UNIVERSITY OF
CAMBRIDGE

Faculty of Economics

Cambridge Working Paper Economics

Econometric Analysis of Production Networks with Dominant Units

M. Hashem
Pesaran

Cynthia
Fan Yang

9 October 2016

This paper builds on the work of Acemoglu et al. (2012) and considers a production network with unobserved common technological factor and establishes general conditions under which the network structure contributes to aggregate fluctuations. It introduces the notions of strongly and weakly dominant units, and shows that at most a finite number of units in the network can be strongly dominant, while the number of weakly dominant units can rise with N (the cross section dimension). This paper further establishes the equivalence between the highest degree of dominance in a network and the inverse of the shape parameter of the power law. A new extremum estimator for the degree of pervasiveness of individual units in the network is proposed, and is shown to be robust to the choice of the underlying distribution. Using Monte Carlo techniques, the proposed estimator is shown to have satisfactory small sample properties. Empirical applications to US input-output tables suggest the presence of production sectors with a high degree of pervasiveness, but their effects are not sufficiently pervasive to be considered as strongly dominant.

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

October 9, 2016

Abstract

This paper builds on the work of Acemoglu et al. (2012) and considers a production network with unobserved common technological factor and establishes general conditions under which the network structure contributes to aggregate fluctuations. It introduces the notions of strongly and weakly dominant units, and shows that at most a finite number of units in the network can be strongly dominant, while the number of weakly dominant units can rise with N (the cross section dimension). This paper further establishes the equivalence between the highest degree of dominance in a network and the inverse of the shape parameter of the power law. A new extremum estimator for the degree of pervasiveness of individual units in the network is proposed, and is shown to be robust to the choice of the underlying distribution. Using Monte Carlo techniques, the proposed estimator is shown to have satisfactory small sample properties. Empirical applications to US input-output tables suggest the presence of production sectors with a high degree of pervasiveness, but their effects are not sufficiently pervasive to be considered as strongly dominant.

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

---

1 Introduction

Following the seminal contribution of Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), there has been considerable interest in production networks and the role that individual production units (firms/sectors) can play in propagation of shocks across the economy. (Horvath (1998, 2000), Gabaix (2011), Acemoglu, Akcigit, and Kerr (2016), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2016), and Siavash (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 production and price networks with unobserved common technological factor and establish conditions under which the network structure contributes to aggregate fluctuations. 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 the total number of sectors. 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. We also show that the largest δ, denoted by $\delta_{(1)}$, is the inverse of β, the shape parameter of the Pareto distribution, the form assumed by Acemoglu et al. (2012) and others in the literature.

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) derive only a lower bound on the decay rate of sector-specific shocks on aggregate outcomes and 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 highest δ (abstracting from unobserved common shocks). 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 $\delta_{(1)}$ is an extremum estimator of the shape parameter of the Pareto distribution, β, as it depends on the largest outdegree which is adjusted by cross-sectional averages. In addition, our approach also allows us to estimate the degrees of dominance, δ_i, for a given sector i so long as $\delta_{(i)} > 1/2$.

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 $\delta_{(1)}$, 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 the highest δ lying between 0.77 and 0.82. Our results 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 latter procedure produces varying results under different cut-off points and different orders of interconnections assumed. To provide more reliable estimates, we also conduct panel estimation and find that the highest δ 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 both the estimates of δ and the identities of the pervasive sectors are rather stable throughout the whole period from 1972 to 2007, with the wholesale trade sector identified repeatedly as the most dominant sector for almost all years and for the pooled sample periods as well. Most importantly, our results suggest that no sector in the US economy is strongly dominant, which requires the value of δ to be close to unity, whilst we obtain at most 0.76 with a standard error of 0.037. Overall, our estimates and analyses support the view that sector-specific shocks have some macro effects, but we do not find such effects to be sufficiently strong to be long-lasting.

The rest of the paper is organized as follows. Section 2 presents an economic model of production and price networks that generalizes the model by Acemoglu et al. (2012). Section 3 introduces the concept of the degree of dominance, characterizes its properties in a network, and derives exact conditions under which micro (sectoral) shocks can lead to aggregate fluctuations in the framework of spatial models. Section 4 introduces the extremum estimator for the degree of dominance in the context of balanced and unbalanced panels, and derives its asymptotic distribution. It also compares the proposed estimator with the inverse of the estimator of the shape parameter of the power law distribution. Section 5 sets up the Monte Carlo experiments and discusses the results. Section 6 presents the empirical application to the US production network, and finally Section 7 concludes. Some of the mathematical details and a brief discussion of data sources and transformations are provided in the appendix.

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 sequences 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 and price networks

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). 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

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

where u_{it} is the productivity shock composed of a sector-specific shock, ε_{it}, and a common technological factor, f_t. Specifically,

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

where γ_i denotes the extent to which the common factor influences sector i, and is known as a factor loading. 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 $Var(\gamma_i) = \sigma_\gamma^2 > 0$.

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 $0 < \underline{\sigma}^2 < \sigma_i^2 < \bar{\sigma}^2 < K < \infty$. 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, αw_{ij}, denotes the share of labor, and $(1 - \alpha) w_{ij}$ is the share of jth output in the ith 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

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

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

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

The competitive equilibrium solution of the economy for given sector prices, $P_{1t}, P_{2t}, \ldots, P_{Nt}$, and the wage rate, w_t, is given by

$$q_{ij,t} = \frac{(1 - \alpha) w_{ij} P_{it} q_{it}}{P_{jt}}, \tag{6}$$

and

$$l_{it} = \frac{\alpha P_{it} q_{it}}{W_t}. \tag{7}$$

Substituting the above results in (1) and simplifying yields

$$p_{it} = (1 - \alpha) \sum_{j=1}^{N} w_{ij} p_{jt} + \alpha w_t - b_i - \alpha \left(\gamma_i f_t + \varepsilon_{it}\right), \tag{8}$$

where $p_{it} = \log\left(P_{it}\right)$, $w_t = \log\left(W_t\right)$, and

$$b_i = \alpha \log\left(\alpha\right) + (1 - \alpha) \log\left(1 - \alpha\right) + (1 - \alpha) \sum_{j=1}^{N} w_{ij} \log(w_{ij}), \tag{9}$$

for $i = 1, 2, \ldots, N$. In cases where $w_{ij} = 0$, we set $w_{ij} \log\left(w_{ij}\right) = 0$ as well. A dual to the price equation in (8) can also be obtained using (6) in (3) to obtain

$$S_{it} = (1 - \alpha) \sum_{j=1}^{N} w_{ji} S_{jt} + C_{it}, \tag{10}$$

where $C_{it} = P_{it} c_{it}$, and $S_{it} = P_{it} q_{it}$ is the sales of sector i.

In matrix notation the 'price' system, (8), can be written as

$$\mathbf{p}_t = (1 - \alpha)\,\mathbf{W}\mathbf{p}_t + \alpha w_t \boldsymbol{\tau}_N - (\mathbf{b} + \alpha\boldsymbol{\gamma} f_t + \alpha\boldsymbol{\varepsilon}_t)\,, \tag{11}$$

where $\mathbf{p}_t = (p_{1t}, p_{2t}, \ldots, p_{Nt})'$, $\mathbf{W}$ is an $N \times N$ matrix $\mathbf{W} = (w_{ij})$, $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)'$, $\mathbf{b} = (b_1, b_2, \ldots, b_N)'$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. The dual of the price equation, the sales equation, (10), can also be written as

$$\mathbf{S}_t = (1 - \alpha)\,\mathbf{W}'\mathbf{S}_t + \mathbf{C}_t, \tag{12}$$

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 (12) as compared to the price equations in (11).

Finally, aggregating (7) over i, we have

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

or

$$l_t W_t = \alpha \sum_{i=1}^{N} S_{it} = \alpha \boldsymbol{\tau}_N' \mathbf{S}_t. \tag{13}$$

Also using (12)

$$\mathbf{S}_t = [\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}']^{-1}\,\mathbf{C}_t, \tag{14}$$

where $[\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}']^{-1}$ is known as the Leontief inverse.[1] Using this result in (13) now yields the following expression for the total wage bill,

$$l_t W_t = \alpha \boldsymbol{\tau}_N' [\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}']^{-1}\,\mathbf{C}_t. \tag{15}$$

Similarly, solving (11) for the log-price vector, $\mathbf{p}_t$, and using Lemma 1 in Appendix A again we have

$$\begin{aligned} \mathbf{p}_t = {} & \alpha w_t\,[\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}]^{-1}\,\boldsymbol{\tau}_N \\ & + \alpha\,[\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}]^{-1}\left(-\alpha^{-1}\mathbf{b} - \boldsymbol{\gamma} f_t - \boldsymbol{\varepsilon}_t\right). \end{aligned} \tag{16}$$

Let

$$\bar{p}_t = N^{-1} \sum_{i=1}^{N} p_{it} = N^{-1}\boldsymbol{\tau}_N'\mathbf{p}_t, \tag{17}$$

be the aggregate log price index, and use (16) to obtain

$$\begin{aligned} \bar{p}_t = {} & \left\{ \frac{\alpha}{N}\boldsymbol{\tau}_N'\,[\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}]^{-1}\,\boldsymbol{\tau}_N \right\} w_t \\ & - \frac{\alpha}{N}\boldsymbol{\tau}_N'\,[\mathbf{I}_N - (1 - \alpha)\,\mathbf{W}]^{-1}\,\boldsymbol{\xi}_t, \end{aligned} \tag{18}$$

---

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

where $\boldsymbol{\xi}_t = \alpha^{-1}\mathbf{b} + \boldsymbol{\gamma}f_t + \boldsymbol{\varepsilon}_t$. But since $w_{ij} \geq 0$, $\mathbf{W}\boldsymbol{\tau}_N = \boldsymbol{\tau}_N$, and $0 < \alpha < 1$, then $[\mathbf{I}_N - (1-\alpha)\mathbf{W}]^{-1}\boldsymbol{\tau}_N = \boldsymbol{\tau}_N/\alpha$, and hence (18) can also be written as

$$w_t - \overline{p}_t = \boldsymbol{v}'\boldsymbol{\xi}_t, \tag{19}$$

where

$$\boldsymbol{v} = \frac{\alpha}{N}\left[\mathbf{I}_N - (1-\alpha)\mathbf{W}'\right]^{-1}\boldsymbol{\tau}_N. \tag{20}$$

Acemoglu et al. (2012) refer to $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)'$ as the 'influence vector' (their equation (4)), and show that $v_i \geq 0$, and $\boldsymbol{\tau}_N'\boldsymbol{v} = 1$. Writing (19) in terms of the common and sector-specific shocks we now have

$$w_t - \overline{p}_t = \alpha^{-1}\left(\boldsymbol{v}'\mathbf{b}\right) + \left(\boldsymbol{v}'\boldsymbol{\gamma}\right)f_t + \boldsymbol{v}'\boldsymbol{\varepsilon}_t, \tag{21}$$

where, using (9),

$$\boldsymbol{v}'\mathbf{b} = \alpha\log(\alpha) + (1-\alpha)\log(1-\alpha) + (1-\alpha)\sum_{i=1}^{N}\sum_{j=1}^{N}v_i w_{ij}\log(w_{ij}).$$

The above (log) real-wage equation, (21), generalizes equation (3) in Acemoglu et al. (2012) by allowing for time variations in prices and technologies.

Equations (15) and (21) are dual of each other. (15) gives the total wage bill in terms of a weighted sum of consumption expenditures, with the weights given by $\alpha\left[\mathbf{I}_N - (1-\alpha)\mathbf{W}\right]^{-1}\boldsymbol{\tau}_N$. Whilst (21) gives the log of the real wage rate in terms of the common and aggregate sectoral shocks, namely f_t and $\boldsymbol{v}'\boldsymbol{\varepsilon}_t$. The key issue is how much of the cyclical fluctuations in (log) real wages is due to common shocks $(\boldsymbol{v}'\boldsymbol{\gamma})^2 Var(f_t)$ and how much is due to the sectoral shocks, $Var(\boldsymbol{v}'\boldsymbol{\varepsilon}_t)$. Recall that common and sectoral shocks are assumed to be uncorrelated. Acemoglu et al. (2012) focus on the asymptotic properties of $\boldsymbol{v}'\boldsymbol{\varepsilon}_t$, as $N \to \infty$. Since $Var(\boldsymbol{v}'\boldsymbol{\varepsilon}_t) = \boldsymbol{v}'Var(\boldsymbol{\varepsilon}_t)\boldsymbol{v}$, then it follows that

$$\overline{\sigma}^2(\boldsymbol{v}'\boldsymbol{v}) \geq Var(\boldsymbol{v}'\boldsymbol{\varepsilon}_t) \geq \underline{\sigma}^2(\boldsymbol{v}'\boldsymbol{v}),$$

and the asymptotic properties of $Var(\boldsymbol{v}'\boldsymbol{\varepsilon}_t)$ is governed by that of $\boldsymbol{v}'\boldsymbol{v}$. Acemoglu et al. (2012, p.2009) show that[2]

$$\boldsymbol{v}'\boldsymbol{v} \geq k_{\alpha,0}N^{-1} + k_{\alpha,1}N^{-2}\sum_{j=1}^{N}d_j^2, \tag{22}$$

where $k_{\alpha,0}$ and $k_{\alpha,1}$ are finite constants that depend on α, and d_j is defined by

$$d_j = \sum_{i=1}^{N}w_{ij}, \tag{23}$$

which is the jth column sum of $\mathbf{W}$. Therefore, to analyze the limiting behavior of $Var(\boldsymbol{v}'\boldsymbol{\varepsilon}_t)$, 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.

---

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

3   A spatial representation of the price network

In this section we consider the price network equation system given by (11) and write it more generally as

$$\mathbf{x}_t = \rho \mathbf{W} \mathbf{x}_t - \mathbf{b}(\rho, \mathbf{W}) - (1 - \rho)\left(\boldsymbol{\gamma} f_t + \boldsymbol{\varepsilon}_t\right), \tag{24}$$

where

$$\mathbf{x}_t = \mathbf{p}_t - w_t \boldsymbol{\tau}_N, \tag{25}$$

$\rho = 1 - \alpha$, and the ith element of the $N \times 1$ vector $\mathbf{b}(\rho, \mathbf{W})$ is given by (9), which can be written as

$$b_i(\rho, \mathbf{W}) = (1 - \rho)\log(1 - \rho) + \rho \log(\rho) + \rho \sum_{j=1}^{N} w_{ij} \log(w_{ij}), \text{ for } i = 1, 2, ..., N. \tag{26}$$

Equation (24) represents a first-order spatial autoregressive (SAR(1)) model with an unobserved common factor, where the price-specific intercepts, b_i, depend on ρ and the weight matrix $\mathbf{W}$. This model can also be used to derive an expression for $w_t - \bar{p}_t$ which is taken as a measure of GDP in the literature. Averaging (25) and using (17) we have

$$\bar{x}_{N,t} = \frac{1}{N} \boldsymbol{\tau}_N' \mathbf{x}_t = -(w_t - \bar{p}_t).$$

Also using (21)

$$-\bar{x}_{N,t} = \alpha^{-1}\left(\boldsymbol{v}'\mathbf{b}\right) + \left(\boldsymbol{v}'\boldsymbol{\gamma}\right)f_t + \boldsymbol{v}'\boldsymbol{\varepsilon}_t.$$

Therefore, one can derive the limiting properties of $\boldsymbol{v}'\boldsymbol{\varepsilon}_t$, which has been the subject of intensive investigation by Acemoglu et al. (2012), using the SAR model (24).

There are two advantages in directly focusing on the SAR model. First, it allows us to relate to the network and spatial econometrics literature, and derive conditions under which α, $\boldsymbol{\gamma}$ and σ_i^2 can be estimated from panel data on wage rate and prices. The direct use of the SAR model also enables us to provide exact bounds on $Var\left(w_t - \bar{p}_t\right)$ rather than the different lower bounds obtained by Acemoglu et al. (2012) for different degrees of higher-order interconnections. Instead, by considering the SAR model 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.

We also face a number of challenges in the empirical analysis of the SAR model given by (24). Estimation of this model can be carried out using standard techniques, such as ML or GMM, only when the spatial weight matrix, $\mathbf{W}$, has bounded row and column norms, namely if

$$\|\mathbf{W}\|_\infty < K < \infty, \text{ and } \|\mathbf{W}\|_1 < K < \infty.$$

Since $\mathbf{W}$ is row-standardized then it follows that $\|\mathbf{W}\|_\infty \leq 1$, and we only need to consider the behavior of the column norm of $\mathbf{W}$. The presence of the common factor, f_t, also complicates the estimation problem but does not pose significant difficulties if $\mathbf{W}$ is row and column bounded. See the recent contributions of Bai and Li (2013) and Yang (2016) on estimation of SAR models with unobserved common factors.

3.1 Characterization of networks in terms of strongly and weakly dominant units

Suppose now that $\|\mathbf{W}\|_1$ is not bounded in N, but continue to assume that $\mathbf{W}$ is row-standardized and $w_{ij} \geq 0$, for all i and j. Denote the jth column of $\mathbf{W}$ by $\mathbf{w}_{\cdot j}$ and the associated column sum by $d_j = \boldsymbol{\tau}'_N \mathbf{w}_{\cdot j}$, also defined by (23). In the network literature d_j is known as the weighted outdegree of unit j, which is proportional to the number of units in the network that connect to unit j. We rule out units with zero outdegrees and assume throughout that $d_j > 0$, for all j.

