ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Econometric analysis of production networks with dominant units[☆]

M. Hashem Pesaran[a,b], Cynthia Fan Yang[c,*]

[a] University of Southern California, USA
[b] Trinity College, Cambridge, UK
[c] Florida State University, USA

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C12
C13
C23
C67
E32

Keywords:
Aggregate fluctuations
Strongly and weakly dominant units
Spatial models
Outdegrees
Degree of pervasiveness
Extremum estimator
Power law
Input–output tables
US economy

ABSTRACT

This paper introduces the notions of strongly and weakly dominant units for networks, and shows that pervasiveness of shocks to a network is measured by the degree of dominance of its most pervasive unit; shown to be equivalent to the inverse of the shape parameter of the power law fitted to the network outdegrees. New cross-section and panel extremum estimators of the degree of dominance in networks are proposed, and their asymptotic properties investigated. The small sample properties of the proposed estimators are examined by Monte Carlo experiments, and their use is illustrated by an empirical application to US input–output tables.

1. Introduction

Over the past decade, there has been renewed interest in production networks and the role that individual production units (firms/sectors) can play in propagation of shocks across the economy. This literature builds on the multisectoral model of real business cycles pioneered by Long and Plosser (1983), and draws from a variety of studies on social and economic networks, including network games, cascades, and micro foundations of macro volatility. Notable theoretical contributions in this area include Acemoglu et al. (2012), Horvath (1998, 2000), Gabaix (2011), Acemoglu et al. (2016b), and Siavash (2018). Empirical evidence for such propagation mechanism is presented in Foerster et al. (2011), Acemoglu et al. (2016a), and Carvalho et al. (2016). One important issue in this literature relates to conditions under which

☆ We acknowledge helpful comments from the guest editor (Jiti Gao) and the three anonymous reviewers, Daron Acemoglu, Vasco Carvalho, Ron Smith, Martin Weidner, and participants at the Max King Conference at Monash University, the 2017 Asian Meeting of the Econometric Society, the third Vienna Workshop on High Dimensional Time Series in Macroeconomics and Finance, the fourth annual conference of the International Association for Applied Econometrics (IAAE), and seminar participants at University of Southern California and Emory University. We would also like to thank Ida Johnsson for help with the compilation of input–output tables. Yang gratefully acknowledges the IAAE travel grant. Part of this paper was written when Yang was a doctoral student at the University of Southern California.

* Correspondence to: Department of Economics, Florida State University, 281 Bellamy Building, Tallahassee, FL 32306, USA.
E-mail addresses: pesaran@usc.edu (M.H. Pesaran), cynthia.yang@fsu.edu (C.F. Yang).

https://doi.org/10.1016/j.jeconom.2020.03.014
0304-4076/© 2020 Elsevier B.V. All rights reserved.

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 et al. (2016c).

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

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

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

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

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

Notations. The total number of cross section units (sectors) in the economy is denoted by N, which is then decomposed into m dominant units and n non-dominant units. The number of dominant units is also decomposed into strongly dominant units and weakly dominant units. (See Definition 1). δ_i denotes the degree of dominance (or pervasiveness) of unit i in a network, where $i = 1, 2, \ldots, N$, and $0 \leq \delta_i \leq 1$. If $\{f_N\}_{N=1}^{\infty}$ is any real sequence and $\{g_N\}_{N=1}^{\infty}$ is a sequence of positive real numbers, then $f_N = O(g_N)$ if there exists a positive finite constant K such that $|f_N|/g_N \leq K$ for all N. $f_N = o(g_N)$ if $f_N/g_N \to 0$, as $N \to \infty$. If $\{f_N\}_{N=1}^{\infty}$ and $\{g_N\}_{N=1}^{\infty}$ are both positive sequences of real numbers, then $f_N = \ominus(g_N)$ if there exists $N_0 \geq 1$ and positive finite constants K_0 and K_1, such that $\inf_{N \geq N_0}(f_N/g_N) \geq K_0$, and $\sup_{N \geq N_0}(f_N/g_N) \leq K_1$. $\varrho(\mathbf{A})$ is the spectral radius of the $N \times N$ matrix $\mathbf{A} = \left(a_{ij}\right)$, defined as $\varrho(\mathbf{A}) = \max\{|\lambda_i|, i = 1, 2, \ldots, N\}$**,** where λ_i is an eigenvalue of $\mathbf{A}$ and $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq \cdots \geq |\lambda_N(\mathbf{A})|$. $\|\mathbf{A}\|_\infty = \max\limits_{1 \leq i \leq N} \sum_{j=1}^{N} \left|a_{ij}\right|$ and $\|\mathbf{A}\|_1 = \max\limits_{1 \leq j \leq N} \sum_{i=1}^{N} \left|a_{ij}\right|$ are the maximum row sum norm and the maximum column sum norm of matrix $\mathbf{A}$, respectively. Generic finite positive constants are denoted by K and C_i for $i = 0, 1, 2, \ldots$.

