02	18.75	15.68	14.54	13.27	7.56	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-0.73	-0.21	-0.16	-0.15	-0.10	14.11	11.77	10.93	9.99	5.44	3.60	4.00	4.05	4.30	4.95	8.30	11.20	12.40	14.65	44.45
6	-1.82	-0.72	-0.49	-0.30	-0.07	8.76	7.07	6.50	5.85	3.10	4.60	4.30	4.40	4.45	4.65	14.50	24.30	29.45	36.80	89.10
10	-2.00	-0.80	-0.54	-0.35	-0.09	7.14	5.63	5.15	4.63	2.45	5.75	5.25	5.20	5.20	5.35	20.40	37.60	44.90	55.85	98.75
20	-1.89	-0.69	-0.44	-0.26	-0.05	5.19	3.97	3.63	3.25	1.72	6.15	5.05	4.80	4.85	4.90	38.80	66.05	75.60	85.90	100.00
$\delta_{(2)} = 0.75$																				
1	-2.94	-1.97	-1.66	-1.32	-0.17	16.42	14.77	14.13	13.29	7.74	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-3.19	-1.33	-0.89	-0.53	-0.07	13.63	11.30	10.48	9.54	5.22	3.05	3.40	3.10	3.80	4.40	45.10	57.60	63.90	71.95	99.40
6	-2.19	-0.88	-0.61	-0.40	-0.12	8.94	7.19	6.61	5.96	3.16	5.30	5.20	5.35	5.65	5.65	87.15	95.20	97.15	99.00	100.00
10	-1.76	-0.59	-0.35	-0.17	0.00	7.01	5.55	5.08	4.57	2.42	5.80	4.95	4.70	4.90	4.65	97.30	99.55	99.90	100.00	100.00
20	-1.72	-0.55	-0.31	-0.14	0.02	5.03	3.87	3.54	3.18	1.69	6.30	5.10	4.90	4.80	5.20	100.00	100.00	100.00	100.00	100.00

Notes: The Data Generating Process (DGP) is given by the exponent specification (87). For Experiment A.1, there is one strongly dominant unit and the rest of the units are non-dominant: $\delta_{\max} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$. For Experiment A.2, there are two strongly dominant units and the rest are non-dominant: $\delta_{(1)} = \delta_{(2)} = 1$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$. For Experiment A.3, there are one strongly dominant unit and one weakly dominant unit, and the rest are non-dominant: $\delta_{(1)} = 1$ and $\delta_{(2)} = 0.75$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$. $\delta_{(i)}$ denotes the i^{th} largest δ, i.e., $\delta_{\max} = \delta_{(1)} \geq \delta_{(2)} \geq \delta_{(3)} \geq ...$, which are estimated by (75). The standard error of $\hat{\delta}_{(i)}$ is computed using (78), for $T \geq 2$. The nominal size of the test is 5%, and power is computed at 0.9 if the true value is 1, and at 1 if the true value is 0.75. The number of replications is to 2,000.

Table 2: Frequencies with which the dominant unit or units are jointly selected, under Exponent DGP for Experiments A.1 to A.3

	Empirical frequency (percent)				
$T \backslash N$	100	300	500	1,000	450,000
Experiment A.1: One strongly dominant unit, $\delta_{\max} = 1$					
1	97.25	99.55	99.80	99.90	100.0
2	100.0	100.0	100.0	100.0	100.0
6	100.0	100.0	100.0	100.0	100.0
10	100.0	100.0	100.0	100.0	100.0
20	100.0	100.0	100.0	100.0	100.0
Experiment A.2: Two strongly dominant units, $\delta_{(1)} = \delta_{(2)} = 1$					
1	94.20	99.10	99.65	99.85	100.0
2	100.0	100.0	100.0	100.0	100.0
6	100.0	100.0	100.0	100.0	100.0
10	100.0	100.0	100.0	100.0	100.0
20	100.0	100.0	100.0	100.0	100.0
Experiment A.3: One strongly dominant unit and one weakly dominant unit, $\delta_{(1)} = 1$, $\delta_{(2)} = 0.75$					
1	61.55	75.90	79.70	85.45	98.75
2	86.30	91.90	93.65	95.65	100.0
6	97.75	99.10	99.50	99.85	100.0
10	99.65	99.95	100.0	100.0	100.0
20	99.95	100.0	100.0	100.0	100.0

Notes: This table complements Table 1 and reports the frequencies with which the dominant units are jointly selected across $2,000$ replications. See also the notes to Table 1.