Definition 1. (δ-dominant) We shall refer to unit j as δ_j-*dominant* if its weighted outdegree, d_j, is of order N^{δ_j}, or more specifically if

$$d_j = \kappa_j N^{\delta_j}, \text{ for } j = 1, 2, ..., N, \tag{27}$$

for $\kappa_j > 0$ and $0 \leq \delta_j \leq 1$. In particular, 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 (or pervasiveness) of unit j in the network.[3]

In the standard case where the column norm of $\mathbf{W}$ is bounded we must have $\delta_j = 0$ for all j, that is, all units are non-dominant. $\mathbf{W}$ will have an unbounded column norm 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

$$\boldsymbol{\tau}'_N \mathbf{d} = \boldsymbol{\tau}'_N \mathbf{W}' \boldsymbol{\tau}_N = N. \tag{28}$$

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

$$\sum_{j=1}^{N} \kappa_j N^{\delta_j} = N, \tag{29}$$

for a fixed N and as $N \to \infty$. Let

$$S_N = N^{-1} \sum_{j=1}^{N} \kappa_j N^{\delta_j}, \tag{30}$$

and note that S_N must converge to 1, and it is therefore necessary that $S_N > 0$ and is bounded in N from above. In Appendix B we show that

$$S_N \leq \kappa_{\max} \left(1 + \frac{N-1}{N} \sum_{i=1}^{N} \delta_i \right),$$

and since $\kappa_{\max} < K$, it then follows that S_N will be bounded in N if the following summability condition is met

$$\sum_{j=1}^{N} \delta_j < K < \infty. \tag{31}$$

---

[3] δ_j can also be viewed as the power exponent of the outdegree of unit j in the network.

To ensure that $S_N > 0$ for all N and as $N \to \infty$, let $\kappa_{\min} = \min(\kappa_1, \kappa_2, ..., \kappa_N)$, and $\delta_{\min} = \min(\delta_1, \delta_2, ..., \delta_N)$, and note that since by assumption $\kappa_{\min} > 0$, then we must also have

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

which in turn implies

$$\kappa_{\min}N^{\delta_{\min}} < 1. \tag{32}$$

It is clear that this condition 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. We refer to this latter condition as the min condition and summarize the above findings 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 that power function, (27), with δ_j being the degree of dominance of unit j in the network. Then $\{\delta_j\}$ must be summable, and not all units of the network can be δ-dominant, with $\delta_j > 0$ for all j.*

Consider now the case where a finite number of units, say m, are δ-dominant (with $\delta_j > 0$) and the rest are non-dominant (namely with $\delta_j = 0$). In this case we have

$$N = \sum_{j=1}^{m} \kappa_j N^{\delta_j} + \sum_{j=m+1}^{N} \kappa_j \geq m\kappa_{\min}N^{\delta_{m,\min}} + (N - m)\kappa_{\min},$$

and it follows that

$$mN^{\delta_{m,\min}-1} - \frac{m}{N} \leq \frac{1 - \kappa_{\min}}{\kappa_{\min}}. \tag{33}$$

This condition is satisfied for all values of $0 \leq \delta_{m,\min} \leq 1$, and the m dominant units can be strongly or weakly dominant. To summarize, it is possible for the network to contain many weakly dominant units so long as the degrees of pervasiveness of the units satisfy the summability and min conditions given by (31) and (32), respectively.

In the case where m is allowed to rise with N, it is clear from Proposition 1 that m can not rise at the same rate as N. Let

$$m = cN^\epsilon, \ \ 0 \leq \epsilon < 1, \text{ and } c > 0. \tag{34}$$

Then condition (33) can be written as

$$N^{\delta_{m,\min}+\epsilon-1} - N^{\epsilon-1} \leq \frac{1-\kappa_{\min}}{c\kappa_{\min}},$$

which is met for all values of N (and as $N \to \infty$) if $\delta_{m,\min} + \epsilon \leq 1$, since $\epsilon < 1$. Note that we must have $\epsilon = 0$ if $\delta_{m,\min} = 1$, which re-iterates that the number of strongly dominant units must be fixed and cannot rise with N. If $\delta_{m,\min} < 1$, then the number of weakly dominant units could rise with N but at a slower rate such that $\epsilon \leq 1 - \delta_{m,\min}$ is satisfied. Notice that the greater the value of $\delta_{m,\min}$, the lower the rate at which the number of weakly dominant units can expand with N. The next proposition summarizes these findings.

Proposition 2. *Consider the network represented by* $\mathbf{W}$, *and assume that* $\mathbf{W}$ *is row-standardized,* $w_{ij} \geq 0$, *and no unit has zero outdegree. Then the number of strongly dominant units must be fixed and cannot rise with* N. *The number of weakly dominant units can rise with* N *at a rate of* ϵ *as long as* $0 \leq \epsilon \leq 1 - \delta_{m,\min}$ *holds, where* $\delta_{m,\min} = \min(\delta_1, \delta_2, ..., \delta_m) < 1$, *and* δ_j *is the degree of dominance of unit* j *in the network. (see Definition 1).*

Remark 1. *Propositions 1 and 2 imply that there are three possible network structures in terms of* δ-*dominance. The first one is a network consisting of a fixed number of strongly dominant units and the remaining units are non-dominant. The second possibility is that there exist a (large) number of weakly dominant units, the number of which could increase with* N *but at a slower rate, and the rest are non-dominant units. The third scenario is a combination of the first two. For example, one could have a fixed number of strongly dominant units, and a large number of weakly dominant units with degrees of dominance,* δ_j, *that decay exponentially.*

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.*

3.2 The relationship between the two measures of dominance

The exponent $\delta_{\max}$ 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 (27) in (22) to obtain

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

where $(N-m)^{-1} \sum_{j=m+1}^{N} \kappa_j^2 = O(1)$. Also, recall that $m = cN^{\epsilon}$ and

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

where as before $\delta_{\max} = \max(\delta_1, \delta_2, ..., \delta_N)$, and $\kappa_{\max} = \max(\kappa_1, \kappa_2, ..., \kappa_N)$.

Then the limiting behavior of $\boldsymbol{v}'\boldsymbol{v}$ will be determined by $N^{2(\delta_{\max}-1)+\epsilon}$, 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 $\boldsymbol{v}'\boldsymbol{v}$ does not converge to zero as $N \to \infty$, the 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 $Var(\overline{x}_{N,t})$ 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 $N^{2(\delta_{\max}-1)+\epsilon}$ established as a lower bound for $\boldsymbol{v}'\boldsymbol{v}$, we have

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

or $\delta_{\max} \geq 1/\beta - \tilde{\epsilon}$, where $\tilde{\epsilon} = \epsilon_\beta + \epsilon/2 > 0$.

It also follows that $\delta_{\max} \geq 1/\varsigma - \tilde{\epsilon}_\varsigma$, where $\tilde{\epsilon}_\varsigma > 0$, 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 as we shall see are not needed since it is possible to derive an exact rate for the asymptotic behavior of $\boldsymbol{v}'\boldsymbol{v}$.

3.3   Derivation of an exact measure of dominance

In what follows we shall derive an exact expression for $Var\left(\overline{x}_{N,t}\right)$ that captures first- and higher-order network interconnections and from which a universal value for $\delta_{\max}$ can be derived. Suppose as before the network contains a total of N units, of which m units are dominant and n units are non-dominant. Here to simplify the exposition we confine our analysis to networks with one dominant unit, that is, $m = 1$ and $n = N - 1$. We provide derivations for networks with multiple dominant units in Appendix C.

Consider the SAR model, (24), and without loss of generality assume that the first element of $\mathbf{x}_t$, namely x_{1t}, is the dominant unit. Accordingly, we write (24) in partitioned form as (setting $w_{11} = 0$)

$$\left(\begin{array}{c} x_{1t} \\ \mathbf{x}_{2t} \end{array}\right) = \left(\begin{array}{cc} 0 & \rho\mathbf{w}'_{12} \\ \rho\mathbf{w}_{21} & \rho\mathbf{W}_{22} \end{array}\right)\left(\begin{array}{c} x_{1t} \\ \mathbf{x}_{2t} \end{array}\right) + \left(\begin{array}{c} g_{1t} \\ \mathbf{g}_{2t} \end{array}\right), \tag{35}$$

where $\mathbf{x}_{2t} = (x_{2t}, x_{3t}, ..., x_{Nt})'$, $\mathbf{w}_{12} = (w_{12}, w_{13}, ..., w_{1N})'$, $\mathbf{w}_{21} = (w_{21}, w_{31}, ..., w_{N1})'$, $\mathbf{W}_{22}$ is the $n \times n$ weight matrix associated with the n non-dominant units, $\mathbf{g}_{2t} = (g_{2t}, g_{3t}, ..., g_{Nt})'$, and

$$g_{it} = -b_i - (1-\rho)(\gamma_i f_t + \varepsilon_{it}), \tag{36}$$

for $i = 1, 2, ..., N$. Furthermore, note that since

$$\mathbf{W}\boldsymbol{\tau}_N = \left(\begin{array}{cc} 0 & \mathbf{w}'_{12} \\ \mathbf{w}_{21} & \mathbf{W}_{22} \end{array}\right)\left(\begin{array}{c} 1 \\ \boldsymbol{\tau}_n \end{array}\right) = \left(\begin{array}{c} 1 \\ \boldsymbol{\tau}_n \end{array}\right),$$

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 ith row sum of $\mathbf{W}_{22}$ is given by $1 - w_{i1} < 1$, 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 matrix $\mathbf{A}$. Under the assumption that $|\rho| < 1$, by Lemma 1 in Appendix A the system of equations (35) has a unique solution given by

$$\begin{aligned} \left(\begin{array}{c} x_{1t} \\ \mathbf{x}_{2t} \end{array}\right) &= \left(\begin{array}{cc} 1 & -\rho\mathbf{w}'_{12} \\ -\rho\mathbf{w}_{21} & \mathbf{I}_n - \rho\mathbf{W}_{22} \end{array}\right)^{-1}\left(\begin{array}{c} g_{1t} \\ \mathbf{g}_{2t} \end{array}\right) \\ &= \mathbf{S}^{-1}(\rho)\mathbf{g}_t. \end{aligned} \tag{37}$$

11

For future reference also note that the $(1,1)^{th}$ element of $\mathbf{S}^{-1}(\rho)$ is given by ϕ_1^{-1}, where

$$\phi_1 = 1 - \rho^2 \mathbf{w}'_{12} \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} \mathbf{w}_{21} \neq 0. \tag{38}$$

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 $0 < \delta_1 \leq 1$. In more general settings where w_{ij} can also take negative values one could use

$$d_1 = \sum_{i=2}^{N} |w_{i1}| = \kappa_1 N^{\delta_1}, \tag{40}$$

which also implies that

$$||\mathbf{w}_{21}||_2 = \left(\sum_{i=2}^{N} w_{i1}^2 \right)^{\frac{1}{2}} \leq d_1^{1/2} = \kappa_1^{1/2} N^{\delta_1/2}. \tag{41}$$

The above exponent formulation is to be contrasted to the power law specification which has been used in the literature. Further discussion is provided in Section 4.

The system of equations (35) can now be solved for $\mathbf{x}_{2t}$ in terms of x_{1t}, namely (note that $|\varrho(\rho \mathbf{W}_{22})| < 1$, since $|\rho| < 1$)

$$\mathbf{x}_{2t} = \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} (\rho \mathbf{w}_{21} x_{1t} + \mathbf{g}_{2t}), \tag{42}$$

and[4]

$$x_{1t} = \phi_1^{-1} \left[g_{1t} + \rho \mathbf{w}'_{12} \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} \mathbf{g}_{2t} \right]. \tag{43}$$

Using the above in (42), we now have

$$\mathbf{x}_{2t} = (\rho/\phi_1) \left[g_{1t} + \rho \mathbf{w}'_{12} \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} \mathbf{g}_{2t} \right] \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} \mathbf{w}_{21} + \left(\mathbf{I}_n - \rho \mathbf{W}_{22} \right)^{-1} \mathbf{g}_{2t}. \tag{44}$$

The first term of $\mathbf{x}_{1t}$ 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.

Consider now the aggregate effects defined as the simple cross-section average of x_{it},

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

Using (42) and (43)

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

---

[4] In deriving (44), it is required that $\phi_1 \neq 0$. 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

$$\overline{x}_{N,t} = -a_N + \theta_N x_{1t} - (1 - \rho)\psi_N f_t - (1 - \rho)\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}, \tag{45}$$

where

$$a_N = N^{-1}\boldsymbol{\tau}_n' (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\mathbf{b}_2,$$

$$\theta_N = N^{-1}\left[1 + \rho\boldsymbol{\tau}_n' (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\mathbf{w}_{21}\right], \tag{46}$$

$$\psi_N = N^{-1}\boldsymbol{\tau}_n' (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\boldsymbol{\gamma}_2, \tag{47}$$

and

$$\mathbf{v}_N' = N^{-1}\boldsymbol{\tau}_n' (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}. \tag{48}$$

The first term of (45), a_N, is an intercept which is bounded in N. To see this, recall that $\|\mathbf{W}_{22}\|_\infty \leq 1$, $|\rho| < 1$, and $\mathbf{W}_{22}$ is column bounded, then it follows from Lemma 2 in Appendix A that $(\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}$ has bounded row and column norms. 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. $\mathbf{v}_N$ is the influence vector associated with the non-dominant units. It is analogous to the influence vector defined by (20) which applies to all units. Taking these terms in turn we have

$$Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) = N^{-2}\boldsymbol{\tau}_n'\mathbf{H}_{22}\boldsymbol{\tau}_n,$$

where $\mathbf{H}_{22} = (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\mathbf{V}_{22,\varepsilon} (\mathbf{I}_n - \rho\mathbf{W}_{22}')^{-1}$ and $\mathbf{V}_{22,\varepsilon} = diag\left(\sigma_2^2, \sigma_3^2, ..., \sigma_N^2\right)$. Notice that

$$\left(\frac{N-1}{N^2}\right)\lambda_n\left(\mathbf{H}_{22}\right) \leq Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) \leq \left(\frac{N-1}{N^2}\right)\lambda_1\left(\mathbf{H}_{22}\right), \tag{49}$$

where $\lambda_1\left(\mathbf{H}_{22}\right)$ and $\lambda_n\left(\mathbf{H}_{22}\right)$ denote the largest and smallest eigenvalue of $\mathbf{H}_{22}$, respectively. We first demonstrate that $\lambda_1\left(\mathbf{H}_{22}\right) < K < \infty$. Since $\sup_i\left(\sigma_i^2\right) < K < \infty$, and we have shown that $(\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}$ has bounded row and column norms, then

$$\|\mathbf{H}_{22}\|_\infty \leq \left\|(\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\right\|_\infty \|\mathbf{V}_{22,\varepsilon}\|_\infty \left\|(\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}\right\|_1 < K < \infty,$$

and likewise $\|\mathbf{H}_{22}\|_1 < K < \infty$. Hence, $\lambda_1\left(\mathbf{H}_{22}\right) \leq \min\left(\|\mathbf{H}_{22}\|_\infty, \|\mathbf{H}_{22}\|_1\right) < K < \infty$. Next, it is easily seen that $\lambda_n\left(\mathbf{H}_{22}\right) > 0$, as $\mathbf{V}_{22,\varepsilon}$ is positive definite and $(\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}$ is nonsingular. Using these results in (49), we establish that $Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right)$ is bounded both above and below by N^{-1} asymptotically, and thus

$$Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) = \ominus\left(N^{-1}\right). \tag{50}$$

Therefore, the effects of the idiosyncratic shocks from the non-dominant sectors on $\overline{x}_{N,t}$ vanish as $N \to \infty$. It is worth noting that we establish the rate of convergence, denoted by $\ominus(\cdot)$, by showing that the upper and lower bounds are identical asymptotically in terms of N, and the symbol $\ominus(\cdot)$ should be distinguished from the $O(\cdot)$ notation, which provides only an upper bound on the growth rate of a function.

Using (43) we also have

$$Cov\left(x_{1t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) = -(1 - \rho)\rho\phi_1^{-1}N^{-1}\mathbf{w}_{12}'\mathbf{H}_{22}\boldsymbol{\tau}_n, \tag{51}$$

and
$$Cov\left(x_{1t}, f_t\right) = -(1 - \rho)\left[\phi_1^{-1}\gamma_1 + \rho\phi_1^{-1}h_2\right]Var\left(f_t\right), \tag{52}$$

where
$$h_2 = \boldsymbol{\gamma}_2'\left(\mathbf{I}_n - \rho\mathbf{W}_{22}\right)^{-1}\mathbf{w}_{12}.$$

Overall (recalling that f_t and ε_{it} are independently distributed), we have

$$
\begin{aligned}
Var\left(\overline{x}_{N,t}\right) = \ & \theta_N^2 Var\left(x_{1t}\right) - 2(1 - \rho)\theta_N Cov\left(x_{1t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) + (1 - \rho)^2 Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) \\
& + (1 - \rho)^2 q_N Var\left(f_t\right),
\end{aligned} \tag{53}
$$

where
$$q_N = \psi_N^2 + 2\psi_N\theta_N\phi_1^{-1}\gamma_1 + 2\rho\psi_N\theta_N\phi_1^{-1}h_2.$$

Also, using (43) we have

$$
\begin{aligned}
Var\left(x_{1t}\right) = \ & \phi_1^{-2}(1 - \rho)^2\left\{\left[\gamma_1^2 + \rho^2 h_2^2\right]Var(f_t) + \sigma_1^2\right\} \\
& + \phi_1^{-2}\rho^2(1 - \rho)^2\mathbf{w}_{12}'\mathbf{H}_{22}\mathbf{w}_{21},
\end{aligned} \tag{54}
$$

which is easily seen to be bounded in N.