2. Production network

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

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

where l_{it} is the labor input, $q_{ij,t}$ is the amount of output of sector j used in production of sector i, w_{ij} is the share of sector j's output in the total intermediate input use of sector i, and ρ is capital's share of output ($0 < \rho < 1$). We assume that $w_{ij} \geq 0$ for all i and j, and the input shares of all sectors sum up to one, namely, $\sum_{j=1}^{N} w_{ij} = 1$, for all $i = 1, 2, \ldots, N$. Finally, u_{it} is the productivity shock to sector i, composed of common and idiosyncratic components. Specifically, u_{it} is decomposed into r (r is finite) common factors $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{rt})'$, with factor loadings $\boldsymbol{\gamma}_i$, and a sector-specific shock, ε_{it}:

$$u_{it} = \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it}. \tag{2}$$

Following the literature, and without loss of generality, we shall assume that $\mathbf{f}_t$ and ε_{it} are uncorrelated. Examples of common factors include common technological shocks, regulatory changes and organizational innovations that may affect production in all sectors. The factor loadings, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ir})'$, for $i = 1, 2, \ldots, N$, are fixed constants that measure the relative importance of the common factors for sector i. Following Bailey et al. (2016), we use α_ℓ, for $\ell = 1, 2, \ldots, r$, to denote the cross-section exponent of $\gamma_{i\ell}$, which measures the degree of pervasiveness of factor $f_{\ell t}$, over the N sectors in the economy. More specifically, α_ℓ is defined by

$$\sum_{i=1}^{N} |\gamma_{i\ell}| = \ominus\left(N^{\alpha_\ell}\right), \tag{3}$$

with $0 \leq \alpha_\ell \leq 1$. The standard factor model sets $\alpha_\ell = 1$, and treats the common factors as "strong" or "fully pervasive", in the sense that changes in $f_{\ell t}$ affect all sectors of the economy. But in what follows we shall also consider cases where one or more of the factors are weak in the sense that $\alpha_\ell < 1$ for some ℓ. If $\alpha_{\max} = \max(\alpha_1, \alpha_2, \ldots, \alpha_r) = 1$, there exists at least one "strong" or "pervasive" factor. If $\alpha_{\max} < 1$, factors are weak but could be influential if $\alpha_{\max}$ is close to unity. Following Acemoglu et al. (2012), we shall assume that the sector-specific shocks are cross-sectionally independent with zero means and finite variances, $Var(\varepsilon_{it}) = \sigma_i^2$, such that $0 < \underline{\sigma}^2 < \sigma_i^2 < \bar{\sigma}^2 < K < \infty$. The independence assumption is not necessary and can be relaxed by assuming that ε_{it} are cross-sectionally weakly dependent. We also assume that ε_{it} are serially uncorrelated, although this is not essential either for our main theoretical results.