Table 3: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{\max}$, under Pareto DGP for Experiment B.1 ($\beta = 1$)

	$T=1$					$T=2$				
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed				**Log-log regression** $\left(\widehat{\beta}_{GI}\right)$						
cut-off value										
10%	1.11	1.02	1.01	1.00	1.00	1.11	1.04	1.02	1.01	1.00
	(0.50)	(0.26)	(0.20)	(0.14)	(0.01)	(0.35)	(0.19)	(0.14)	(0.10)	(0.00)
20%	1.04	1.01	1.00	1.00	1.00	1.06	1.02	1.01	1.00	1.00
	(0.33)	(0.18)	(0.14)	(0.10)	(0.00)	(0.24)	(0.13)	(0.10)	(0.07)	(0.00)
30%	1.02	1.00	1.00	1.00	1.00	1.04	1.01	1.00	1.00	1.00
	(0.26)	(0.15)	(0.12)	(0.08)	(0.00)	(0.19)	(0.11)	(0.08)	(0.06)	(0.00)
Infeasible					*Using true* $d_{\min,t}$					
cut-off value	1.03	1.00	1.00	1.00	1.00	1.05	1.02	1.01	1.00	1.00
24%	(0.30)	(0.17)	(0.13)	(0.09)	(0.00)	(0.22)	(0.12)	(0.09)	(0.06)	(0.00)
Assumed				**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$						
cut-off value										
10%	1.24	1.07	1.04	1.02	1.00	1.15	1.05	1.03	1.01	1.00
	(0.39)	(0.20)	(0.15)	(0.10)	(0.00)	(0.26)	(0.14)	(0.10)	(0.07)	(0.00)
20%	1.11	1.03	1.02	1.01	1.00	1.07	1.02	1.01	1.00	1.00
	(0.25)	(0.13)	(0.10)	(0.07)	(0.00)	(0.17)	(0.09)	(0.07)	(0.05)	(0.00)
30%	1.06	1.01	1.01	1.00	1.00	1.02	0.99	0.99	0.98	0.99
	(0.19)	(0.11)	(0.08)	(0.06)	(0.00)	(0.13)	(0.07)	(0.06)	(0.04)	(0.00)
Infeasible					*Using true* $d_{\min,t}$					
cut-off value	1.09	1.03	1.01	1.01	1.00	1.04	1.01	1.00	1.00	1.00
24%	(0.23)	(0.12)	(0.09)	(0.07)	(0.00)	(0.15)	(0.08)	(0.07)	(0.05)	(0.00)
Estimated					**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$					
cut-off value	44%	38%	37%	35%	24%	37%	33%	31%	29%	22%
	1.02	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	1.00
	(0.17)	(0.10)	(0.08)	(0.06)	(0.00)	(0.13)	(0.08)	(0.06)	(0.04)	(0.00)
					$\widehat{\beta}_{\max} = 1/\ \widehat{\delta}_{\max}$					
	1.04	1.03	1.02	1.02	1.00	1.01	1.01	1.00	1.00	1.00
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.08)	(0.05)	(0.04)	(0.04)	(0.01)

Notes: The DGP follows the Pareto tail distribution given by (70) with $\beta = 1$. $d_{\min,t}$ denotes the assumed lower bound for the Pareto distribution. The cut-off value refers to the percentage of the largest observations (sorted in descending order) that are assumed to follow the Pareto distribution. The infeasible cut-off value is computed by (91) assuming the true value of $d_{\min,t}$ is known. All estimates are averaged across $2,000$ replications. Standard errors are in parentheses. $\hat{\beta}_{GI}$ is the Gabaix-Ibragimov estimator obtained by running the log-log regression, (62). $\widehat{\beta}_{MLE}$ is computed by (63). $\hat{\beta}_{CSN}$ is calculated by applying the joint MLE procedure described in Clauset et al. (2009). $\hat{\delta}_{\max}$ is computed according to (75), and its standard error by (78). The standard error for the inverse of $\hat{\delta}_{\max}$ is computed by the delta method. (N/A) indicates that the standard error of $\hat{\delta}_{\max}$ cannot be computed when $T = 1$.