A number of results can now be obtained from (53). 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 as $N \to \infty$.

In the case where there is no common factor but the network includes a dominant unit, using (53) and (50) we have

$$Var\left(\overline{x}_{N,t}\right) = \theta_N^2 Var\left(x_{1t}\right) - 2(1 - \rho)\theta_N Cov\left(x_{1t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) + O\left(N^{-1}\right). \tag{55}$$

Recall that $Var(x_{1t})$ is bounded in N, and consider the limiting properties of θ_N defined by (46). We first note that $\left(\mathbf{I}_n - \rho\mathbf{W}_{22}\right)^{-1}$ is row and column bounded, and $\boldsymbol{\phi}_n' = \boldsymbol{\tau}_n'\left(\mathbf{I}_n - \rho\mathbf{W}_{22}\right)^{-1}$, being the $n \times 1$ vector of column sums of $\left(\mathbf{I}_n - \rho\mathbf{W}_{22}\right)^{-1}$, will also have bounded elements. Furthermore, since

$$\boldsymbol{\phi}_n' = (\phi_2, \phi_3, ..., \phi_N) = \boldsymbol{\tau}_n' + \rho\boldsymbol{\tau}_n'\mathbf{W}_{22} + \rho^2\boldsymbol{\tau}_n'\mathbf{W}_{22}^2 + ...,$$

recalling that $\rho > 0$ and $w_{ij} \geq 0$, then $\phi_i > 1$ for $i = 2, 3, ..., N$, and

$$N^{-1} + \phi_{\min}\rho N^{-1}d_1 \leq \theta_N \leq N^{-1} + \phi_{\max}\rho N^{-1}d_1, \tag{56}$$

where $\phi_{\min} = \min(\phi_2, \phi_3, ..., \phi_N) > 1$, $\phi_{\max} = \max(\phi_2, \phi_3, ..., \phi_N) < K < \infty$, and $\boldsymbol{\tau}_n'\mathbf{w}_{21} = \sum_{j=2}^{N}w_{j1} = d_1$. Therefore, the asymptotic behavior of θ_N depends on the way the outdegree of the dominant unit, namely d_1, varies with N. Using the specification as before in (27), $d_1 = \kappa_1 N^{\delta_1}$, it follows that

$$N^{-1} + \phi_{\min}\rho\kappa_1 N^{\delta_1 - 1} \leq \theta_N \leq N^{-1} + \phi_{\max}\rho\kappa_1 N^{\delta_1 - 1}, \tag{57}$$

which leads to

$$\theta_N = \ominus(N^{\delta_1 - 1}), \ 0 < \delta_1 \leq 1. \tag{58}$$

Finally, consider the second term of (55) and note from (51) that

$$\left| Cov\left(x_{1t}, \mathbf{v}'_n \boldsymbol{\varepsilon}_{2t}\right) \right| \leq \left| \frac{(1-\rho)\rho}{\phi_1} \right| N^{-1} \left\| \mathbf{w}'_{12} \right\|_\infty \left\| (\mathbf{I}_n - \rho \mathbf{W}_{22})^{-1} \right\|_\infty \left\| \mathbf{V}_{22,\varepsilon} \right\|_\infty \left\| \boldsymbol{\phi}_n \right\|_\infty = O\left(N^{-1}\right),$$

since $\left\| \mathbf{w}'_{12} \right\|_\infty = \left\| \mathbf{w}_{12} \right\|_1 = \sum_{i=2}^{N} w_{1i} = 1$, $\left\| (\mathbf{I}_n - \rho \mathbf{W}_{22})^{-1} \right\|_\infty < K$, $\left\| \mathbf{V}_{22,\varepsilon} \right\|_\infty = \bar{\sigma}^2 < K$, and $\left\| \boldsymbol{\phi}_n \right\|_\infty = \phi_{\max} < K$. Using the above results in (55) we now have

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

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

$$Var\left(\overline{x}_{N,t}\right) = \ominus(N^{2\delta_1 - 2}) + \ominus\left(N^{-1}\right), \tag{59}$$

and clearly,

$$Var\left(\overline{x}_{N,t}\right) = \ominus(N^{2\delta_1 - 2}), \text{ if } \delta_1 \geq 1/2. \tag{60}$$

As noted earlier, this result 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 θ_N to the first- and higher-order network connections discussed in Acemoglu et al. (2012). Expanding the terms of the inverse $(\mathbf{I}_n - \rho \mathbf{W}_{22})^{-1}$, θ_N can also be written as

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

where $N^{-1} \rho \boldsymbol{\tau}'_n \mathbf{w}_{21} = N^{-1} \rho d_1$ 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 (56) and (58) all these higher order interconnections (individually and together) at most behave as $O(N^{\delta_1 - 1})$.

Therefore, the rate at which micro shocks influence the macro economy depends on δ_1, which measures the strength of the dominant unit. But to ensure a non-vanishing variance, $Var\left(\overline{x}_{N,t}\right) > 0$, we need a value of $\delta_1 = 1$. For all other values of $\delta_1 < 1$, $Var\left(\overline{x}_{N,t}\right) \to 0$, as $N \to \infty$. Also 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. *It should be noted from (59) that δ_1 can not be consistently estimated if $\delta_1 < 1/2$. This finding 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

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

15

and the rate of convergence of $\overline{x}_{N,t}$ is determined by the strength of the factor as given by ψ_N^2. Using (47) we have (recall that $\varrho(\mathbf{W}_{22}) \leq 1$),

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

By a similar line of reasoning as before, it is then easily seen that $\psi_N = \ominus\left(N^{\delta_\gamma - 1}\right)$, 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 the dominant unit. Then for $\delta_1 > 1/2$ and $\delta_\gamma > 1/2$ we have[5]

$$
Var\left(\overline{x}_{N,t}\right) = \ominus\left(N^{2\delta_1 - 2}\right) + \ominus\left(N^{2\delta_\gamma - 2}\right) + \ominus\left(N^{-1}\right). \tag{61}
$$

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 (61) 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 the dominant units, respectively.

4 Degrees of pervasiveness of units in a network: estimation and inference

In this section we consider the problem of estimating the degrees of pervasiveness 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 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 allows identification of more than one dominant units in the network.

4.1 Power law type estimators

Suppose that we have observations on the outdegrees, d_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ...T$. Denote the degree of dominance of unit i by δ_i, and the associated ordered values by $\delta_{(i)}$, where $\delta_{(1)} \geq \delta_{(2)} \geq ... \geq \delta_{(N)}$. The power law estimate of $\delta_{(1)}$ is given by $1/\hat{\beta}$, where $\hat{\beta}$ is an estimator of the shape parameter of the power law distribution

---

[5]Note that for values of δ_1 and $\delta_\gamma < 1/2$ the third term in (61) will dominate the network and the common factor effects.

fitted to the outdegrees that lie above a given minimum cut-off value, $d_{\min}$. (see Section 3.2). A random variable D is said to follow a power law distribution if its complementary cumulative density function (CCDF), or tail distribution, satisfies

$$\Pr\left(D \geq d\right) \propto d^{-\beta}, \tag{62}$$

where $\beta > 0$ is a constant known as the shape parameter, or the exponent, of the power law, and $\propto$ denotes asymptotic equivalence.[6] 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

$$f_D\left(d\right) \propto d^{-(\beta+1)}. \tag{63}$$

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

$$\Pr\left(D \geq d\right) = \left(d/d_{\min}\right)^{-\beta}, \, d \geq d_{\min}, \tag{64}$$

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.[7]

The first popular procedure is to run the following log-log regression (also known as the rank-size regression or Zipf regression),

$$\ln i = a - \beta \ln d_{(i)}, \, i = 1, 2, ..., N_{\min}, \tag{65}$$

where a is a constant, and $d_{(1)} \geq d_{(2)} \geq ... \geq d_{(N_{\min})}$ are the $N_{\min}$ largest ordered observations on d such that $d_{(N_{\min})} \geq d_{\min}$. d in our case are the outdegrees of units in a network. It can be shown that the ordinary least squares (OLS) estimator of β from regression (65) converges in probability to the true β if $N_{\min}$ is sufficiently large. This procedure, however, is fraught with two pitfalls. First, $d_{\min}$, the lower bound above which D obeys the Pareto distribution, is rarely known in practice, and hence the choice of $N_{\min}$, i.e., the number of cut-off observations used in the regression, is often made by graphical inspection or decided arbitrarily by the investigator, which may lead to unreliable results. Second, both the estimate of β and its standard error are subject to small sample bias (Goldstein, Morris, and Yen, 2004; Gabaix and Ioannides, 2004). Thus, to correct this bias, Gabaix and Ibragimov (2011) advocate shifting the rank by $1/2$ and estimating β using the following regression

$$\ln\left(i - 1/2\right) = a - \beta \ln d_{(i)}, \, i = 1, 2, ..., N_{\min}. \tag{66}$$

---

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

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

Hereafter, we refer to the estimator of β from (66) as $\hat{\beta}_{LL}$, or the log-log regression estimator with Gabaix and Ibragimov (2011) correction. The standard error of $\hat{\beta}_{LL}$ can be estimated by

$$\hat{\sigma}\left(\hat{\beta}_{LL}\right) = \sqrt{\frac{2}{N_{\min}}}\hat{\beta}_{LL}. \tag{67}$$

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[8]

$$\hat{\beta}_{MLE} = \frac{N_{\min}}{\sum_{i=1}^{N_{\min}} \ln d_{(i)} - N_{\min} \ln d_{(N_{\min})}}, \tag{68}$$

and its standard error can be estimated by

$$\hat{\sigma}\left(\hat{\beta}_{MLE}\right) = \frac{\hat{\beta}_{MLE}}{\sqrt{N_{\min}}}, \tag{69}$$

if $N_{\min}$ is large. This estimator would be the most efficient one if the underlying distribution is indeed Pareto and if $d_{\min}$ is known. Otherwise, it may be less robust than the log-log regression, and also suffers from small sample bias and the same problem of sensitivity to the choice of $N_{\min}$.

Rather than estimating β based on some arbitrarily chosen value of $N_{\min}$, some authors have proposed joint estimation of β and $d_{\min}$. Among them a notable contribution is by Clauset, Shalzi and Newman (2009, CSN), who demonstrate the importance of $\hat{d}_{\min}$ (the estimated $d_{\min}$) by showing that $\hat{\beta}_{MLE}$ greatly underestimates β if $\hat{d}_{\min} < d_{\min}$ and slightly overestimates β if $\hat{d}_{\min} > d_{\min}$. They 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 $d \geq d_{\min}$ and is obtained by the maximum likelihood estimation. The MLE in (68) is then computed using the estimated value of $d_{\min}$. Hereafter, we call this estimator feasible maximum likelihood estimator and denote it by $\hat{\beta}_{CSN}$.[9] It should be noted that this method relies on the assumption that the true distribution above $d_{\min}$ is exactly Pareto. Other goodness-of-fit measures can also be considered. (See Clauset, Shalzi and Newman, 2009, and references therein.)

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 $\delta_{(1)}$, and how these estimates compare to the extremum estimator that we now consider.

---

[8] See, for example, Appendix B of Newman (2005).
[9] The code implementing this method can be downloaded from http://tuvalu.santafe.edu/~aaronc/powerlaws/.

4.2 Extremum type estimators

In this sub-section we propose an extremum estimator of $\delta_{(1)}$ (or $1/\beta$) which is applicable more generally as compared to the power law type estimators. Specifically, it applies to the general specification of the outdegrees given by (27). We begin with the following balanced panel data model

$$d_{it} = \kappa N^{\delta_i} \exp(\upsilon_{it}), \; i = 1, 2, ..., N; \; t = 1, 2, ..., T, \tag{70}$$

where d_{it} is the outdegree of unit i measured at time t, $\kappa > 0$, and δ_i measures the degree of dominance of unit i. We also assume that υ_{it} are $IID(0, \sigma_v^2)$, and distributed independently of δ_i, which we take as fixed constants. Furthermore, given the constraint (28), we must have

$$\kappa \sum_{i=1}^{N} N^{\delta_i} \exp(\upsilon_{it}) = N,$$

which for N sufficiently large yields (noting that υ_{it} and δ_i are by assumption independently distributed)[10]

$$\kappa = \frac{exp\left(-\frac{\sigma_v^2}{2}\right)}{\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} N^{\delta_i}} > 0, \tag{71}$$

Taking the log transformation of (70) we have

$$\ln d_{it} = \ln \kappa + \delta_i \ln N + \upsilon_{it}, \; i = 1, 2, ..., N; \; t = 1, 2, ..., T. \tag{72}$$

Averaging across i and t now yields

$$(TN)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \ln d_{it} = \ln \kappa + \left(N^{-1} \sum_{i=1}^{N} \delta_i\right) \ln N + (TN)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \upsilon_{it}. \tag{73}$$

But from the summability condition, (31), 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

$$\left(N^{-1} \sum_{i=1}^{N} \delta_i\right) \ln N \to 0, \text{ as } N \to \infty. \tag{74}$$

This result holds irrespective of whether the number of δ_i's that are non-zero, say m, is fixed or rising with N, but at a slower rate than N (Propositions 1 and 2).

Furthermore, considering that by assumption υ_{it} are IID over i and t, then the last term of (73) also tends to zero, and $\ln \kappa$ can be estimated by

$$\widehat{\ln \kappa} = (TN)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \ln d_{it}. \tag{75}$$

---

[10]Recall that under the summability condition (31), $\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} N^{\delta_i}$ exists and is strictly positive.

19

Now for each i, averaging (72) over $t = 1, 2, ..., T$, we can propose the following estimator of δ_i:

$$\hat{\delta}_i = \frac{T^{-1} \sum_{t=1}^{T} \ln d_{it} - (TN)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \ln d_{it}}{\ln N}. \tag{76}$$

The largest of these estimates, which we refer as the *extremum* estimator, can be used as the estimator of the degree of pervasiveness of the most dominant unit in the network, the second largest as the estimator of the degree of pervasiveness of the second most dominant unit and so on. Recall that only estimates of $\hat{\delta}_i > 1/2$ should be considered, since any unit with an estimate of δ below $1/2$ will not have any network effects (see Remark 3).

It is worth noting that in the pure cross section case with $T = 1$, the extremum estimator reduces to

$$\hat{\delta}_{(1)} = \frac{\ln d_{(1)} - N^{-1} \sum_{i=1}^{N} \ln d_i}{\ln N}, \tag{77}$$

where $d_{(1)}$ denotes the largest value of $d_i > 0$. Recall that $\delta_{(1)}$ is the inverse of the shape parameter of the Pareto distribution, β, when N becomes large. The inverse of $\hat{\delta}_{(1)}$ serves as an extremum type estimator of β, since it relies on the largest d_i, which is adjusted by the sample mean. We will analyze how well $\hat{\delta}_{(1)}$ estimates $1/\beta$ in the next section. Also note that our approach allows us to estimate $\delta_{(2)}, \delta_{(3)}, ...\delta_{(M)}$ for an a priori given value of M, such that $\delta_{(M)} > 1/2$. We refer to $\hat{\delta}_{(i)}$, for $i = 1, 2, ..., M$, as the extrema estimators.

To investigate the asymptotic properties of the extrema estimators, we first note that under (72), $\widehat{\ln \kappa}$ and $\hat{\delta}_i$ can be written as

$$\widehat{\ln \kappa} - \ln \kappa = \bar{\delta} \ln N + \bar{v}, \tag{78}$$

$$\hat{\delta}_i - \delta_i = \bar{\delta} + \frac{\bar{v}_i - \bar{v}}{\ln N}, \tag{79}$$

where $\bar{\delta} = N^{-1} \sum_{i=1}^{N} \delta_i$, and

$$\bar{v}_i = T^{-1} \sum_{t=1}^{T} v_{it}, \quad \bar{v} = N^{-1} \sum_{i=1}^{N} \bar{v}_i.$$

It is now easily seen that

$$Cov(\hat{\delta}_i, \hat{\delta}_j) = -\frac{1}{(\ln N)^2 N} \frac{\sigma_v^2}{T}, \text{ for all } i \neq j,$$

$$Var\left(\hat{\delta}_i\right) = \frac{\sigma_v^2}{(\ln N)^2 T} \left(1 - \frac{1}{N}\right). \tag{80}$$

Note that for any given i, $Var\left(\hat{\delta}_i\right) \to 0$ for a fixed T as $N \to \infty$, and if both N and $T \to \infty$. But when T is fixed, the rate of convergence of $Var\left(\hat{\delta}_i\right)$ is given by $(\ln N)^{-2}$ and will be rather slow. Also, it is already established that $\bar{\delta} \to 0$, as $N \to \infty$, and as a result $\hat{\delta}_i \to_p \delta_i$, which establishes that for any finite $T \geq 1$, $\hat{\delta}_i$ is a $\ln N$ consistent estimator of $\hat{\delta}_i$. The estimators of δ for two different units (i, j), are asymptotically independent and their covariance converges to zero at a fast rate of $(\ln N)^{-2} N^{-1}$, even if T is fixed.

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

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

Ignoring lower order terms in N we now have

$$\frac{(\ln N)\sqrt{T}\left(\hat{\delta}_i - \delta_i - \bar{\delta}\right)}{\sigma_v} \to_d N(0,1), \text{ as } N \to \infty.$$

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

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

But given the summability condition, (31), for the test to be free of the nuisance parameter it is sufficient that the following condition on the relative expansion rates of N and T holds

$$(\ln N)\, N^{-1}\sqrt{T} \to 0, \tag{81}$$

as $N, T \to \infty$, jointly. It is clear that this condition is met if T is fixed as $N \to \infty$. However, when T takes moderate to large values, N needs to be sufficiently large relative to T. It is clear that N and T can rise at the same rate. But by setting $T = O\left(N^\phi\right)$, it also follows that condition (81) will be satisfied so long as $\phi < 2$, which allows T to rise faster than N.