The amount of final goods, c_{it}, are defined by

$$c_{it} = q_{it} - \sum_{j=1}^{N} q_{ji,t}, \quad i = 1, 2, \ldots, N, \tag{4}$$

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

$$u\left(c_{1t}, c_{2t}, \ldots, c_{Nt}\right) = A \prod_{i=1}^{N} c_{it}^{1/N}, \ A > 0. \tag{5}$$

We further assume that the aggregate labor supply, l_t, is given exogenously and labor markets clear, $l_t = \sum_{i=1}^{N} l_{it}$.

Let $P_{1t}, P_{2t}, \ldots, P_{Nt}$ be the sectoral equilibrium prices, $Wage_t$ the equilibrium wage rate, and denote their logarithms by $p_{it} = \log(P_{it})$, $\omega_t = \log(Wage_t)$. Then it can be shown by using similar arguments as in Appendix A of Acemoglu et al. (2012) that in the competitive equilibrium the logarithm of real wage, which is taken as a measure of GDP (or real value added), is given by

$$\omega_t - \bar{p}_t = \mu + \boldsymbol{v}_N' \mathbf{u}_t, \tag{6}$$

Please cite this article as: M.H. Pesaran and C.F. Yang, Econometric analysis of production networks with dominant units. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.03.014.

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

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

$\mathbf{p}_t = \left(p_{1t}, p_{2t,}, \ldots, p_{Nt}\right)'$, $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})'$, and

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

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

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

The (log) real-wage equation, (6), generalizes equation (3) in Acemoglu et al. (2012) by allowing for time variations in prices. By normalizing $\bar{p}_t$ such that $\bar{p}_t = -\mu$ and ignoring common factors, Acemoglu et al. (2012) concentrate on $\omega_t = \boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t$ as a measure of aggregate output, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. The authors refer to $\boldsymbol{\upsilon}_N$ as the "influence vector" (their Eq. (4)), and show that $\upsilon_i \geq 0$, and $\mathbf{1}_N'\boldsymbol{\upsilon}_N = 1$.[1] They measure aggregate volatility by the standard deviation of aggregate output, namely $\left[Var\left(\boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t\right)\right]^{1/2}$, and focus on the asymptotic properties of $\boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t$, as $N \to \infty$. Since $Var\left(\boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t\right) = \boldsymbol{\upsilon}_N' Var\left(\boldsymbol{\varepsilon}_t\right)\boldsymbol{\upsilon}_N$, it follows that

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

and hence the asymptotic properties of $Var\left(\boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t\right)$ is governed by $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$. The same conclusion also follows if we allow for common factors as in (2) so long as the factors are weak, in the sense that $\alpha_{\max} = \max(\alpha_1, \alpha_2, \ldots, \alpha_r) < 1$.[2] This result follows by noting that in the presence of common factors

$$\left(\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N\right)\lambda_N\left(\boldsymbol{\Sigma}_{u,N}\right) \leq Var\left(\boldsymbol{\upsilon}_N'\mathbf{u}_t\right) \leq \left(\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N\right)\lambda_1\left(\boldsymbol{\Sigma}_{u,N}\right),$$

where $\lambda_N\left(\boldsymbol{\Sigma}_{u,N}\right)$ and $\lambda_1\left(\boldsymbol{\Sigma}_{u,N}\right)$ are, respectively, the smallest and largest eigenvalues of $\boldsymbol{\Sigma}_{u,N} = E\left(\mathbf{u}_t\mathbf{u}_t'\right)$, and $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{Nt})' = \boldsymbol{\Gamma}_i\mathbf{f} + \boldsymbol{\varepsilon}_t$, with $\boldsymbol{\Gamma}_i = \left(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N\right)'$. In the case where $\mathbf{u}_t$ is cross-sectionally weakly dependent, all eigenvalues of $\boldsymbol{\Sigma}_{u,N}$ will be bounded in N and the asymptotic behavior of $Var\left(\boldsymbol{\upsilon}_N'\mathbf{u}_t\right)$ continues to be determined by that of $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$.