Table 4: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{\max}$, under Pareto DGP for Experiment B.2 ($\beta = 1.3$)

	$T=1$					$T=2$				
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
	Log-log regression $\left(\widehat{\beta}_{GI}\right)$									
Assumed cut-off value										
10%	1.44	1.33	1.31	1.30	1.30	1.42	1.34	1.32	1.30	1.30
	(0.65)	(0.34)	(0.26)	(0.18)	(0.01)	(0.45)	(0.24)	(0.19)	(0.13)	(0.01)
20%	1.35	1.31	1.30	1.29	1.30	1.36	1.32	1.31	1.30	1.30
	(0.43)	(0.24)	(0.18)	(0.13)	(0.01)	(0.30)	(0.17)	(0.13)	(0.09)	(0.00)
30%	1.31	1.29	1.29	1.29	1.29	1.32	1.30	1.29	1.29	1.29
	(0.34)	(0.19)	(0.15)	(0.11)	(0.00)	(0.24)	(0.14)	(0.11)	(0.07)	(0.00)
Infeasible	*Using true* $d_{\min,t}$									
cut-off value	1.37	1.31	1.30	1.30	1.30	1.37	1.32	1.31	1.30	1.30
16%	(0.49)	(0.27)	(0.20)	(0.14)	(0.01)	(0.34)	(0.19)	(0.14)	(0.10)	(0.00)
	Maximum Likelihood Estimation $\left(\widehat{\beta}_{MLE}\right)$									
Assumed cut-off value										
10%	1.61	1.39	1.35	1.32	1.30	1.48	1.35	1.33	1.31	1.30
	(0.51)	(0.25)	(0.19)	(0.13)	(0.01)	(0.33)	(0.17)	(0.13)	(0.09)	(0.00)
20%	1.44	1.34	1.32	1.31	1.30	1.37	1.32	1.31	1.30	1.30
	(0.32)	(0.17)	(0.13)	(0.09)	(0.00)	(0.22)	(0.12)	(0.09)	(0.06)	(0.00)
30%	1.34	1.28	1.26	1.26	1.25	1.28	1.25	1.25	1.24	1.25
	(0.24)	(0.13)	(0.10)	(0.07)	(0.00)	(0.17)	(0.09)	(0.07)	(0.05)	(0.00)
Infeasible	*Using true* $d_{\min,t}$									
cut-off value	1.49	1.35	1.33	1.31	1.30	1.39	1.33	1.31	1.31	1.30
16%	(0.37)	(0.19)	(0.15)	(0.10)	(0.00)	(0.24)	(0.13)	(0.10)	(0.07)	(0.00)
Estimated	**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$									
cut-off value	39%	32%	30%	28%	17%	33%	28%	26%	24%	17%
	1.31	1.30	1.30	1.30	1.30	1.31	1.30	1.30	1.30	1.30
	(0.23)	(0.14)	(0.11)	(0.08)	(0.00)	(0.18)	(0.11)	(0.08)	(0.06)	(0.00)
	$\widehat{\beta}_{\max} = 1/\ \widehat{\delta}_{\max}$									
	1.27	1.27	1.27	1.27	1.27	1.24	1.25	1.25	1.25	1.27
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.08)	(0.05)	(0.04)	(0.03)	(0.00)

Notes: The DGP follows the Pareto tail distribution given by (70) with $\beta = 1.3$. See also the notes to Table 3.