To obtain a consistent estimator of σ_v^2, using (75) and (76) in (72) we have

$$
\begin{aligned}
\hat{v}_{it} &= \ln d_{it} - \widehat{\ln \kappa} - \hat{\delta}_i \ln N \tag{82}\\
&= -\left(\widehat{\ln \kappa} - \ln \kappa\right) - \left(\hat{\delta}_i - \delta_i\right)\ln N + v_{it}.
\end{aligned}
$$

Now using (78) and (79)

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

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

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2}{N(T-1)}. \tag{83}$$

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

$$\mathfrak{D}_{(1)} = \frac{(\ln N)\left(\hat{\delta}_{(1)} - \delta_{(1)}^0\right)}{\hat{\sigma}_v\left(\frac{1}{T} - \frac{1}{TN}\right)^{1/2}}, \tag{84}$$

21

where $\hat{\delta}_{(1)} = \max_i \left(\hat{\delta}_i \right)$. It then follows that $\mathfrak{D}_{(1)} \to_d N(0,1)$ as $N \to \infty$, for a fixed $T > 1$.

The test is also valid in cases where N and T are both large, if $\ln(N) \sqrt{T}/N \to 0$, as N and $T \to \infty$. When $T = 1$, δ_i can still be consistently estimated by (76), but to obtain its distribution we need to impose further restrictions on δ_i over i.

4.3 Misspecification analysis

The exponent specification of the outdegrees given in (70) is closely related to (27) in that $\kappa_j = \kappa \exp(\upsilon_{it}) > 0$, and it 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 type specification. If the observations on outdegrees indeed follow a power law tail distribution with the shape parameter, β, we can show that our proposed extremum estimator, $\hat{\delta}_{(1)}$, is still consistent for $\delta_{(1)} = 1/\beta$ in this case. More importantly, the extremum estimator of β (given by $1/\hat{\delta}_{(1)}$), does not require specification of a cut-off value $d_{\min}$, and is robust to the distribution of d_{it} below $d_{\min}$.

To see this, suppose that the observations on the outdegrees, $d_{it} \geq 0$, for $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, are independent draws from the following distribution

$$
\begin{aligned}
f(d_{it}) &\propto d_{it}^{-1-\beta}, \text{ if } d_{it} \geq d_{\min}, \\
&\propto \psi(d_{it}), \text{ if } d_{it} < d_{\min},
\end{aligned}
\tag{85}
$$

where $d_{it} > 0$ and follows a Pareto distribution with the shape parameter β for values of d_{it} above $d_{\min}$, and an arbitrary non-Pareto distribution, $\psi(d_{it})$, for values of d_{it} 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 $d_{\min}$.

In what follows we focus on the pure cross section case where $T = 1$ and consider the extremum estimator, $\hat{\delta}_{(1)}$, which can be written as

$$
\hat{\delta}_{(1)} = \frac{\ln d_{(1)} - N^{-1} \sum_{i=1}^N \ln d_i}{\ln N} = \frac{z_{(1)} - N^{-1} \sum_{i=1}^N z_i}{\ln N},
\tag{86}
$$

where $z_i = \ln(d_i/d_{\min})$, for all i, and $z_{(i)} = \ln(d_{(i)}/d_{\min})$, with $d_{(i)}$ being the ith largest value of $d_i > 0$. 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 $E(z|z \geq 0) = 1/\beta$, and $Var(z|z \geq 0) = 1/\beta^2$, for $\beta > 0$.[11] We also assume that $E(z_i|z_i < 0)$ exists, which is a mild moment condition to impose on $\psi(d_i)$ for

---

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

$\ln(d_i/d_{\min}) < 0$. The following proposition summarizes the consistency property of $\hat{\delta}_{(1)}$ 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 (85) 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*

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

where $\hat{\delta}_{(1)}$ is defined by (86).

A proof is provided in Appendix D.

Remark 4. *The convergence of $\hat{\delta}_{(1)}$ 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 $d_{\min}$ or the shape of $\psi(d)$, the non-Pareto part of the distribution.*

The above analysis can be readily extended to the case where $T > 1$, but it should be emphasized that when $T > 1$, one should order $d_{i,t}$ for each period t and then add up $\ln d_{(i),t}$ across t, that is,

$$\hat{\delta}_{(1)} = \frac{T^{-1}\sum_{t=1}^{T} \ln d_{(1),t} - (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} \ln d_{(i),t}}{\ln N}. \tag{87}$$

Notice that

$$\widehat{\ln \kappa} = (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} \ln d_{(i),t} = (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} \ln d_{it}.$$

Assuming the mean of $\ln d_{it}$ exists and is the same for all i and t, we have $\widehat{\ln \kappa} \to_p \mu_d$ for any fixed T as $N \to \infty$, or if N and $T \to \infty$, jointly.

Although the identities of the dominant units may change over time, $\hat{\delta}_{(1)}$ given by (87) provides a systematic measure of the degree of pervasiveness of the network. If, on the other hand, one is interested in identifying the dominant units, then estimating $\delta_{(i)}$ for each period t would be more suitable. The asymptotic distribution of $\hat{\delta}_{(1)}$ when the underlying distribution takes the form of (85) is left for future research, but we conjecture that it can be developed when N and T are large.

Conversely, if the true distribution of the outdegrees follows the exponent formulation given in (70), but a researcher assumes a power law tail distribution and applies the strategies we discussed in Section 4.1, then the estimators of $\delta_{(1)}$ based on the inverse of the estimated shape parameter β could be strongly biased. Monte Carlo evidence is provided in the next section to support this claim.

To sum up, compared to the power law type 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 distribution assumption 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.

4.4 Extension of the extremum type estimators to 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 that 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

$$\hat{\delta}_i = \frac{T_i^{-1} \sum_{t=T_i^0}^{T_i^1} \ln d_{it} - N^{-1} \sum_{i=1}^{N} \left(T_i^{-1} \sum_{t=T_i^0}^{T_i^1} \ln d_{it} \right)}{\ln N}, \tag{88}$$

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

$$Var\left(\hat{\delta}_i \right) = \frac{\sigma_v^2}{(\ln N)^2} \left(\frac{1}{T_i} - \frac{1}{NT_i} \right). \tag{89}$$

The estimator of the most dominant unit is given by $\hat{\delta}_{(1)} = \max_i \left(\hat{\delta}_i \right)$, and as in the balanced panel case, the asymptotic distribution of $\hat{\delta}_{(1)}$ is given by

$$\mathfrak{D}_{(1)} = \frac{(\ln N)\left(\hat{\delta}_{(1)} - \delta_{(1)}^0 \right)}{\hat{\sigma}_v \left(\frac{1}{T_{(1)}} - \frac{1}{NT_{(1)}} \right)^{1/2}}, \tag{90}$$

where $T_{(1)}$ refers to the sample size of the most dominant unit, and

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N} (T_i - 1)^{-1} \sum_{t=T_i^0}^{T_i^1} \hat{v}_{it}^2}{N}. \tag{91}$$

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

5   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. We also investigate the small sample properties of the extremum estimator for unbalanced panels. We consider two types of data generating processes (DGPs) for the outdegrees (d_{it}): a power law and an exponent specification.

The DGP for the exponent specification is given by

$$\ln d_{it} = \ln \kappa + \delta_i \ln N + v_{it}, \ i = 1, 2, ..., N; \ t = 1, 2, ..., T, \tag{92}$$

where v_{it}, the idiosyncratic errors, are generated as $v_{it} \sim IIDN(0, 1)$, with κ set as

$$\kappa = \frac{exp\left(-\frac{1}{2}\right)}{N^{-1} \sum_{i=1}^{N} N^{\delta_i}} > 0, \tag{93}$$

to ensure that d_{it} add up to N across i for each t. The following two sets of experiments are carried out under this DGP assuming balanced panels.

Experiment A. In this experiment we consider networks with a finite number of dominant units, and a large number of non-dominant units. Specifically, we consider

- A.1. One strongly dominant unit: $\delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$.

- A.2. Two strongly dominant units: $\delta_{(1)} = \delta_{(2)} = 1$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$.

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

Experiment B. In this experiment we allow all units to be weakly dominant, and ensure $\bar{\delta} \to 0$ at a sufficiently fast rate by assuming that individual δ_i decays exponentially. In particular we consider $\delta_{(i)} = 0.9^i$, for $i = 1, 2, ..., N$.[12]

All experiments are replicated $2,000$ times for all combinations of $N = 100, 300, 500, 1,000$, and $T = 1, 2, 4, 6, 8, 10, 20, 50, 100$. We also provide simulation results for a very large data set with $N = 450,000$, which can arise when using inter-firm level sales data.[13] The M largest estimates of δ considered by the investigator are denoted by $\hat{\delta}_{(i)}$, for $i = 1, 2, ..., M$, which are calculated according to (76). When $T > 1$, the variance of $\hat{\delta}_{(i)}$ is computed by (80), with σ_v^2 estimated by (83).

Furthermore, as shown in Section 3.2, the largest value of δ in a network is related to the inverse of β, the shape parameter of the Pareto distribution considered by Acemoglu et al. (2012). We exploit this relationship and compare the inverse of the extremum estimator of $\delta_{(1)}$ with the estimator of β based on the Pareto distribution. Specifically,

---

[12]Note that for all the above experiments the denominator of (93), $N^{-1} \sum_{i=1}^{N} N^{\delta_i}$, converges to a finite positive constant. Also see Appendix B.

[13]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.

we report the average estimates of β and the inverse of $\hat{\delta}_{(1)}$, across $2,000$ replications for Experiment A.1.

The second type of DGP that we consider takes the form of a power law tail distribution.[14] In the first step, random variable y_{it} is drawn from the following distribution that obeys an exact Pareto distribution above $y_{\min}$ and an exponential distribution below $y_{\min}$:

$$f(y_{it}) = \begin{cases} C_t(y_{it}/y_{\min,t})^{-(\beta+1)}, & \text{for } y_{it} \geq y_{\min,t} \\ \\ C_t e^{-(\beta+1)(y_{it}/y_{\min,t}-1)}, & \text{for } y_{it} < y_{\min,t} \end{cases}, \tag{94}$$

for $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, where C_t is given by

$$C_t = \left[\frac{y_{\min,t}\left(e^{\beta+1} - 1\right)}{\beta + 1} + \frac{y_{\min,t}}{\beta} \right]^{-1}, \tag{95}$$

which ensures that $f(y_{it})$ integrates to 1 over its full support, $y_{it} > 0$. Then, observations on d_{it} are generated as multiples of y_{it},

$$d_{it} = \frac{N}{\sum_{i=1}^{N} y_{it}} y_{it}, \tag{96}$$

so that the outdegrees add up to N. We denote the lower bound of the Pareto distribution for d_{it} as $d_{\min,t}$, which is given by

$$d_{\min,t} = \frac{N}{\sum_{i=1}^{N} y_{it}} y_{\min,t}. \tag{97}$$

It is worth noting that under the distribution given by (94), the probability that d_{it} is greater than or equal to $d_{\min,t}$ is given by

$$\Pr\left(d_{it} \geq d_{\min,t}\right) = \Pr\left(y_{it} \geq y_{\min,t}\right) = \frac{1}{\beta}\left(\frac{e^{\beta+1} - 1}{\beta + 1} + \frac{1}{\beta}\right)^{-1}, \tag{98}$$

which is time-invariant and depends only on the value of β.

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.

The inverse transformation sampling method is used to generate y_{it} such that its distribution satisfies (94). To this end we first generate u_{it} as $IIDU[0, 1]$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, and set

$$u_{\min,t} = C_t \left(\frac{y_{\min,t}}{\beta + 1}\right)\left(e^{\beta+1} - 1\right), \tag{99}$$

and then generate y_{it} as

$$y_{it} = \begin{cases} -\dfrac{y_{\min,t}}{\beta + 1} \ln\left[1 - \dfrac{(\beta + 1)u_{it}}{C_t e^{\beta+1} y_{\min,t}}\right], & \text{if } u_{it} < u_{\min,t} \\ \\ \left[\dfrac{\beta\left(y_{\min,t} - u_{it}\right) + C_t y_{\min,t}}{C_t y_{\min,t}^{\beta+1}}\right]^{-1/\beta}, & \text{if } u_{it} \geq u_{\min,t} \end{cases}. \tag{100}$$

---

[14]This DGP follows (3.10) in Clauset et al. (2009).

Under the Pareto tail DGP given by (94), we carry out the following experiment.

Experiment C. We consider four values of β, $\beta = 1.0, 1.1, 1.3, 1.5$, and set $y_{\min,t} = y_{\min} = 15$. The sample sizes are combinations of $N = 100, 300, 500, 1,000, 450,000$ and $T = 1, 4$, and the number of replications for each experiment is $R = 2,000$. We assess the performance of four estimators of β in this experiment: the log-log regression estimator $(\hat{\beta}_{LL})$ according to (66) for different given cut-off values, $d_{\min,t}$, the maximum likelihood estimator $(\hat{\beta}_{MLE})$ given by (68) for different $d_{\min,t}$, the CSN estimator $(\hat{\beta}_{CSN})$ which estimates β jointly with the cut-off value, and the inverse of $\hat{\delta}_{(1)}$, the extremum estimator, given by (76).

It is also of interest to see how the estimators of the shape parameter of the Pareto distribution perform compared to the extremum estimator when the DGP follows the exponent representation given by (92). For this purpose, we also apply the four estimation methods just mentioned to the exponent DGP under the setup of Experiment A.1 and Experiment B. (In Experiment B, we also consider $\delta_{(i)} = 0.75^i$, for $i = 1, 2, ..., N$.)

Finally, we investigate the small sample properties of the extremum estimates of $\delta_{(1)}$ when the panel data under consideration is unbalanced. We focus on Experiment A and generate an unbalanced panel where T_i (the number of time series observations for unit i) lies between 2 and 4. To ensure that the most important dominant units are present across the years, only units in the bottom 95th percentile of the distribution of δ were subject to missing observations. In the case of these units, we dropped the first and the last observations with a 50% probability. This randomization process is repeated for all the $2,000$ replications. The estimates of δ_i are computed using (88), and their variances (when $T > 1$) using (89).

5.1 MC results

The results for Experiments A.1 to A.3 are summarized in Tables 1 and 2, in which bias $(\times 100)$ and root mean squared error $(\text{RMSE}(\times 100))$, as well as size $(\times 100)$ and power $(\times 100)$ are reported. We first note that the bias and RMSE of the estimators decline as N and/or T rises. The bias and RMSE reduction is particularly pronounced as T rises, even by a few time periods. This is in line with the theoretical derivations which establish that along the cross-sectional dimension the rate of convergent 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 $\hat{\delta}_{(1)}$ 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 $\delta_{(1)} = 0.90$) 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.

Similar results are obtained for Experiment B (Table 3). Recall that in the case of this experiment, the degrees of dominance are assumed to decay exponentially across the N units. This table provides results for the four largest values of δ, namely $0.9, 0.9^2, 0.9^3$, and 0.9^4. For other values of $\delta_{(i)}$, for $i = 5, 6, ..., N$, the estimates fall below $1/2$ and have no consequence for the shock diffusion within the network. The results in Table 3 confirm the validity of our theoretical derivations for the case where there are many dominant units whose degrees of dominance decay exponentially.

Turning now to the results in Table 4, we find that the most dominant units can almost always be correctly identified when there are a finite number of dominant units, especially when $T > 2$. By contrast, the results in Table 5 show that the probability of correct identification is lower when the network consists of a large number of weakly dominant units whose degree of pervasiveness decay exponentially. As expected, the more clustered are the degrees of pervasiveness across units, the more difficult it is to differentiate one unit from another.

Our theoretical results suggest that β, the shape parameter of the Pareto distribution, should be close to the inverse of $\delta_{(1)}$. This is supported by the simulation results summarized in Table 6 when N is not too large. Notice that as the network expands along the cross-sectional dimension, $\hat{\beta}_{LL}$, which is obtained by the log-log regression (66) assuming a 20% cut-off value ($N_{\min}/N = 20\%$), deviates more and more from its true value, and also from the inverse of $\hat{\delta}_{(1)}$. In particular, when $N = 450,000$, the estimate of β is 2.11 (with a standard error of 0.01), which far exceeds the true value of 1. By contrast, the inverse of $\hat{\delta}_{(1)}$ equals 1.00 for $T \geq 2$.

Tables 7 to 10 summarize the results for Experiment C where the DGP exhibits the Pareto tail behavior given by (85). 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 Section 4.3. It is also worth noting that the log-log regression estimator ($\hat{\beta}_{LL}$) and the MLE ($\hat{\beta}_{MLE}$) are very sensitive to the choice of the cut-off values.[15] The feasible MLE, $\hat{\beta}_{CSN}$, performs better, but this could be partly explained by the fact that it assumes a known distribution.[16]

However, when the DGP is given by (92), the exponent model, $\hat{\beta}_{LL}$, $\hat{\beta}_{MLE}$, and $\hat{\beta}_{CSN}$ estimators all show severe biases particularly for N large (see Tables 11 to 13). For instance, the simulation mean of $\hat{\beta}_{LL}$ is 2.39 assuming a 10% cut-off value when $N = 450,000$ for Experiment A.1, as compared to its true value of 1. The maximum likelihood estimates and the feasible MLE suffer from similar severe biases.

Finally, in the case of unbalanced panels, it can be seen from Table 14 that the extremum estimator continues to perform well, and is reasonably robust to alternative network structures under different specifications of the distribution of outdegrees. But in the case of unbalanced panels, we need to assume that the outdegrees of the units with the highest degrees of dominance are observed for at least two periods.