Acemoglu et al. (2012, p.2009) derive a lower bound for $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$ and show that[3]

$$\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N \geq c_0 N^{-1} + c_1 N^{-2}\sum_{j=1}^{N} d_j^2, \tag{9}$$

where c_0 and c_1 are finite constants that do not depend on N, and d_j is the outdegree of the jth unit defined by $d_j = \sum_{i=1}^{N} w_{ij}$. In their analysis, Acemoglu et al. (2012) consider the limiting behavior of $N^{-2}\sum_{j=1}^{N} d_j^2$. But as we shall see below, 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. To fully understand the limiting behavior of $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$ we also need to investigate the limiting properties of the upper bound to $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$ which is not addressed by Acemoglu et al. (2012).

3. Price network as a dual to the production network

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

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

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

[2] See Chudik et al. (2011) Theorem 3.1.

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

and

$$l_{it} = \frac{(1-\rho)\, P_{it} q_{it}}{Wage_t}. \tag{11}$$

Substituting the above results in (1) and simplifying yields

$$p_{it} = \rho \sum_{j=1}^{N} w_{ij} p_{jt} + (1-\rho)\, \omega_t - b_i - (1-\rho)\, u_{it}, \tag{12}$$

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

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

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

$$\mathbf{p}_t = \rho \mathbf{W} \mathbf{p}_t + (1-\rho)\, \omega_t \mathbf{1}_N - [\mathbf{b} + (1-\rho)\, \mathbf{u}_t], \tag{14}$$

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

A dual to the price equation in (12) can also be obtained using (10) in (4) to obtain

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

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

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

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

Aggregating (11) over i, we have

$$Wage_t \sum_{i=1}^{N} l_{it} = (1-\rho) \sum_{i=1}^{N} P_{it} q_{it},$$

or

$$l_t Wage_t = (1-\rho) \sum_{i=1}^{N} S_{it} = (1-\rho)\, \mathbf{1}_N' \mathbf{S}_t. \tag{17}$$

Also using (16)

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

where $\left(\mathbf{I}_N - \rho \mathbf{W}'\right)^{-1}$ is known as the Leontief inverse.[4] Using (18) in (17) now yields the following expression for the total wage bill,

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

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

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

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

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

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

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

where $\boldsymbol{\upsilon}_N$ is the influence vector given by (8). Now let $\mathbf{x}_t = \mathbf{p}_t - \omega_t \mathbf{1}_N$, and rewrite (14) in terms of log price-wage ratios, $\mathbf{x}_t$, as

$$\mathbf{x}_t = \rho \mathbf{W} \mathbf{x}_t - \mathbf{b} - (1-\rho)\, \mathbf{u}_t. \tag{23}$$

Eq. (23) represents a first-order spatial autoregressive (SAR(1)) model.

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

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

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

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

$$\omega_t - \bar{p}_t = -\bar{x}_{N,t} = (1-\rho)^{-1}\left(\boldsymbol{\upsilon}_N'\mathbf{b}\right) + \boldsymbol{\upsilon}_N'\mathbf{u}_t, \tag{24}$$

which fully specifies the dependence of aggregate output on the productivity shocks.

Note that Eqs. (19) and (24) are dual of each other. (19) gives the total wage bill in terms of a weighted sum of consumption expenditures, with the weights given by $(1-\rho)(\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{1}_N$, whilst (24) gives the log of the real wage rate in terms of the aggregate shocks. Recall that $\mathbf{u}_t = \boldsymbol{\Gamma}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\Gamma} = \left(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N\right)'$, and common and sectoral shocks are assumed to be uncorrelated. The key issue is how much of the cyclical fluctuations in (log) real wages, $Var\left(\bar{x}_{N,t}\right)$, is due to common shocks, $Var\left(\boldsymbol{\upsilon}_N'\boldsymbol{\Gamma}\mathbf{f}_t\right)$, and how much is due to sectoral shocks, $Var\left(\boldsymbol{\upsilon}_N'\boldsymbol{\varepsilon}_t\right)$.