Table 5: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{\max}$, under Exponent DGP for Experiment A.1 ($\beta = 1$)

	$T=1$					$T=2$				
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed cut-off value	**Log-log regression** $\left(\widehat{\beta}_{GI}\right)$									
10%	0.98	1.10	1.20	1.36	2.39	0.97	1.10	1.20	1.37	2.39
	(0.44)	(0.29)	(0.24)	(0.19)	(0.02)	(0.31)	(0.20)	(0.17)	(0.14)	(0.01)
20%	1.11	1.28	1.39	1.54	2.11	1.11	1.29	1.39	1.55	2.11
	(0.35)	(0.23)	(0.20)	(0.15)	(0.01)	(0.25)	(0.17)	(0.14)	(0.11)	(0.01)
30%	1.17	1.34	1.44	1.56	1.91	1.18	1.35	1.45	1.57	1.91
	(0.30)	(0.20)	(0.17)	(0.13)	(0.01)	(0.22)	(0.14)	(0.12)	(0.09)	(0.01)
Assumed cut-off value	**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$									
10%	1.53	1.74	1.84	1.95	2.11	1.44	1.71	1.82	1.93	2.11
	(0.48)	(0.32)	(0.26)	(0.19)	(0.01)	(0.32)	(0.22)	(0.18)	(0.14)	(0.01)
20%	1.52	1.64	1.68	1.73	1.79	1.46	1.62	1.67	1.72	1.79
	(0.34)	(0.21)	(0.17)	(0.12)	(0.01)	(0.23)	(0.15)	(0.12)	(0.09)	(0.00)
30%	1.42	1.49	1.51	1.54	1.58	1.38	1.48	1.51	1.54	1.58
	(0.26)	(0.16)	(0.12)	(0.09)	(0.00)	(0.18)	(0.11)	(0.09)	(0.06)	(0.00)
Estimated	**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$									
cut-off value	39%	29%	24%	18%	2%	36%	26%	21%	16%	1%
	1.37	1.58	1.69	1.85	2.83	1.36	1.59	1.71	1.87	2.87
	(0.24)	(0.19)	(0.17)	(0.15)	(0.04)	(0.17)	(0.13)	(0.12)	(0.11)	(0.03)
	$\hat{\beta}_{\max} = 1/\ \widehat{\delta}_{\max}$									
	1.06	1.04	1.03	1.02	1.01	1.04	1.02	1.02	1.01	1.00
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.16)	(0.13)	(0.12)	(0.10)	(0.05)

Notes: The DGP is given by the exponent specification, (87). There is one strongly dominant unit and the rest are non-dominant: $\delta_{\max} = \delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$, where $\delta_{(i)}$ denotes the i^{th} largest δ. The true value of β is $\beta = 1$. See also the notes to Table 3 for other details.

Table 6: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{\max}$, under Exponent DGP for Experiment A.1 ($\beta = 1.3$)

	$T=1$					$T=2$				
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed cut-off value	**Log-log regression** $\left(\widehat{\beta}_{GI}\right)$									
10%	1.43	1.53	1.61	1.76	2.40	1.38	1.51	1.61	1.76	2.40
	(0.64)	(0.39)	(0.32)	(0.25)	(0.02)	(0.44)	(0.28)	(0.23)	(0.18)	(0.01)
20%	1.45	1.59	1.67	1.79	2.11	1.44	1.60	1.68	1.79	2.11
	(0.46)	(0.29)	(0.24)	(0.18)	(0.01)	(0.32)	(0.21)	(0.17)	(0.13)	(0.01)
30%	1.44	1.57	1.64	1.72	1.92	1.44	1.58	1.65	1.73	1.92
	(0.37)	(0.23)	(0.19)	(0.14)	(0.01)	(0.26)	(0.17)	(0.13)	(0.10)	(0.01)
Assumed cut-off value	**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$									
10%	1.84	1.89	1.95	2.01	2.11	1.70	1.85	1.92	1.99	2.11
	(0.58)	(0.35)	(0.28)	(0.20)	(0.01)	(0.38)	(0.24)	(0.19)	(0.14)	(0.01)
20%	1.65	1.70	1.72	1.75	1.79	1.58	1.67	1.71	1.74	1.79
	(0.37)	(0.22)	(0.17)	(0.12)	(0.01)	(0.25)	(0.15)	(0.12)	(0.09)	(0.00)
30%	1.50	1.52	1.54	1.55	1.58	1.45	1.51	1.53	1.55	1.58
	(0.27)	(0.16)	(0.13)	(0.09)	(0.00)	(0.19)	(0.11)	(0.09)	(0.06)	(0.00)
Estimated	**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$									
cut-off value	38%	26%	22%	16%	2%	32%	22%	18%	13%	1%
	1.50	1.70	1.80	1.95	2.83	1.51	1.72	1.83	1.99	2.87
	(0.28)	(0.22)	(0.19)	(0.17)	(0.04)	(0.21)	(0.17)	(0.15)	(0.13)	(0.03)
	$\widehat{\beta}_{\max} = 1/\ \widehat{\delta}_{\max}$									
	1.35	1.36	1.35	1.35	1.31	1.38	1.34	1.34	1.33	1.31
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.21)	(0.17)	(0.15)	(0.14)	(0.07)