---

[15] 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).

[16] It is important to note that the validity of the feasible MLE procedure proposed by Clauset et al. (2009) critically depends on how close is the assumed specification of the distribution of d_{it} below $d_{\min,t}$ to the true underlying distribution.

6   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 $\delta_{(1)}$) obtained by Acemoglu et al. (2012) for the most dominant sector. We provide estimates based on the US input-ouput 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 $\hat{\delta}_{(1)} = \max_i(\hat{\delta}_i)$, where $\hat{\delta}_i$ is computed using (76). 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 $\hat{\beta}_{LL}$ assuming a 20% cut-off value (the percentage above which the degree sequences are assumed to follow the Pareto distribution). The estimates $\hat{\delta}_{(1)}$ and the inverse of $\hat{\beta}$ for the years 1972, 1977, 1982, 1987, 1992, 1997 and 2002 are given in Tables 15 and 16. For the inverse of $\hat{\beta}$, Tables 15 and 16 report estimates based on the first-order and second-order interconnections, respectively.[17] We estimate β by the three approaches discussed above, namely the log-log regression with Gabaix and Ibragimov (2011) correction ($\hat{\beta}_{LL}$) given by (66), the MLE ($\hat{\beta}_{MLE}$) given by (68), and the feasible MLE ($\hat{\beta}_{CSN}$) estimator. For the log-log 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.[18]

The results in Tables 15 and 16 show that the yearly estimates of $\hat{\delta}_{(1)}$ 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 $\hat{\delta}_{(1)}$ to be close to unity. In contrast, the estimates of $\delta_{(1)}$ 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 $\hat{\beta}_{LL}$, 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 $\hat{\beta}_{LL}$ and

---

[17]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 and $\mathbf{w}_{\cdot j}$ is the jth column of $\mathbf{W}$.

[18]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.

$\hat{\beta}_{MLE}$, 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 $\delta_{(1)}$ 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 $\hat{\delta}_{(1)}$ and the inverse of $\hat{\beta}_{LL}$, 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 $\hat{\delta}_{(1)}$ and inverse of $\hat{\beta}_{LL}$ also vary across the years. For example, using the second-order interconnections and a cut-off value of 20%, $\hat{\delta}_{(1)}$ and inverse of $\hat{\beta}_{LL}$ 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 $\delta_{(1)}$. 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 **W** matrices from the input-ouput tables downloaded from the BEA website.[19] The top five largest estimates of δ, denoted by $\hat{\delta}_{(1)} \geq \hat{\delta}_{(2)} \geq ... \geq \hat{\delta}_{(5)}$, for each of the years 1972 to 2007 are given in Table 17. The identities of the associated sectors are given in Table 18. 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 15, no sector is strongly dominant. The highest $\hat{\delta}_{(1)}$ 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.[20] Accordingly, to provide more reliable estimates of $\delta_{(1)}, \delta_{(2)}, ..., \delta_{(5)}$ and the associated sectoral identities, we also give 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 19, which also gives standard

---

[19]The **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 transformations can be found in Appendix E.

[20]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.

errors in parentheses, computed using (89). 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 $\hat{\delta}_{(1)} = 0.763$ (0.037) for the first sub-sample (1972-1992), and $\hat{\delta}_{(1)} = 0.716$ (0.045) for the second sub-sample (1997-2007). The estimates do not differ significantly and identify the wholesale trade sector 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 0.605 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).

7 Concluding remarks

This paper builds on the seminal contribution of Acemoglu et al. (2012) and derives exact conditions under which micro (sectoral) shocks can have aggregate effects. The paper also presents a simple nonparametric estimator of the degree of pervasiveness of micro shocks that compares favorably with the parametric estimates based on Pareto distribution fitted to the outdegrees. The paper also presents a simple test of the degree of pervasiveness of the most dominant units in the network, which are shown to have satisfactory size and power properties when N is large, even if T is quite small. The analysis of this paper has been 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.[21] 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.

---

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

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.

Gabaix, X. and Y. M. Ioannides (2004). The evolution of city size distributions. *Handbook of regional and urban economics 4*, 2341–2378.

Goldstein, M. L., S. A. Morris, and G. G. Yen (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems 41*, 255–258.

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.

Horn, R. A. and C. R. Johnson (2012). *Matrix analysis*. Cambridge University Press.

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.

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

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. Unpublished manuscript. Department of Economics, University of Southern California.

Table 1: Bias, RMSE, size and power of the extremum estimator for the dominant units in Experiments A.1 and A.2

Experiment A.1: One strongly dominant unit, $\delta_{(1)} = 1$

	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
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
4	-1.26	-0.54	-0.39	-0.26	-0.08	10.71	8.65	7.93	7.14	3.79	4.50	4.35	4.35	4.25	4.20	12.45	19.40	22.20	26.35	73.70
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
8	-1.27	-0.55	-0.40	-0.28	-0.09	7.63	6.13	5.62	5.05	2.68	4.90	4.75	4.75	4.55	4.65	19.65	31.45	38.60	48.25	96.05
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
50	-1.00	-0.33	-0.20	-0.10	0.00	3.16	2.45	2.24	2.01	1.07	5.00	4.40	4.60	4.85	4.95	84.25	97.75	99.30	99.80	100.00
100	-1.00	-0.33	-0.20	-0.10	0.00	2.38	1.79	1.62	1.45	0.77	7.80	6.25	5.90	5.65	5.20	98.60	100.00	100.00	100.00	100.00

Experiment A.2: Two strongly dominant units, $\delta_{(1)} = \delta_{(2)} = 1$

$\delta_{(1)} = 1$

	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
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
4	3.85	4.06	3.94	3.71	2.08	9.63	8.25	7.68	7.00	3.78	2.80	3.70	4.15	4.25	4.40	19.85	33.25	38.40	45.15	93.00
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
8	1.98	2.55	2.55	2.46	1.41	6.67	5.78	5.41	4.95	2.68	3.00	4.15	4.55	4.80	5.30	30.55	51.15	60.65	71.65	99.60
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
50	-0.30	0.70	0.86	0.93	0.60	2.55	2.17	2.07	1.94	1.09	1.90	2.45	3.05	4.10	5.30	92.90	99.95	100.00	100.00	100.00
100	-0.73	0.36	0.54	0.65	0.45	1.91	1.48	1.42	1.35	0.77	2.55	2.20	3.30	4.05	5.45	99.95	100.00	100.00	100.00	100.00

$\delta_{(2)} = 1$

	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
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
4	-8.42	-5.85	-5.16	-4.47	-2.27	12.20	9.26	8.37	7.43	3.88	7.40	5.40	5.10	4.80	4.50	2.40	3.80	5.45	7.35	53.80
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
8	-6.43	-4.24	-3.68	-3.15	-1.56	8.93	6.59	5.92	5.23	2.71	8.55	5.75	5.20	4.45	4.15	2.90	10.10	14.75	23.70	92.05
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
50	-3.70	-2.04	-1.66	-1.33	-0.60	4.45	2.87	2.49	2.13	1.07	17.60	8.05	6.45	5.80	4.45	55.40	93.80	97.85	99.40	100.00
100	-3.23	-1.66	-1.31	-1.02	-0.43	3.70	2.21	1.88	1.58	0.78	27.55	12.10	9.40	7.50	5.10	91.45	99.90	100.00	100.00	100.00

Notes: The Data Generating Process (DGP) is given by (92). For Experiment A.1, the DGP includes one strongly dominant unit and the rest are non-dominant: $\delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$. For Experiment A.2, the DGP includes two strongly dominant units and the rest are non-dominant: $\delta_{(1)} = \delta_{(2)} = 1$, with $\delta_{(i)} = 0$ for $i = 3, 4, ..., N$. $\delta_{(1)}$ and $\delta_{(2)}$ denote the largest and second-largest of δ respectively ($\delta_{(1)} \geq \delta_{(2)}$), which are estimated by (76). The standard errors of $\hat{\delta}_{(i)}$ cannot be computed when $T = 1$, otherwise they are computed by (80). The power is calculated at 0.9 if $\delta_{(1)} = 1$. The number of replications is 2,000.

Table 2: Bias, RMSE, size and power of the extremum estimator for the dominant units in Experiment A.3

Experiment A.3: One strongly dominant unit and one weakly dominant unit, $\delta_{(1)} = 1$, $\delta_{(2)} = 0.75$

$\delta_{(1)} = 1$

$T\backslash N$	Bias (×100)					RMSE (×100)					Size (×100)					Power (×100)				
	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
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
4	-1.65	-0.66	-0.46	-0.30	-0.08	10.31	8.45	7.80	7.06	3.79	3.60	3.65	3.95	4.00	4.20	10.55	18.35	21.70	25.75	73.70
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
8	-2.00	-0.80	-0.55	-0.35	-0.09	7.74	6.15	5.63	5.06	2.68	5.45	4.85	4.90	4.55	4.65	16.25	30.00	37.15	47.60	96.05
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
50	-1.75	-0.58	-0.35	-0.17	0.00	3.47	2.50	2.26	2.02	1.07	7.70	4.60	4.60	4.85	4.95	76.95	97.10	99.10	99.75	100.00
100	-1.75	-0.58	-0.35	-0.17	0.00	2.78	1.85	1.65	1.46	0.77	12.95	7.05	6.45	6.05	5.25	96.60	99.95	100.00	100.00	100.00

$\delta_{(2)} = 0.75$

$T\backslash N$	Bias (×100)					RMSE (×100)					Size (×100)					Power (×100)				
	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
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
4	-2.42	-0.96	-0.65	-0.41	-0.10	10.74	8.71	8.02	7.24	3.87	4.40	4.35	4.45	4.45	4.70	71.85	83.65	88.75	94.30	100.00
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
8	-1.95	-0.73	-0.48	-0.29	-0.06	7.81	6.21	5.69	5.12	2.72	5.40	4.95	4.70	4.75	4.80	94.20	98.70	99.55	99.85	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
50	-1.75	-0.58	-0.35	-0.17	0.00	3.53	2.54	2.30	2.06	1.09	9.35	5.30	4.90	5.10	4.80	100.00	100.00	100.00	100.00	100.00
100	-1.71	-0.55	-0.32	-0.15	0.01	2.78	1.86	1.66	1.47	0.78	12.60	6.45	5.95	5.45	5.30	100.00	100.00	100.00	100.00	100.00

Notes: The DGP is given by (92). For Experiment A.3, the DGP includes 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 $i = 3, 4, ..., N$. $\delta_{(1)}$ and $\delta_{(2)}$ denote the largest and second-largest of δ respectively $(\hat{\delta}_{(1)} \geq \hat{\delta}_{(2)})$, which are estimated by (76). The standard errors of $\hat{\delta}_{(i)}$ cannot be computed when $T = 1$, otherwise they are computed by (80). The power is calculated at 0.9 if $\delta_{(1)} = 1$, and at 1 if $\delta = 0.75$. The power is calculated at 0.9 if $\delta_{(1)} = 1$, and at 1 if $\delta = 0.75$. The number of replications is 2, 000.

Table 3: Bias, RMSE, size and power of the extremum estimator for the top four dominant units in Experiment B

	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
$\delta_{(1)} = 0.9$																				
1	4.04	5.60	5.44	4.93	1.28	15.42	13.87	13.08	12.00	6.70	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-2.43	1.17	1.63	1.76	0.37	11.82	9.84	9.24	8.54	5.01	1.15	1.95	2.30	2.45	3.20	14.45	35.70	44.85	55.50	97.85
4	-6.01	-1.26	-0.43	0.10	0.03	10.50	7.41	6.83	6.26	3.64	3.25	1.25	1.55	1.85	3.35	20.40	57.05	70.60	83.20	99.95
6	-7.24	-2.10	-1.14	-0.45	-0.03	10.34	6.60	5.95	5.36	3.05	7.80	2.75	2.20	2.50	4.15	25.25	71.90	84.50	94.25	100.00
8	-7.92	-2.52	-1.47	-0.72	-0.09	10.27	6.08	5.35	4.76	2.67	13.70	4.30	3.60	3.35	4.45	30.55	83.30	93.15	98.30	100.00
10	-8.19	-2.67	-1.60	-0.81	-0.09	10.21	5.79	5.03	4.44	2.45	19.15	5.70	4.30	4.00	5.40	37.30	89.80	97.00	99.60	100.00
20	-8.80	-2.96	-1.80	-0.93	-0.05	9.88	4.77	3.91	3.29	1.72	44.65	10.80	6.75	4.80	4.90	64.15	99.60	100.00	100.00	100.00
50	-8.98	-3.00	-1.80	-0.90	0.00	9.45	3.86	2.87	2.20	1.07	83.25	23.20	10.90	6.30	4.95	96.25	100.00	100.00	100.00	100.00
100	-9.00	-3.00	-1.80	-0.90	0.00	9.25	3.47	2.42	1.71	0.77	98.40	40.05	20.40	9.30	5.25	99.95	100.00	100.00	100.00	100.00
$\delta_{(2)} = 0.9^2 = 0.81$																				
1	-1.74	1.40	1.78	1.85	0.39	11.46	10.04	9.57	8.92	5.70	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-6.33	-1.54	-0.68	-0.10	0.08	10.89	7.90	7.35	6.80	4.36	0.25	0.30	0.30	0.25	1.75	4.85	22.40	32.70	48.10	98.50
4	-8.18	-2.69	-1.61	-0.80	-0.07	10.87	6.82	6.13	5.57	3.57	2.60	0.95	0.75	0.70	2.95	10.10	49.05	65.70	82.85	99.95
6	-8.46	-2.79	-1.67	-0.84	-0.10	10.45	6.11	5.43	4.90	3.02	6.85	1.85	1.45	1.45	4.35	17.55	72.15	84.95	94.05	100.00
8	-8.65	-2.85	-1.70	-0.84	-0.04	10.31	5.67	4.93	4.40	2.67	13.10	2.60	2.10	1.90	4.20	26.20	84.10	94.20	98.65	100.00
10	-8.74	-2.87	-1.70	-0.82	0.01	10.10	5.29	4.55	4.04	2.40	17.55	3.50	3.00	2.75	4.40	33.95	91.75	97.45	99.75	100.00
20	-8.87	-2.92	-1.73	-0.85	0.02	9.73	4.52	3.68	3.12	1.69	43.90	8.60	5.65	4.90	5.20	66.45	99.85	100.00	100.00	100.00
50	-8.97	-2.99	-1.80	-0.90	0.00	9.44	3.87	2.89	2.23	1.09	84.85	23.10	12.90	7.65	4.80	95.65	100.00	100.00	100.00	100.00
100	-8.96	-2.97	-1.77	-0.87	0.01	9.22	3.46	2.41	1.71	0.78	98.60	39.50	20.40	9.50	5.30	99.95	100.00	100.00	100.00	100.00
$\delta_{(3)} = 0.9^3 = 0.729$																				
1	-3.59	-0.05	0.52	0.82	0.35	10.32	8.70	8.32	7.84	5.24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-6.86	-1.81	-0.86	-0.20	0.17	10.43	7.19	6.66	6.21	4.28	0.15	0.05	0.20	0.25	1.05	2.55	19.10	31.80	48.25	99.10
4	-8.26	-2.55	-1.44	-0.63	0.00	10.45	6.21	5.59	5.13	3.38	1.90	0.35	0.20	0.30	2.45	7.00	51.85	69.95	84.75	100.00
6	-8.45	-2.67	-1.52	-0.68	0.05	10.21	5.78	5.09	4.60	2.94	6.55	1.15	0.95	1.15	3.30	16.40	74.45	87.25	95.10	100.00
8	-8.52	-2.67	-1.52	-0.66	0.09	10.00	5.39	4.69	4.20	2.60	11.50	1.60	1.20	1.45	3.95	24.45	85.95	94.95	98.90	100.00
10	-8.61	-2.73	-1.57	-0.70	0.06	9.96	5.22	4.49	3.99	2.39	17.50	3.25	2.60	2.40	4.35	34.15	92.50	97.85	99.75	100.00
20	-8.75	-2.81	-1.63	-0.75	0.07	9.66	4.52	3.71	3.17	1.74	43.70	9.00	5.10	4.20	4.80	65.30	99.80	100.00	100.00	100.00
50	-8.94	-2.96	-1.77	-0.87	0.01	9.39	3.82	2.84	2.20	1.08	84.30	22.60	12.10	6.65	4.20	96.05	100.00	100.00	100.00	100.00
100	-8.98	-2.99	-1.79	-0.89	0.00	9.21	3.43	2.36	1.65	0.74	99.25	39.15	19.10	8.50	3.90	100.00	100.00	100.00	100.00	100.00
$\delta_{(4)} = 0.9^4 = 0.6561$																				
1	-4.04	-0.36	0.23	0.57	0.28	9.30	7.78	7.54	7.21	5.03	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	-7.01	-1.83	-0.88	-0.21	0.00	10.19	6.90	6.42	6.01	4.18	0.05	0.10	0.05	0.05	0.90	1.45	18.30	30.45	47.05	99.05
4	-8.08	-2.47	-1.38	-0.57	0.05	10.10	5.97	5.37	4.93	3.31	1.45	0.25	0.20	0.50	2.20	7.05	52.80	71.15	85.65	100.00
6	-8.36	-2.56	-1.43	-0.60	0.11	9.98	5.52	4.86	4.38	2.82	4.55	1.00	0.80	0.85	3.15	15.20	74.85	89.65	96.85	100.00
8	-8.54	-2.70	-1.55	-0.70	0.04	9.84	5.14	4.46	4.00	2.58	9.45	1.45	1.00	0.80	3.15	24.00	88.10	95.45	99.25	100.00
10	-8.73	-2.84	-1.66	-0.79	0.00	9.85	4.92	4.16	3.67	2.29	15.60	2.55	1.70	1.30	3.80	32.60	94.35	98.75	99.85	100.00
20	-8.97	-3.02	-1.82	-0.94	-0.05	9.77	4.53	3.68	3.12	1.71	46.00	8.10	5.35	3.85	5.10	64.50	99.65	100.00	100.00	100.00
50	-8.90	-2.94	-1.75	-0.87	0.01	9.34	3.80	2.86	2.24	1.12	83.95	22.15	11.70	7.05	5.60	96.70	100.00	100.00	100.00	100.00
100	-8.96	-2.97	-1.78	-0.88	0.01	9.21	3.46	2.41	1.71	0.78	98.55	38.85	20.60	9.45	5.90	99.90	100.00	100.00	100.00	100.00

Notes: The DGP is given by (92), where the true values of δ are generated as $\delta_{(i)} = 0.9^i$, for $i = 1, 2, ..., N$. $\delta_{(i)}$ denotes the ith largest δ, namely, $\delta_{(1)} \geq \delta_{(2)} \geq \delta_{(3)} \geq \delta_{(4)} \geq ...$, which are estimated by (76). The standard errors of $\hat{\delta}_{(i)}$ cannot be computed when $T = 1$, otherwise they are computed by (80). The power is calculated at $\delta_{(i)} - 0.2$ when the true value is $\delta_{(i)}$, for $i = 1, 2, 3, 4$. The number of replications is 2,000.