There are two advantages in directly focusing on the price network model, (23). First, it allows us to relate the production network to the literature on spatial econometrics that should facilitate the econometric analysis of production networks, and allows us to address more easily the issues of identification and estimation of the structural parameters, including capital's share ρ, factor loadings $\boldsymbol{\gamma}_i$ and error variances σ_i^2, for $i = 1, 2, \ldots, N$.[5] The direct use of the SAR model, (23), also enables us to provide exact bounds on $Var\left(\bar{x}_{N,t}\right) = Var\left(\omega_t - \bar{p}_t\right)$ rather than the lower bounds obtained by Acemoglu et al. (2012). Instead, by considering the price network explicitly we are able to show that at most only a few sectors can have significant aggregate effects, and these sectors are those with outdegrees that rise with N. The rate at which the outdegrees rise with N could differ across sectors and it is important that such sectors are identified and their empirical contribution to aggregate fluctuations evaluated.

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

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

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

$$d_j = \kappa_j N^{\delta_j}, \text{ for } j = 1, 2, \ldots, N, \tag{25}$$

where κ_j is a strictly positive random variable defined on $0 < \underline{\kappa} \leq \kappa_j \leq \bar{\kappa} < K$, where $\underline{\kappa}$ and $\bar{\kappa}$ are fixed constants. The unit j is said to be *strongly dominant* if $\delta_j = 1$, *weakly dominant* if $0 < \delta_j < 1$, and *non-dominant* if $\delta_j = 0$. We refer to δ_j as the degree of dominance of unit j in the network.

Remark 1. It is worth noting that δ_j is identified by requiring that κ_j is a strictly positive random variable bounded in N, and δ_j is a fixed constant that does not vary with N.

In the standard case where the column sum of $\mathbf{W}$ is bounded in N we must have $\delta_j = 0$ for all j, that is, all units are non-dominant. $\mathbf{W}$ will have an unbounded column sum if $\delta_j > 0$ for at least one j. But due to the bounded nature of the rows of $\mathbf{W}$, not all columns of $\mathbf{W}$ can be δ-dominant with $\delta_j > 0$ for all j. To see this, let $\mathbf{d} = (d_1, d_2, \ldots, d_N)' = \mathbf{W}'\mathbf{1}_N$, and note that

$$\mathbf{1}_N'\mathbf{d} = \mathbf{1}_N'\mathbf{W}'\mathbf{1}_N = N. \tag{26}$$

[5] For example, see the recent contributions of Bai and Li (2013) and Yang (2020) on estimation of SAR models with unobserved common factors.

Please cite this article as: M.H. Pesaran and C.F. Yang, Econometric analysis of production networks with dominant units. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.03.014.

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

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

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

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

which in turn implies

$$\underline{\kappa} N^{\delta_{\min}} \leq 1. \tag{27}$$

Since by assumption $\underline{\kappa} > 0$ and $\delta_{\min} \geq 0$, it is clear that (27) cannot be satisfied for all values of N unless $\delta_{\min} = 0$, which establishes that not all units in a given network can be dominant. This result is summarized in the following proposition.

Proposition 1. *Consider the network represented by* $\mathbf{W} = (w_{ij} \geq 0)$*, and assume that* $\mathbf{W}$ *is row-standardized. Suppose that the outdegrees of the network,* $d_j = \sum_{i=1}^{N} w_{ij}$*, are non-zero (*$d_j > 0$*) and follow the power function, (25), with* δ_j *being the degree of dominance of unit j in the network. Then not all units of the network can be* δ*-dominant, with* $\delta_j > 0$ *for all j.*

Let $S_N = N^{-1} \sum_{j=1}^{N} \kappa_j N^{\delta_j}$, and note that since $\underline{\kappa} > 0$ and hence

$$S_N = N^{-1} \sum_{j=1}^{N} \kappa_j e^{\delta_j \ln N} \geq \underline{\kappa} N^{-1} \sum_{j=1}^{N} e^{\delta_j \ln N}. \tag{28}$$

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

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

which if substituted in (28) yields

$$S_N \geq \underline{\kappa} \left[1 + \sum_{s=1}^{\infty} \frac{\left(\sum_{j=1}^{N} \delta_j^s \right) (\ln N)^s}{s! N} \right]. \tag{29}$$