Notes: The DGP is given by the exponent specification, (87). There is one strongly dominant unit and the rest of the units are non-dominant: $\delta_{\max} = 1/1.3 = 0.77$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$, where $\delta_{(i)}$ denotes the i^{th} largest δ. The true value of β is $\beta = 1.3$. See also the notes to Table 5.

Table 7: Yearly estimates of the degree of dominance, $\delta_{\max}$, and inverse of the shape parameter of power law, β, based on the first-order interconnections, using US input-ouput tables compiled by Acemoglu et al. (2012)

			$\widehat{\delta}_{\max}$ based on the inverse of $\widehat{\beta}$ using the first-order interconnections							
			Inverse of $\widehat{\beta}_{GI}$			Inverse of $\widehat{\beta}_{MLE}$			Inverse of $\widehat{\beta}_{CSN}$	
			Assumed cut-off value			Assumed cut-off value				Estimated
Year	N	$\widehat{\delta}_{\max}$	10%	20%	30%	10%	20%	30%		cut-off value
1972	483	0.767	0.694	0.727	0.832	0.736	0.829	1.135	0.728	16.8%
			(0.142)	(0.104)	(0.098)	(0.106)	(0.095)	(0.145)	(0.081)	
1977	524	0.778	0.677	0.725	0.804	0.715	0.852	1.009	0.726	13.6%
			(0.133)	(0.100)	(0.091)	(0.099)	(0.099)	(0.114)	(0.086)	
1982	529	0.788	0.717	0.739	0.818	0.719	0.786	1.039	0.741	15.3%
			(0.139)	(0.101)	(0.092)	(0.099)	(0.084)	(0.119)	(0.082)	
1987	510	0.804	0.667	0.731	0.814	0.724	0.849	1.028	0.742	13.3%
			(0.132)	(0.102)	(0.093)	(0.101)	(0.099)	(0.118)	(0.090)	
1992	476	0.824	0.672	0.758	0.842	0.738	0.891	1.002	0.706	9.5%
			(0.137)	(0.110)	(0.100)	(0.107)	(0.110)	(0.114)	(0.105)	
1997[a]	474	0.778	0.625	0.698	0.791	0.617	0.909	0.982	0.670	13.1%
			(0.129)	(0.101)	(0.094)	(0.090)	(0.137)	(0.131)	(0.085)	
2002	417	0.765	0.639	0.687	0.759	0.685	0.756	0.930	0.730	19.4%
			(0.139)	(0.107)	(0.096)	(0.106)	(0.092)	(0.113)	(0.081)	