Table 4: Frequencies with which the dominant units are jointly selected in Experiments A.1, A.2 and A.3

	Empirical frequency (percent)									
	Experiment A.1: One strongly dominant unit, $\delta_{(1)} = 1$					Experiment A.2: Two strongly dominant units, $\delta_{(1)} = \delta_{(2)} = 1$				
$T\backslash N$	100	300	500	1,000	450,000	$T\backslash N$	100	300	500	1,000	450,000
1	97.25	99.55	99.80	99.90	100.0	1	94.20	99.10	99.65	99.85	100.0
2	100.0	100.0	100.0	100.0	100.0	2	100.0	100.0	100.0	100.0	100.0
4	100.0	100.0	100.0	100.0	100.0	4	100.0	100.0	100.0	100.0	100.0
6	100.0	100.0	100.0	100.0	100.0	6	100.0	100.0	100.0	100.0	100.0
8	100.0	100.0	100.0	100.0	100.0	8	100.0	100.0	100.0	100.0	100.0
10	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	20	100.0	100.0	100.0	100.0	100.0
50	100.0	100.0	100.0	100.0	100.0	50	100.0	100.0	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0	100.0	100.0

	Experiment A.3: One strongly dominant unit and one weakly dominant unit, $\delta_{(1)} = 1$, $\delta_{(2)} = 0.75$				
$T\backslash N$	100	300	500	1,000	450,000
1	61.55	75.90	79.70	85.45	98.75
2	86.30	91.90	93.65	95.65	100.0
4	94.90	97.75	98.40	99.20	100.0
6	97.75	99.10	99.50	99.85	100.0
8	99.10	99.90	99.95	99.95	100.0
10	99.65	99.95	100.0	100.0	100.0
20	99.95	100.0	100.0	100.0	100.0
50	100.0	100.0	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	100.0

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

Table 5: Frequencies with which each of the top four dominant units are selected in Experiment B

	Empirical frequency (percent)				
$T \backslash N$	100	300	500	1,000	450,000
	$\delta_{(1)} = 0.9$				
1	40.40	48.40	51.85	56.15	78.05
2	54.00	60.65	64.00	68.00	87.65
4	65.85	72.70	75.25	79.00	95.20
6	72.20	80.20	83.30	86.50	97.85
8	76.20	83.90	86.50	89.35	99.25
10	79.60	85.85	88.25	90.85	99.80
20	89.95	94.85	95.85	97.25	99.95
50	98.35	99.65	99.95	100.0	100.0
100	99.90	100.0	100.0	100.0	100.0
	$\delta_{(2)} = 0.9^2 = 0.81$				
1	23.10	27.15	29.15	31.25	54.95
2	31.55	38.20	41.60	46.00	73.20
4	39.70	48.95	53.00	57.95	88.95
6	48.05	59.00	63.65	69.15	94.20
8	53.65	65.35	70.70	75.75	97.55
10	59.45	69.95	75.15	80.40	98.95
20	78.20	87.60	89.70	92.90	99.90
50	94.90	98.75	99.45	99.80	100.0
100	99.65	99.90	99.90	100.0	100.0
	$\delta_{(3)} = 0.9^3 = 0.729$				
1	14.80	18.85	21.80	25.25	48.30
2	21.90	28.55	31.75	35.35	65.45
4	34.55	42.55	46.60	52.30	84.95
6	40.40	50.80	55.15	61.00	90.90
8	47.70	58.90	64.10	70.70	95.20
10	54.95	65.40	70.95	76.50	97.70
20	73.20	83.50	86.75	90.75	99.95
50	92.30	97.20	98.45	99.35	100.0
100	99.00	99.85	99.85	99.95	100.0
	$\delta_{(4)} = 0.9^4 = 0.6561$				
1	10.90	14.50	17.05	18.80	41.80
2	17.75	23.85	25.90	28.80	59.80
4	29.65	36.20	39.80	45.15	79.35
6	37.30	47.50	51.55	56.55	87.80
8	43.35	54.15	59.00	65.65	92.45
10	50.80	60.65	65.55	71.30	95.75
20	68.00	78.00	82.30	87.15	99.85
50	89.05	94.60	96.20	97.75	100.0
100	97.50	99.40	99.75	99.90	100.0

Notes: This table complements Table 3 and reports the frequencies with which each of the top four dominant units are selected across 2,000 replications. The DGP is given by (92), where the true values of δ are generated as $\delta_{(i)} = 0.9^i$, for $i = 1, 2, ..., N$. See also the notes to Table 3.

Table 6: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment A.1

$T\backslash N$	100		300		500		1,000		450,000	
	$\widehat{\beta}_{LL}$ [20%]	Inverse of $\widehat{\delta}_{(1)}$	$\widehat{\beta}_{LL}$ [20%]	Inverse of $\widehat{\delta}_{(1)}$	$\widehat{\beta}_{LL}$ [20%]	Inverse of $\widehat{\delta}_{(1)}$	$\widehat{\beta}_{LL}$ [20%]	Inverse of $\widehat{\delta}_{(1)}$	$\widehat{\beta}_{LL}$ [20%]	Inverse of $\widehat{\delta}_{(1)}$
1	1.11	1.06	1.28	1.04	1.39	1.03	1.54	1.02	2.11	1.01
	(0.35)	(N/A)	(0.23)	(N/A)	(0.20)	(N/A)	(0.15)	(N/A)	(0.01)	(N/A)
2	1.11	1.04	1.29	1.02	1.39	1.02	1.55	1.01	2.11	1.00
	(0.25)	(0.17)	(0.17)	(0.13)	(0.14)	(0.12)	(0.11)	(0.11)	(0.01)	(0.05)
4	1.12	1.03	1.30	1.01	1.40	1.01	1.55	1.01	2.11	1.00
	(0.18)	(0.12)	(0.12)	(0.09)	(0.10)	(0.08)	(0.08)	(0.07)	(0.00)	(0.04)
6	1.12	1.02	1.30	1.01	1.40	1.01	1.55	1.01	2.11	1.00
	(0.14)	(0.09)	(0.10)	(0.07)	(0.08)	(0.07)	(0.06)	(0.06)	(0.00)	(0.03)
8	1.13	1.02	1.30	1.01	1.40	1.01	1.55	1.01	2.11	1.00
	(0.13)	(0.08)	(0.08)	(0.06)	(0.07)	(0.06)	(0.05)	(0.05)	(0.00)	(0.03)
10	1.13	1.02	1.30	1.01	1.40	1.01	1.55	1.00	2.11	1.00
	(0.11)	(0.07)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.00)	(0.02)
20	1.13	1.01	1.30	1.01	1.40	1.00	1.55	1.00	2.11	1.00
	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.00)	(0.02)
50	1.13	1.01	1.30	1.00	1.40	1.00	1.55	1.00	2.11	1.00
	(0.05)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.00)	(0.01)
100	1.13	1.01	1.30	1.00	1.40	1.00	1.55	1.00	2.11	1.00
	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)

Notes: The DGP is given by (92). For Experiment A.1, the DGP includes one strongly dominant unit and the rest are non-dominant: $\delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$. $\hat{\beta}_{LL}$ denotes the estimates of the shape parameter of the Pareto distribution obtained by the log-log regression with Gabaix and Ibragimov (2011) correction using the OLS regression defined by (66) assuming a 20% cut-off value ($N_{\min}/N = 20\%$) as in Acemoglu et al. (2012), which means that the Pareto tail is taken as the top 20% of all units. $\hat{\delta}_{(1)} = \max_i(\hat{\delta}_i)$, and $\hat{\delta}_i$ is calculated according to (76). All estimates are averaged across 2,000 replications. Standard errors are in parentheses. The standard errors for $\hat{\beta}_{LL}$ and $\hat{\delta}_{(1)}$ are computed by (67) and (80), respectively, and for the inverse of $\hat{\delta}_{(1)}$ by the delta method. (N/A) indicates that the standard error of $\hat{\delta}_{(1)}$ cannot be computed when $T = 1$.

Table 7: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment C ($\beta = 1$)

		$T = 1$					$T = 4$			
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed				**Log-log regression** $\left(\widehat{\beta}_{LL}\right)$						
cut-off value										
10%	1.11	1.02	1.01	1.00	1.00	1.13	1.05	1.03	1.02	1.00
	(0.50)	(0.26)	(0.20)	(0.14)	(0.01)	(0.25)	(0.14)	(0.10)	(0.07)	(0.00)
20%	1.04	1.01	1.00	1.00	1.00	1.08	1.03	1.02	1.01	1.00
	(0.33)	(0.18)	(0.14)	(0.10)	(0.00)	(0.17)	(0.09)	(0.07)	(0.05)	(0.00)
30%	1.02	1.00	1.00	1.00	1.00	1.05	1.02	1.01	1.00	1.00
	(0.26)	(0.15)	(0.12)	(0.08)	(0.00)	(0.14)	(0.08)	(0.06)	(0.04)	(0.00)
Infeasible				*Using true $d_{\min,t}$*						
cut-off value	1.03	1.00	1.00	1.00	1.00	1.07	1.03	1.01	1.01	1.00
24%	(0.30)	(0.17)	(0.13)	(0.09)	(0.00)	(0.15)	(0.09)	(0.07)	(0.05)	(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	1.11	1.04	1.02	1.01	1.00
	(0.39)	(0.20)	(0.15)	(0.10)	(0.00)	(0.18)	(0.09)	(0.07)	(0.05)	(0.00)
20%	1.11	1.03	1.02	1.01	1.00	1.05	1.02	1.01	1.00	1.00
	(0.25)	(0.13)	(0.10)	(0.07)	(0.00)	(0.12)	(0.07)	(0.05)	(0.04)	(0.00)
30%	1.06	1.01	1.01	1.00	1.00	1.00	0.98	0.98	0.98	0.99
	(0.19)	(0.11)	(0.08)	(0.06)	(0.00)	(0.09)	(0.05)	(0.04)	(0.03)	(0.00)
Infeasible				*Using true $d_{\min,t}$*						
cut-off value	1.09	1.03	1.01	1.01	1.00	1.03	1.01	1.00	1.00	1.00
24%	(0.23)	(0.12)	(0.09)	(0.07)	(0.00)	(0.11)	(0.06)	(0.05)	(0.03)	(0.00)
Estimated				**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$						
cut-off value	44%	38%	37%	35%	24%	33%	29%	28%	26%	20%
	1.02	1.00	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00
	(0.17)	(0.10)	(0.08)	(0.06)	(0.00)	(0.10)	(0.06)	(0.04)	(0.03)	(0.00)
				Inverse of $\widehat{\delta}_{(1)}$						
	1.04	1.03	1.02	1.02	1.00	1.00	1.00	0.99	1.00	1.00
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.06)	(0.04)	(0.04)	(0.03)	(0.01)

Notes: The DGP follows the Pareto tail distribution given by (85) with $\beta = 1$. $d_{\min,t}$ denotes the lower bound for the Pareto distribution and is given by (97). 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 (98) assuming the true value of $d_{\min,t}$ is known. All estimates are averaged across 2,000 replications. Standard errors are in parentheses. $\hat{\beta}_{LL}$ is obtained by running the log-log regression (66) with Gabaix and Ibragimov (2011) correction. $\widehat{\beta}_{MLE}$ is computed by (68). $\hat{\beta}_{CSN}$ is calculated by applying the joint MLE procedure described in Clauset et al. (2009). $\hat{\delta}_{(1)}$ is computed according to (76), and its standard error by (80). The standard error for the inverse of $\hat{\delta}_{(1)}$ is computed by the delta method. (N/A) indicates that the standard error of $\hat{\delta}_{(1)}$ cannot be computed when $T = 1$.

Table 8: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment C ($\beta = 1.1$)

		$T=1$					$T=4$				
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed				Log-log regression $\left(\widehat{\beta}_{LL}\right)$						
cut-off value										
10%	1.22	1.12	1.11	1.10	1.10	1.23	1.15	1.13	1.12	1.10
	(0.55)	(0.29)	(0.22)	(0.16)	(0.01)	(0.28)	(0.15)	(0.11)	(0.08)	(0.00)
20%	1.15	1.11	1.10	1.10	1.10	1.18	1.13	1.12	1.11	1.10
	(0.36)	(0.20)	(0.16)	(0.11)	(0.01)	(0.19)	(0.10)	(0.08)	(0.06)	(0.00)
30%	1.12	1.10	1.10	1.09	1.10	1.15	1.12	1.11	1.10	1.10
	(0.29)	(0.16)	(0.13)	(0.09)	(0.00)	(0.15)	(0.08)	(0.06)	(0.04)	(0.00)
Infeasible				Using true $d_{\min,t}$						
cut-off value	1.14	1.10	1.10	1.10	1.10	1.17	1.13	1.12	1.11	1.10
21%	(0.36)	(0.20)	(0.15)	(0.11)	(0.01)	(0.18)	(0.10)	(0.08)	(0.05)	(0.00)
Assumed				Maximum Likelihood Estimation $\left(\widehat{\beta}_{MLE}\right)$						
cut-off value										
10%	1.37	1.18	1.14	1.12	1.10	1.21	1.14	1.12	1.11	1.10
	(0.43)	(0.21)	(0.16)	(0.11)	(0.01)	(0.19)	(0.10)	(0.08)	(0.06)	(0.00)
20%	1.22	1.13	1.12	1.11	1.10	1.15	1.12	1.11	1.10	1.10
	(0.27)	(0.15)	(0.11)	(0.08)	(0.00)	(0.13)	(0.07)	(0.06)	(0.04)	(0.00)
30%	1.16	1.11	1.10	1.09	1.09	1.09	1.08	1.07	1.07	1.08
	(0.21)	(0.12)	(0.09)	(0.06)	(0.00)	(0.10)	(0.06)	(0.04)	(0.03)	(0.00)
Infeasible				Using true $d_{\min,t}$						
cut-off value	1.22	1.13	1.12	1.11	1.10	1.14	1.11	1.11	1.10	1.10
21%	(0.27)	(0.14)	(0.11)	(0.08)	(0.00)	(0.12)	(0.07)	(0.05)	(0.04)	(0.00)
Estimated				Feasible MLE $\left(\widehat{\beta}_{CSN}\right)$						
cut-off value	42%	36%	34%	32%	22%	31%	28%	27%	25%	19%
	1.13	1.10	1.10	1.10	1.10	1.12	1.11	1.10	1.10	1.10
	(0.19)	(0.11)	(0.09)	(0.06)	(0.00)	(0.11)	(0.06)	(0.05)	(0.04)	(0.00)
				Inverse of $\widehat{\delta}_{(1)}$						
	1.12	1.11	1.11	1.10	1.09	1.08	1.08	1.08	1.08	1.09
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.06)	(0.04)	(0.04)	(0.03)	(0.01)

Notes: The data are generated to follow the Pareto tail distribution given by (85) with $\beta = 1.1$. See notes to Table 7 for other details.