Notes: Estimates are obtained using the data sets provided by Acemoglu et al. (2012), which are based on the US input-output account data by the Bureau of Economic Analysis (BEA). N is the total number of sectors in a given year and the standard errors are in parentheses. $\hat{\delta}_{\max}$ is the largest estimate of δ computed using (69). The first-order degree sequence is used in the estimation of the shape parameter of the power law, β. $\hat{\beta}_{GI}$ is obtained by the log-log regression with Gabaix and Ibragimov (2011) correction using the OLS regression defined by (62). $\hat{\beta}_{MLE}$ is the maximum likelihood estimate (MLE) of β computed by (63). A 10% cut-off value, for example, means that the Pareto tail is taken to be the top 10% of all sectors in terms of outdegrees in each year. Acemoglu et al. (2012) report $\hat{\beta}_{GI}$ estimates only based on a 20% cut-off point. $\hat{\beta}_{CSN}$ is the feasible MLE proposed by Clauset et al. (2009) and its estimated cut-off values are reported in the last column of the table.
[a] From the year 1997 and thereafter, the BEA input-output tables are based on the North American Industry Classification System (NAICS), while for the earlier years they are based on the Standard Industrial Classification (SIC) system.

Table 8: Yearly estimates of the degree of dominance, $\delta_{\max}$, and inverse of the shape parameter of power law, β, based on the second-order interconnections, using US input-ouput tables compiled by Acemoglu et al. (2012)

			$\widehat{\delta}_{\max}$ based on the inverse of $\widehat{\beta}$ using the second-order interconnections							
			Inverse of $\widehat{\beta}_{GI}$			Inverse of $\widehat{\beta}_{MLE}$			Inverse of $\widehat{\beta}_{CSN}$	
			Assumed cut-off value			Assumed cut-off value				Estimated
Year	N	$\widehat{\delta}_{\max}$	10%	20%	30%	10%	20%	30%		cut-off value
1972	483	0.767	0.719	0.880	1.035	0.873	1.126	1.353	0.973	15.7%
			(0.147)	(0.126)	(0.122)	(0.126)	(0.147)	(0.174)	(0.112)	
1977	524	0.778	0.718	0.870	1.008	0.821	1.058	1.351	0.750	9.4%
			(0.141)	(0.120)	(0.114)	(0.114)	(0.133)	(0.177)	(0.107)	
1982	529	0.788	0.773	0.913	1.013	0.885	1.028	1.329	1.088	23.6%
			(0.150)	(0.125)	(0.114)	(0.122)	(0.116)	(0.158)	(0.097)	
1987	510	0.804	0.686	0.879	1.031	0.883	1.070	1.325	1.110	22.9%
			(0.136)	(0.123)	(0.118)	(0.124)	(0.128)	(0.161)	(0.103)	
1992	476	0.824	0.661	0.869	1.012	0.750	1.014	1.277	0.818	12.2%
			(0.135)	(0.126)	(0.120)	(0.108)	(0.141)	(0.182)	(0.107)	
1997[a]	474	0.778	0.632	0.790	0.955	0.648	1.100	1.202	0.666	12.0%
			(0.130)	(0.115)	(0.113)	(0.095)	(0.192)	(0.187)	(0.088)	
2002	417	0.765	0.620	0.768	0.954	0.721	0.998	1.245	0.772	13.4%
			(0.135)	(0.119)	(0.121)	(0.111)	(0.151)	(0.192)	(0.103)	

Notes: This table differs from Table 7 in that the second-order degree sequence is used to produce the estimates of β. See also the notes to Table 7 for further details.

Table 9: Yearly estimates of the degree of dominance, δ, for the top five pervasive sectors, using US input-ouput tables (our data)

Year	N	$\widehat{\delta}_{(1)}$	$\widehat{\delta}_{(2)}$	$\widehat{\delta}_{(3)}$	$\widehat{\delta}_{(4)}$	$\widehat{\delta}_{(5)}$
1972	446	0.764	0.740	0.701	0.638	0.608
1977	468	0.774	0.704	0.628	0.608	0.590
1982	468	0.786	0.669	0.655	0.635	0.619
1987	457	0.802	0.669	0.657	0.633	0.629
1992	451	0.823	0.678	0.677	0.646	0.631
1997[a]	452	0.775	0.725	0.635	0.622	0.597
2002	408	0.758	0.743	0.639	0.563	0.560
2007	365	0.722	0.649	0.606	0.591	0.550

Notes: Estimates are obtained using the input-output accounts data downloaded from the Bureau of Economic Analysis (BEA) website. See Appendix D for details of the data sources and their transformations. The table reports the five largest yearly estimates of δ, computed using (75), denoted by $\hat{\delta}_{(1)} = \hat{\delta}_{\max}, \hat{\delta}_{(2)}, \cdots, \hat{\delta}_{(5)}$. N is the number of sectors with non-zero outdegrees.
[a] From the year 1997 and thereafter, the BEA input-output tables are based on the North American Industry Classification System (NAICS), while for the previous years they are based on the Standard Industrial Classification (SIC) system.

Table 10: Identities of the top five pervasive sectors based on the yearly estimates of δ

Year	The top five pervasive sectors
1972	Wholesale trade
	Blast furnaces and steel mills
	Real estate
	Miscellaneous business services
	Motor freight transportation & warehousing
1977	Wholesale trade
	Blast furnaces and steel mills
	Real estate
	Petroleum refining
	Industrial inorganic & organic chemicals
1982	Wholesale trade
	Blast furnaces and steel mills
	Petroleum refining
	Private electric services (utilities)
	Advertising
1987	Wholesale trade
	Blast furnaces and steel mills
	Advertising
	Motor freight transportation and warehousing
	Electric services (utilities)
1992	Wholesale trade
	Real estate agents, managers, operators, and lessors
	Blast furnaces and steel mills
	Trucking and courier services, except air
	Advertising
1997[a]	Wholesale trade
	Management of companies and enterprises
	Real estate
	Iron and steel mills
	Truck transportation
2002	Management of companies and enterprises
	Wholesale trade
	Real estate
	Electric power generation, transmission, and distribution
	Iron and steel mills and ferroalloy manufacturing
2007	Wholesale trade
	Management of companies and enterprises
	Other real estate
	Iron and steel mills and ferroalloy manufacturing
	Petroleum refineries

Notes: This table complements Table 9 and reports the identities of those sectors corresponding to the five largest estimates of δ (in descending order) for each year.
[a] From the year 1997 and thereafter, the BEA input-output tables are based on the North American Industry Classification System (NAICS), while for the previous years they are based on the Standard Industrial Classification (SIC) system.

Table 11: Pooled panel estimates of the degree of dominance, δ, for the top five pervasive sectors, using US input-output tables for the two sub-periods 1972 -1992 and 1997-2007

	Sub-sample 1972-1992		Sub-sample 1997-2007	
$\widehat{\delta}_{(1)}$	0.762	Wholesale trade	0.716	Wholesale trade
	(0.036)		(0.045)	
$\widehat{\delta}_{(2)}$	0.667	Blast furnaces	0.683	Management of companies
	(0.036)	and steel mills	(0.045)	and enterprises
$\widehat{\delta}_{(3)}$	0.642	Real estate	0.609	Real estate[a]
	(0.036)		(0.045)	
$\widehat{\delta}_{(4)}$	0.605	Trucking and courier	0.598	Iron and steel mills
	(0.036)	services, except air	(0.045)	
$\widehat{\delta}_{(5)}$	0.605	Miscellaneous business	0.595	Other real estate[a]
	(0.036)	services	(0.045)	
N	548		619	
T	5		3	

Notes: The pooled estimates for the years 1972, 1977, 1982, 1987 and 1992 are based on US input-output data using the Bureau of Economic Analysis (BEA) industry codes, which are in turn based on the Standard Industrial Classification (SIC). For the years 1997, 2002 and 2007, the sectoral classifications are based on the BEA industry codes, which are based on the North American Industry Classification System (NAICS). The table gives the five largest panel estimates of δ together with the identities of the associated sectors. The estimates are denoted by $\hat{\delta}_{(1)} = \hat{\delta}_{\max}, \hat{\delta}_{(2)}, \cdots, \hat{\delta}_{(5)}$, and computed using (83). The standard errors are given in parentheses and computed using (84). N is the total number of sectors with non-zero outdegrees, and T is the number of time periods in the panel.
[a] In the BEA industry classifications, the real estate sector was subdivided into housing and other real estate sectors starting from 2007. Since the pooled estimates are based on unbalanced panels constructed according to BEA codes, real estate and other real estate are considered as two sectors.