Table 9: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment C ($\beta = 1.3$)

			$T = 1$					$T = 4$		
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed					**Log-log regression** $\left(\widehat{\beta}_{LL}\right)$					
cut-off value										
10%	1.44	1.33	1.31	1.30	1.30	1.43	1.34	1.32	1.31	1.30
	(0.65)	(0.34)	(0.26)	(0.18)	(0.01)	(0.32)	(0.17)	(0.13)	(0.09)	(0.00)
20%	1.35	1.31	1.30	1.29	1.30	1.38	1.33	1.31	1.31	1.30
	(0.43)	(0.24)	(0.18)	(0.13)	(0.01)	(0.22)	(0.12)	(0.09)	(0.07)	(0.00)
30%	1.31	1.29	1.29	1.29	1.29	1.34	1.31	1.30	1.29	1.29
	(0.34)	(0.19)	(0.15)	(0.11)	(0.00)	(0.17)	(0.10)	(0.07)	(0.05)	(0.00)
Infeasible					*Using true $d_{\min,t}$*					
cut-off value	1.37	1.31	1.30	1.30	1.30	1.39	1.33	1.32	1.31	1.30
16%	(0.49)	(0.27)	(0.20)	(0.14)	(0.01)	(0.24)	(0.13)	(0.10)	(0.07)	(0.00)
Assumed					**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$					
cut-off value										
10%	1.61	1.39	1.35	1.32	1.30	1.42	1.34	1.32	1.31	1.30
	(0.51)	(0.25)	(0.19)	(0.13)	(0.01)	(0.23)	(0.12)	(0.09)	(0.07)	(0.00)
20%	1.44	1.34	1.32	1.31	1.30	1.35	1.31	1.30	1.30	1.30
	(0.32)	(0.17)	(0.13)	(0.09)	(0.00)	(0.15)	(0.08)	(0.07)	(0.05)	(0.00)
30%	1.34	1.28	1.26	1.26	1.25	1.26	1.24	1.24	1.24	1.25
	(0.24)	(0.13)	(0.10)	(0.07)	(0.00)	(0.11)	(0.07)	(0.05)	(0.04)	(0.00)
Infeasible					*Using true $d_{\min,t}$*					
cut-off value	1.49	1.35	1.33	1.31	1.30	1.36	1.32	1.31	1.30	1.30
16%	(0.37)	(0.19)	(0.15)	(0.10)	(0.00)	(0.17)	(0.09)	(0.07)	(0.05)	(0.00)
Estimated					**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$					
cut-off value	39%	32%	30%	28%	17%	29%	25%	24%	22%	16%
	1.31	1.30	1.30	1.30	1.30	1.33	1.31	1.30	1.30	1.30
	(0.23)	(0.14)	(0.11)	(0.08)	(0.00)	(0.14)	(0.08)	(0.06)	(0.05)	(0.00)
					Inverse of $\widehat{\delta}_{(1)}$					
	1.27	1.27	1.27	1.27	1.27	1.22	1.23	1.24	1.25	1.27
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.06)	(0.04)	(0.03)	(0.02)	(0.00)

Notes: The DGP follows the Pareto tail distribution given by (85) with $\beta = 1.3$. See the notes to Table 7 for other details.

Table 10: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment C $(\beta = 1.5)$

		T = 1					T = 4				
	N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed		**Log-log regression** $\left(\widehat{\beta}_{LL}\right)$									
cut-off value											
10%		1.66	1.53	1.51	1.50	1.50	1.62	1.54	1.52	1.51	1.50
		(0.74)	(0.40)	(0.30)	(0.21)	(0.01)	(0.36)	(0.20)	(0.15)	(0.11)	(0.01)
20%		1.56	1.51	1.50	1.49	1.50	1.57	1.52	1.51	1.50	1.50
		(0.49)	(0.27)	(0.21)	(0.15)	(0.01)	(0.25)	(0.14)	(0.11)	(0.08)	(0.00)
30%		1.50	1.47	1.47	1.47	1.48	1.52	1.49	1.48	1.48	1.48
		(0.39)	(0.22)	(0.17)	(0.12)	(0.01)	(0.20)	(0.11)	(0.09)	(0.06)	(0.00)
Infeasible		*Using true $d_{\min,t}$*									
cut-off value		1.63	1.52	1.51	1.50	1.50	1.60	1.53	1.52	1.51	1.50
13%		(0.66)	(0.35)	(0.27)	(0.19)	(0.01)	(0.32)	(0.17)	(0.13)	(0.09)	(0.00)
Assumed		**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$									
cut-off value											
10%		1.86	1.60	1.56	1.53	1.50	1.63	1.55	1.52	1.51	1.50
		(0.59)	(0.29)	(0.22)	(0.15)	(0.01)	(0.26)	(0.14)	(0.11)	(0.08)	(0.00)
20%		1.64	1.53	1.51	1.49	1.48	1.53	1.49	1.49	1.48	1.48
		(0.37)	(0.20)	(0.15)	(0.11)	(0.00)	(0.17)	(0.10)	(0.07)	(0.05)	(0.00)
30%		1.48	1.42	1.40	1.40	1.39	1.40	1.38	1.38	1.38	1.39
		(0.27)	(0.15)	(0.11)	(0.08)	(0.00)	(0.13)	(0.07)	(0.06)	(0.04)	(0.00)
Infeasible		*Using true $d_{\min,t}$*									
cut-off value		1.80	1.58	1.55	1.52	1.50	1.59	1.53	1.52	1.51	1.50
13%		(0.52)	(0.25)	(0.19)	(0.13)	(0.01)	(0.22)	(0.12)	(0.09)	(0.07)	(0.00)
Estimated		**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$									
cut-off value		36%	29%	26%	24%	14%	26%	22%	21%	19%	13%
		1.50	1.49	1.49	1.49	1.50	1.52	1.51	1.50	1.50	1.50
		(0.28)	(0.17)	(0.14)	(0.10)	(0.01)	(0.16)	(0.10)	(0.08)	(0.06)	(0.00)
		Inverse of $\widehat{\delta}_{(1)}$									
		1.40	1.41	1.41	1.42	1.44	1.35	1.38	1.38	1.40	1.44
		(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.06)	(0.04)	(0.03)	(0.02)	(0.00)

Notes: The DGP follows the Pareto tail distribution given by (85) with $\beta = 1.5$. See the notes to Table 7 for other details.

Table 11: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment A.1 ($\beta = 1$)

			$T = 1$					$T = 4$		
N	100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed				**Log-log regression** $\left(\widehat{\beta}_{LL}\right)$						
cut-off value										
10%	0.98	1.10	1.20	1.36	2.39	0.97	1.11	1.21	1.37	2.39
	(0.44)	(0.29)	(0.24)	(0.19)	(0.02)	(0.22)	(0.14)	(0.12)	(0.10)	(0.01)
20%	1.11	1.28	1.39	1.54	2.11	1.12	1.30	1.40	1.55	2.11
	(0.35)	(0.23)	(0.20)	(0.15)	(0.01)	(0.18)	(0.12)	(0.10)	(0.08)	(0.00)
30%	1.17	1.34	1.44	1.56	1.91	1.19	1.36	1.45	1.57	1.91
	(0.30)	(0.20)	(0.17)	(0.13)	(0.01)	(0.15)	(0.10)	(0.08)	(0.06)	(0.00)
Assumed				**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$						
cut-off value										
10%	1.53	1.74	1.84	1.95	2.11	1.39	1.69	1.80	1.92	2.11
	(0.48)	(0.32)	(0.26)	(0.19)	(0.01)	(0.22)	(0.15)	(0.13)	(0.10)	(0.00)
20%	1.52	1.64	1.68	1.73	1.79	1.43	1.61	1.66	1.71	1.79
	(0.34)	(0.21)	(0.17)	(0.12)	(0.01)	(0.16)	(0.10)	(0.08)	(0.06)	(0.00)
30%	1.42	1.49	1.51	1.54	1.58	1.36	1.47	1.50	1.53	1.58
	(0.26)	(0.16)	(0.12)	(0.09)	(0.00)	(0.12)	(0.08)	(0.06)	(0.04)	(0.00)
Estimated				**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$						
cut-off value	39%	29%	24%	18%	2%	34%	24%	20%	14%	1%
	1.37	1.58	1.69	1.85	2.83	1.36	1.60	1.71	1.87	2.90
	(0.24)	(0.19)	(0.17)	(0.15)	(0.04)	(0.12)	(0.10)	(0.09)	(0.08)	(0.02)
				Inverse of $\widehat{\delta}_{(1)}$						
	1.06	1.04	1.03	1.02	1.01	1.03	1.01	1.01	1.01	1.00
	(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.12)	(0.09)	(0.08)	(0.07)	(0.04)

Notes: The DGP is given by (92). There is one strongly dominant unit and the rest are non-dominant: $\delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$, where $\delta_{(i)}$ denotes the ith largest δ. The true value of β is $\beta = 1$. See the notes to Tables 6 and 7 for other details.

Table 12: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment B where $\delta_{(i)} = 0.9^i$ ($\beta = 1/0.9 = 1.11$)

	N	$T = 1$					$T = 4$				
		100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed					**Log-log regression** $\left(\widehat{\beta}_{LL}\right)$						
cut-off value											
10%		1.06	0.92	0.95	1.04	2.34	1.09	0.94	0.96	1.05	2.34
		(0.48)	(0.24)	(0.19)	(0.15)	(0.02)	(0.24)	(0.12)	(0.10)	(0.07)	(0.01)
20%		0.99	1.00	1.06	1.19	2.09	1.02	1.01	1.07	1.19	2.09
		(0.31)	(0.18)	(0.15)	(0.12)	(0.01)	(0.16)	(0.09)	(0.08)	(0.06)	(0.00)
30%		0.98	1.04	1.12	1.25	1.90	1.00	1.05	1.12	1.25	1.90
		(0.25)	(0.16)	(0.13)	(0.10)	(0.01)	(0.13)	(0.08)	(0.06)	(0.05)	(0.00)
Assumed					**Maximum Likelihood Estimation** $\left(\widehat{\beta}_{MLE}\right)$						
cut-off value											
10%		1.13	1.10	1.22	1.44	2.11	1.01	1.06	1.20	1.42	2.11
		(0.36)	(0.20)	(0.17)	(0.14)	(0.01)	(0.16)	(0.10)	(0.08)	(0.07)	(0.00)
20%		1.04	1.18	1.29	1.45	1.79	0.99	1.16	1.28	1.43	1.79
		(0.23)	(0.15)	(0.13)	(0.10)	(0.01)	(0.11)	(0.07)	(0.06)	(0.05)	(0.00)
30%		1.01	1.16	1.25	1.36	1.57	0.97	1.15	1.24	1.36	1.57
		(0.18)	(0.12)	(0.10)	(0.08)	(0.00)	(0.09)	(0.06)	(0.05)	(0.04)	(0.00)
Estimated					**Feasible MLE** $\left(\widehat{\beta}_{CSN}\right)$						
cut-off value		49%	39%	36%	30%	1%	42%	38%	35%	28%	1%
		0.96	1.13	1.24	1.40	2.82	0.96	1.12	1.23	1.39	2.85
		(0.15)	(0.11)	(0.10)	(0.08)	(0.04)	(0.08)	(0.05)	(0.05)	(0.04)	(0.02)
					Inverse of $\widehat{\delta}_{(1)}$						
		1.09	1.06	1.06	1.07	1.10	1.20	1.13	1.12	1.12	1.11
		(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.16)	(0.11)	(0.10)	(0.09)	(0.05)

Notes: The DGP is given by (92). The true values of δ are generated as $\delta_{(i)} = 0.9^i$, for $i = 1, 2, ..., N$, where $\delta_{(i)}$ denotes the ith largest δ. The true value of β is $\beta = 1/0.9 = 1.11$. See the notes to Tables 6 and 7 for other details.

Table 13: Estimates of the shape parameter, β, of the power law and inverse of the exponent, $\delta_{(1)}$, across 2,000 Monte Carlo replications for Experiment B where $\delta_{(i)} = 0.75^i$ ($\beta = 1/0.75 = 1.33$)

	N	$T = 1$					$T = 4$				
		100	300	500	1,000	450,000	100	300	500	1,000	450,000
Assumed		Log-log regression $\left(\widehat{\beta}_{LL}\right)$									
cut-off value											
10%		1.36	1.39	1.47	1.60	2.39	1.33	1.39	1.47	1.61	2.39
		(0.61)	(0.36)	(0.29)	(0.23)	(0.02)	(0.30)	(0.18)	(0.15)	(0.11)	(0.01)
20%		1.34	1.45	1.54	1.66	2.11	1.34	1.46	1.55	1.67	2.11
		(0.42)	(0.27)	(0.22)	(0.17)	(0.01)	(0.21)	(0.13)	(0.11)	(0.08)	(0.00)
30%		1.31	1.45	1.52	1.63	1.91	1.33	1.46	1.53	1.63	1.91
		(0.34)	(0.22)	(0.18)	(0.13)	(0.01)	(0.17)	(0.11)	(0.09)	(0.07)	(0.00)
Assumed		Maximum Likelihood Estimation $\left(\widehat{\beta}_{MLE}\right)$									
cut-off value											
10%		1.61	1.67	1.75	1.86	2.11	1.42	1.61	1.71	1.84	2.11
		(0.51)	(0.30)	(0.25)	(0.19)	(0.01)	(0.22)	(0.15)	(0.12)	(0.09)	(0.00)
20%		1.46	1.56	1.61	1.68	1.79	1.37	1.53	1.59	1.66	1.79
		(0.33)	(0.20)	(0.16)	(0.12)	(0.01)	(0.15)	(0.10)	(0.08)	(0.06)	(0.00)
30%		1.35	1.43	1.47	1.51	1.58	1.29	1.41	1.45	1.50	1.58
		(0.25)	(0.15)	(0.12)	(0.09)	(0.00)	(0.12)	(0.07)	(0.06)	(0.04)	(0.00)
Estimated		Feasible MLE $\left(\widehat{\beta}_{CSN}\right)$									
cut-off value		41%	29%	24%	18%	1%	31%	22%	18%	14%	1%
		1.31	1.52	1.62	1.78	2.83	1.33	1.55	1.65	1.81	2.89
		(0.23)	(0.18)	(0.16)	(0.14)	(0.04)	(0.13)	(0.10)	(0.09)	(0.08)	(0.02)
		Inverse of $\widehat{\delta}_{(1)}$									
		1.32	1.33	1.33	1.34	1.34	1.40	1.37	1.36	1.35	1.34
		(N/A)	(N/A)	(N/A)	(N/A)	(N/A)	(0.22)	(0.17)	(0.15)	(0.13)	(0.07)

Notes: The DGP is given by (92). The true values of δ are generated as $\delta_{(i)} = 0.75^i$, for $i = 1, 2, ..., N$, where $\delta_{(i)}$ denotes the ith largest δ. The true value of β is $\beta = 1/0.75 = 1.33$. See the notes to Tables 6 and 7 for other details.

Table 14: Bias, RMSE, size and power of the extremum estimator for the dominant units in Experiments A.1, A.2 and A.3 for unbalanced panels

	Experiment	A.1: One strongly dominant unit $\delta_{(1)} = 1$	A.2: Two strongly dominant units $\delta_{(1)} = \delta_{(2)} = 1$		A.3: One strongly and one weakly dominant units $\delta_{(1)} = 1, \delta_{(2)} = 0.75$	
	N		$\delta_{(1)} = 1$	$\delta_{(2)} = 1$	$\delta_{(1)} = 1$	$\delta_{(2)} = 0.75$
Bias	100	-1.25	3.87	-8.41	-1.64	-2.40
($\times 100$)	300	-0.52	4.08	-5.83	-0.65	-0.94
	500	-0.38	3.95	-5.15	-0.45	-0.64
	1,000	-0.26	3.71	-4.47	-0.30	-0.41
	450,000	-0.08	2.08	-2.27	-0.08	-0.10
RMSE	100	10.74	9.64	12.22	10.34	10.72
($\times 100$)	300	8.65	8.27	9.25	8.46	8.71
	500	7.94	7.69	8.37	7.81	8.03
	1,000	7.14	7.01	7.43	7.06	7.24
	450,000	3.79	3.78	3.88	3.79	3.87
Size	100	5.10	3.40	7.25	3.95	4.40
($\times 100$)	300	4.55	3.85	5.55	3.80	4.55
	500	4.25	4.15	5.20	3.90	4.60
	1,000	4.15	4.05	4.65	3.75	4.30
	450,000	4.25	4.40	4.55	4.25	4.70
Power	100	13.00	20.35	2.30	11.30	71.85
($\times 100$)	300	18.80	33.30	3.75	18.00	83.90
	500	22.65	38.85	5.35	22.10	88.80
	1,000	26.50	45.65	7.20	26.10	94.35
	450,000	73.75	93.10	53.80	73.75	100.00

Notes: The unbalanced panels are generated with $T_{max} = 4$. For each Monte Carlo replication, the top 5% of the units in terms of the true degree of dominance do not have missing observations, whereas the rest will have missing data for the first and the last periods with a 50% probability. The DGP is given by (92). For Experiment A.1, the DGP includes one strongly dominant unit and the rest are non-dominant: $\delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, ..., N$. For Experiment A.2, the DGP includes 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, the DGP includes 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 $i = 3, 4, ..., N$. $\delta_{(1)}$ and $\delta_{(2)}$ denote the largest and second-largest of δ respectively ($\delta_{(1)} \geq \delta_{(2)}$), which are estimated by (88), and the standard errors of $\hat{\delta}_{(i)}$ are computed by (89). The power is calculated at 0.9 if $\delta_{(1)} = 1$, and at 1 if $\delta = 0.75$. The number of replications used is 2,000.

Table 15: Yearly estimates of the degree of dominance, δ, 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)

			$\widehat{\delta}_{(1)}$ based on the inverse of $\widehat{\beta}$ using the first-order interconnections							
			Inverse of $\widehat{\beta}_{LL}$			Inverse of $\widehat{\beta}_{MLE}$			Inverse of $\widehat{\beta}_{CSN}$	
			Assumed cut-off value			Assumed cut-off value				Estimated
Year	N	$\widehat{\delta}_{(1)}$	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%
			(0.204)	(0.144)	(0.117)	(0.144)	(0.114)	(0.128)	(0.111)	
1977	524	0.778	0.677	0.725	0.804	0.715	0.852	1.009	0.726	13.6%
			(0.196)	(0.138)	(0.113)	(0.139)	(0.116)	(0.113)	(0.119)	
1982	529	0.788	0.717	0.739	0.818	0.719	0.786	1.039	0.741	15.3%
			(0.194)	(0.137)	(0.112)	(0.137)	(0.106)	(0.115)	(0.111)	
1987	510	0.804	0.667	0.731	0.814	0.724	0.849	1.028	0.742	13.3%
			(0.198)	(0.140)	(0.114)	(0.140)	(0.116)	(0.115)	(0.121)	
1992	476	0.824	0.672	0.758	0.842	0.738	0.891	1.002	0.706	9.5%
			(0.204)	(0.145)	(0.118)	(0.144)	(0.124)	(0.113)	(0.149)	
1997[a]	474	0.778	0.625	0.698	0.791	0.617	0.909	0.982	0.670	13.1%
			(0.206)	(0.145)	(0.119)	(0.146)	(0.151)	(0.134)	(0.127)	
2002	417	0.765	0.639	0.687	0.759	0.685	0.756	0.930	0.730	19.4%
			(0.218)	(0.155)	(0.126)	(0.154)	(0.121)	(0.121)	(0.111)	

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. $\hat{\delta}_{(1)}$ is the largest estimate of δ computed using (76). The first-order degree sequence is used in the estimation of the shape parameter of the power law, β. $\hat{\beta}_{LL}$ is obtained by the log-log regression with Gabaix and Ibragimov (2011) correction using the OLS regression defined by (66) and its standard error is calculated by (67). $\hat{\beta}_{MLE}$ is the maximum likelihood estimate (MLE) of β computed by (68) and its standard error by (69). 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 $\hat{\beta}_{LL}$ 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 16: Yearly estimates of the degree of dominance, δ, 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)

			$\widehat{\delta}_{(1)}$ based on the inverse of $\widehat{\beta}$ using the second-order interconnections							
			Inverse of $\widehat{\beta}_{LL}$			Inverse of $\widehat{\beta}_{MLE}$			Inverse of $\widehat{\beta}_{CSN}$	
			Assumed cut-off value			Assumed cut-off value				Estimated
Year	N	$\widehat{\delta}_{(1)}$	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%
			(0.204)	(0.144)	(0.117)	(0.144)	(0.131)	(0.129)	(0.115)	
1977	524	0.778	0.718	0.870	1.008	0.821	1.058	1.351	0.750	9.4%
			(0.196)	(0.138)	(0.113)	(0.139)	(0.126)	(0.131)	(0.143)	
1982	529	0.788	0.773	0.913	1.013	0.885	1.028	1.329	1.088	23.6%
			(0.194)	(0.137)	(0.112)	(0.137)	(0.113)	(0.119)	(0.089)	
1987	510	0.804	0.686	0.879	1.031	0.883	1.070	1.325	1.110	22.9%
			(0.198)	(0.140)	(0.114)	(0.140)	(0.120)	(0.121)	(0.092)	
1992	476	0.824	0.661	0.869	1.012	0.750	1.014	1.277	0.818	12.2%
			(0.204)	(0.145)	(0.118)	(0.144)	(0.139)	(0.142)	(0.131)	
1997[a]	474	0.778	0.632	0.790	0.955	0.648	1.100	1.202	0.666	12.0%
			(0.206)	(0.145)	(0.119)	(0.146)	(0.174)	(0.156)	(0.132)	
2002	417	0.765	0.620	0.768	0.954	0.721	0.998	1.245	0.772	13.4%
			(0.218)	(0.155)	(0.126)	(0.154)	(0.152)	(0.154)	(0.134)	

Notes: This table differs from Table 15 in that the second-order degree sequence is used to produce the estimates of β. See the notes to Table 15 for further details.

Table 17: 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	0.639	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	0.633
1992	451	0.823	0.678	0.677	0.646	0.631
1997^a	452	0.775	0.725	0.636	0.622	0.597
2002	408	0.758	0.743	0.640	0.564	0.560
2007	365	0.722	0.649	0607	0.591	0.551

Notes: Estimates are obtained using the input-output accounts data downloaded from the Bureau of Economic Analysis (BEA) website. See Appendix E for details of the data sources and their transformations. The table reports the five largest yearly estimates of δ, denoted by $\hat{\delta}_{(1)}, \hat{\delta}_{(2)}, \cdots, \hat{\delta}_{(5)}$. 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 18: 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 17 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 19: 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	
$\widehat{\delta}_{(1)}$	0.763	Wholesale trade	0.716	Wholesale trade
	(0.037)		(0.045)	
$\widehat{\delta}_{(2)}$	0.667	Blast furnaces	0.683	Management of companies
	(0.037)	and steel mills	(0.045)	and enterprises
$\widehat{\delta}_{(3)}$	0.642	Real estate	0.610	Real estate[a]
	(0.037)		(0.045)	
$\widehat{\delta}_{(4)}$	0.605	Trucking and courier	0.598	Iron and steel forging
	(0.037)	services, except air	(0.045)	
$\widehat{\delta}_{(5)}$	0.605	Miscellaneous business	0.595	Other real estate[a]
	(0.037)	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 estimates of δ, denoted by $\hat{\delta}_{(1)}, \hat{\delta}_{(2)}, \cdots, \hat{\delta}_{(5)}$, and the identities of the associated sectors. The standard errors are given in parentheses and computed using (89). 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.

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. The matrix $\mathbf{A}$ as described is known as a (right) stochastic matrix or transition matrix. The proof of $\lambda_1(\mathbf{A}) = 1$ can be found in, for example, Property 10.1.2 in Stewart (2009). Given that $|\rho| < 1$ and $\lambda_1(\mathbf{A}) = 1$, it is readily seen that all eigenvalues of $\mathbf{I}_N - \rho \mathbf{A}$ are positive in absolute value, and hence the invertibility follows. $\square$

Remark 5. *It should be noted that this lemma holds irrespective of the column sums of $\mathbf{A}$ being bounded or not. 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. *Suppose that $\mathbf{A}$ is an $N \times N$ matrix that satisfies $\|\mathbf{A}\|_\infty \leq 1$ and $\|\mathbf{A}\|_1 < K < \infty$, where K is a finite positive constant. Then $\mathbf{B}^{-1}$ has bounded row and column norms, where $\mathbf{B} = \mathbf{I}_N - \rho \mathbf{A}$ and $|\rho| < 1$.*

Proof. Given that $\|\rho \mathbf{A}\|_\infty \leq 1$, $\mathbf{B}^{-1}$ can be expanded as[22]

$$\mathbf{B}^{-1} = \sum_{k=0}^{\infty} (\rho \mathbf{A})^k,$$

and then

$$
\begin{aligned}
\left\| \mathbf{B}^{-1} \right\|_\infty &\leq \left\| \mathbf{I}_N + \rho \mathbf{A} + \rho^2 \mathbf{A}^2 + \cdots \right\|_\infty \\
&\leq 1 + |\rho| \|\mathbf{A}\|_\infty + |\rho|^2 \|\mathbf{A}\|_\infty^2 + \cdots \\
&= \frac{1}{1 - |\rho| \|\mathbf{A}\|_\infty} < K < \infty.
\end{aligned}
$$

Similarly we can show that $\left\| \mathbf{B}^{-1} \right\|_1 < K < \infty$. $\square$

B   Appendix: Derivation of the summability condition

Since $\kappa_i > 0$ for all i, then using (30) we have

$$
\begin{aligned}
S_N &\leq \kappa_{\max} N^{-1} \sum_{i=1}^{N} N^{\delta_i} \\
&= \kappa_{\max} N^{-1} \sum_{i=1}^{N} e^{\delta_i \ln N}, && \text{(B.1)}
\end{aligned}
$$

---

[22]See, for example, Horn and Johnson (2012) Corollary 5.6.16.

where $\kappa_{\max} = \max_i(\kappa_i)$. Now using the Taylor series expansion of $e^{\delta_i \ln N}$ we have

$$
\begin{aligned}
\sum_{i=1}^{N} e^{\delta_i \ln N} &= \sum_{i=1}^{N} \left[1 + \sum_{s=1}^{\infty} \frac{\delta_i^s (\ln N)^s}{s!} \right] \\
&= N + \sum_{s=1}^{\infty} \frac{(\ln N)^s}{s!} \left(\sum_{i=1}^{N} \delta_i^s \right).
\end{aligned}
\tag{B.2}
$$

As $0 \le \delta_i < 1$, it also follows that

$$
\sum_{i=1}^{N} \delta_i^s \le \sum_{i=1}^{N} \delta_i, \text{ for } s \ge 1.
\tag{B.3}
$$

Using (B.2) and (B.3) in (B.1) we have

$$
S_N = N^{-1} \sum_{i=1}^{N} \kappa_i e^{\delta_i \ln N} \le \kappa_{\max} \left[1 + \left(\sum_{i=1}^{N} \delta_i \right) N^{-1} \sum_{s=1}^{\infty} \frac{(\ln N)^s}{s!} \right].
$$

However, since

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

it then follows that

$$
S_N \le \kappa_{\max} \left(1 + \frac{N-1}{N} \sum_{i=1}^{N} \delta_i \right),
\tag{B.4}
$$

and S_N is bounded in N if $\sum_{i=1}^{N} \delta_i$ is bounded in N, as required.

C   Appendix: Multiple dominant units

This appendix extends the analysis in Section 3 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. Let $N = m + n$ denote the total number of units, and then the network equation system,

$$
\mathbf{x}_t = \rho \mathbf{W} \mathbf{x}_t + \mathbf{g}_t,
$$

can be partitioned as

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

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 $\mathbf{g}_{1t} = (g_{1t}, g_{2t}, ..., g_{mt})'$, $\mathbf{g}_{2t} = (g_{m+1,t}, g_{m+2,t}, ..., g_{Nt})'$, where g_{it} is given by

$$g_{it} = -b_i - (1-\rho)(\gamma_i f_t + \varepsilon_{it}),$$

for $i = 1, 2, ..., N$, as in (36). As $\varrho(\mathbf{W}_{22}) \leq 1$ and $|\rho| < 1$, we have

$$\mathbf{x}_{2t} = (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}(\rho\mathbf{W}_{21}\mathbf{x}_{1t} + \mathbf{g}_{2t}). \tag{C.1}$$

Substituting (C.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

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

where

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

$\mathbf{S}_{22}^{-1} = (\mathbf{I}_n - \rho\mathbf{W}_{22})^{-1}$, and $\mathbf{\Phi}_1$ 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 (C.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

$$\overline{\mathbf{x}}_{Nt} = -a_N + \boldsymbol{\theta}_N'\mathbf{x}_{1t} - (1-\rho)\psi_N f_t - (1-\rho)\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t},$$

where

$$a_N = N^{-1}\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{b}_2,$$

$$\boldsymbol{\theta}_N' = N^{-1}\left[\boldsymbol{\tau}_m' + \rho\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\right],$$

$$\psi_N = N^{-1}\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2,$$

$$\mathbf{v}_N' = N^{-1}\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1},$$

with $\mathbf{b}_2 = (b_{m+1}, b_{m+2}, ..., b_N)'$ and $\boldsymbol{\gamma}_2 = (\gamma_{m+1}, \gamma_{m+2}, ..., \gamma_N)'$.

We will derive $Var(\overline{\mathbf{x}}_{Nt})$ and inspect its asymptotic order of magnitude as $N \to \infty$ following similar steps to those in the main text. First, the same arguments apply for $a_N = O(1)$ and

$$Var\left(\mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) = N^{-1}\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}^{-1'}\boldsymbol{\tau}_n = \ominus\left(N^{-1}\right),$$

where $\mathbf{V}_{22,\varepsilon} = diag\left(\sigma_{m+1}^2, \sigma_{m+2}^2, ..., \sigma_N^2\right)$. Then regarding the effects from the dominant units, using (C.2) we have

$$Cov\left(\mathbf{x}_{1t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) = Cov\left(\mathbf{\Phi}_1^{-1}\mathbf{g}_{1t} + \rho\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) \tag{C.4}$$
$$= -N^{-1}\rho(1-\rho)\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}^{-1'}\boldsymbol{\tau}_n,$$

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

and

$$
\begin{aligned}
Var\left(\mathbf{x}_{1t}\right) &= (1-\rho)^2\boldsymbol{\Phi}_1^{-1}\mathbf{V}_{11,\varepsilon}\boldsymbol{\Phi}_1^{-1\prime} + (1-\rho)^2\rho^2\boldsymbol{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{V}_{22,\varepsilon}\mathbf{S}_{22}^{-1\prime}\mathbf{W}_{12}'\boldsymbol{\Phi}_1^{-1\prime} \quad \text{(C.5)}\\
&\quad +(1-\rho)^2Var(f_t)\left(\boldsymbol{\Phi}_1^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1'\boldsymbol{\Phi}_1^{-1\prime} + \rho^2\boldsymbol{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\boldsymbol{\gamma}_2\boldsymbol{\gamma}_2'\mathbf{S}_{22}^{-1\prime}\mathbf{W}_{12}'\boldsymbol{\Phi}_1^{-1\prime}\right),
\end{aligned}
$$

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

Now, we are ready to examine $Var(\overline{\mathbf{x}}_{Nt})$, which is given by

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

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

$$Var\left(\overline{\mathbf{x}}_{Nt}\right) = \boldsymbol{\theta}_N'Var(\mathbf{x}_{1t})\boldsymbol{\theta}_N - 2(1-\rho)\boldsymbol{\theta}_N'Cov\left(\mathbf{x}_{1t}, \mathbf{v}_N'\boldsymbol{\varepsilon}_{2t}\right) + \ominus(N^{-1}). \quad \text{(C.6)}$$

Consider the ith element of $\boldsymbol{\theta}_N$, $\theta_{i,N}$, for $i = 1, 2, ..., m$, and suppose that m is fixed and does not rise with N, then by definition,

$$\theta_{i,N} = N^{-1}\left(1 + \rho\boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1}\mathbf{w}_{\cdot i,21}\right),$$

where $\mathbf{w}_{\cdot i,21}$ is the ith column of $\mathbf{W}_{21}$. Let $\boldsymbol{\phi}_n' = \boldsymbol{\tau}_n'\mathbf{S}_{22}^{-1} = (\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$, then

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

where $\phi_{\min} = \min(\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$ and $\phi_{\max} = \max(\phi_{m+1}, \phi_{m+2}, ..., \phi_N)$, and it follows that

$$N^{-1} + \phi_{\min}N^{-1}\sum_{j=1}^n w_{ji,21} \leq \theta_{i,N} \leq N^{-1} + \phi_{\max}N^{-1}\sum_{j=1}^n w_{ji,21}. \quad \text{(C.7)}$$

Repeating the same argument as for the case of a single dominant unit, we readily have

$1 < \phi_{\min} \leq \phi_{\max} < K < \infty$. Also note that $\mathbf{w}_{\cdot i,21}'\boldsymbol{\tau}_n = \ominus(N^{\delta_i})$, which immediately follows from the definition

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

and m being fixed. Therefore, we establish by (C.7) that

$$\theta_{i,N} = \ominus(N^{\delta_i - 1}), \text{ for } i = 1, 2, ..., m,$$

and then

$$\boldsymbol{\theta}_N'\boldsymbol{\theta}_N = \ominus(N^{2\delta_{\max} - 2}),$$

where $0 < \delta_{\max} = \max(\delta_1, \delta_2, ..., \delta_m) \leq 1$.

Further notice that

$$\boldsymbol{\theta}'_N \boldsymbol{\theta}_N \lambda_m \left[Var(\mathbf{x}_{1t})\right] \leq \boldsymbol{\theta}'_N Var(\mathbf{x}_{1t})\boldsymbol{\theta}_N \leq \boldsymbol{\theta}'_N \boldsymbol{\theta}_N \lambda_1 \left[Var(\mathbf{x}_{1t})\right],$$

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

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

Turning now to the jth element of the covariance term, for $j = 1, 2, ..., m$, we have

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

where $\boldsymbol{\Phi}^{-1}_{j.,1}$ denotes the jth row of $\boldsymbol{\Phi}^{-1}_1$, and using similar line of reasoning as in the main text it is easily verified that

$$\left|Cov\left(x_{jt}, \mathbf{v}'_N \boldsymbol{\varepsilon}_{2t}\right)\right| = O\left(N^{-1}\right),$$

and therefore

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

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 $\psi_N = \ominus(N^{\delta_\gamma-1})$ 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 (61) 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.

D   Appendix: Consistency of $\hat{\delta}_{(1)}$

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

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

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

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

Taking the expectation and variance of $\hat{\delta}_{(1)}$ given by (86), and making use of the above results we now have

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

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

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

$$\lim_{N\to\infty}\frac{\sum_{j=1}^{N}j^{-1}}{\ln N} = 1. \quad \text{(D.4)}$$

Using (D.1) and (D.4) in (D.2) we now have

$$\lim_{N\to\infty}E\left(\hat{\delta}_{(1)}\right) = 1/\beta.$$

Turning to the variance of $\hat{\delta}_{(1)}$, we note that

$$Var\left(\hat{\delta}_{(1)}\right) = \frac{Var\left(z_{(1)}\right) + \left(\frac{1-2N}{N^2}\right)\sum_{i=1}^{N}Var(z_{(i)})}{(\ln N)^2},$$

$$(\ln N)^{-2}Var\left(z_{(1)}\right) \leq Var\left(\hat{\delta}_{(1)}\right) \leq (\ln N)^{-2}\left[Var\left(z_{(1)}\right) + \left(\frac{2N-1}{N^2}\right)NVar\left(z_{(1)}\right)\right],$$

$$(\ln N)^{-2}\delta^2\left(\sum_{j=1}^{N}\frac{1}{j^2}\right) \leq Var\left(\hat{\delta}_{(1)}\right) \leq (\ln N)^{-2}\left(\frac{3N-1}{N}\right)\delta^2\left(\sum_{j=1}^{N}\frac{1}{j^2}\right).$$

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

E   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**.