Since $S_N = 1$, and all the summands over s in (29) are nonnegative as $\delta_j \geq 0$ and $\ln N > 0$, it is necessary that

$$\frac{\left(\sum_{j=1}^{N} \delta_j^s \right) (\ln N)^s}{s! N} \to 0, \text{ as } N \to \infty, \text{ for all } s = 1, 2, 3, \dots. \tag{30}$$

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

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

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

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

and

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

Hence, for conditions in (30) to be met it is sufficient that $\left\{ \delta_j \right\}$ satisfies the following summability assumption.

Assumption 1. The degrees of dominance of all units in a network, $\left\{ \delta_j, j = 1, 2, \dots, N \right\}$, are summable, namely,

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

As we shall see, Assumption 1 plays an important role in the proof of consistency of the extremum estimator proposed in Section 6.2.

Suppose now that m units are strongly dominant with $\delta_j = 1$, and the rest are non-dominant with $\delta_j = 0$. Then using (29) we have

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

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

In the case where m units are dominant with $\delta_j > 0$, then m must be finite if the summability condition given by (31) is to hold. For example, suppose that only m units are dominant. Then $\sum_{j=1}^{N}\delta_j \geq m\delta_{\min} > 0$, and from the summability condition (31) we have $K > \sum_{j=1}^{N}\delta_j \geq m\delta_{\min}$, from which it follows that $m \leq K/\delta_{\min}$ and in consequence m must be bounded in N. These findings are summarized in the next proposition.

Proposition 2. *Consider the network represented by* $\mathbf{W} = (w_{ij} \geq 0)$*, and assume that* $\mathbf{W}$ *is row-standardized, and the outdegrees of the network,* $d_j = \sum_{i=1}^{N} w_{ij}$*, are non-zero* ($d_j > 0$)*. Then the number of strongly dominant units must be fixed and cannot rise with N. Moreover, under Assumption 1 the number of dominant units with* $\delta_j \neq 0$ *must be finite, where* δ_j *is the degree of dominance of unit j in the network.*

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

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

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

The degree of network pervasiveness, $\delta_{\max}$, defined in Definition 2, is related to β, the shape parameter of the power law assumed by Acemoglu et al. (2012, Definition 2) for the outdegree sequence, $\{d_1, d_2, \ldots, d_N\}$. To see this, we first use the specification of the outdegrees given by (25) in (9) to obtain

$$\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N \geq c_0 N^{-1} + c_1 N^{-2}\sum_{j=1}^{m}\kappa_j^2 N^{2\delta_j} + c_1\frac{N-m}{N^2}\left(\sum_{j=m+1}^{N}\frac{\kappa_j^2}{N-m}\right),$$

where $(N-m)^{-1}\sum_{j=m+1}^{N}\kappa_j^2 = O(1)$. Also, recall that m must be finite if $\{\delta_j\}$ is summable (Proposition 2). Therefore, recalling that $\kappa_j \leq \bar{\kappa}$, then

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

and the limiting behavior of $\boldsymbol{\upsilon}_N'\boldsymbol{\upsilon}_N$ will be determined by that of $N^{2(\delta_{\max}-1)}$, namely the cross section exponent of the strongest of the dominant units, $\delta_{\max}$.

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

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

or $\delta_{\max} \geq 1/\beta - \epsilon_\beta$. Therefore, $\delta_{\max}$ can be viewed as measuring the inverse of β, a result that we formally establish in Section 6.

We are now in a position to consider the exact rate at which $Var(\bar{x}_{N,t})$ varies with N. We will show that it is governed by the pervasiveness of the network, measured by $\delta_{\max}$, and the maximum of the exponents of the factors, $\alpha_{\max} = \max(\alpha_1, \alpha_2, \ldots, \alpha_r)$, where α_ℓ ($\ell = 1, 2, \ldots, r$) is defined by (3). For unit-specific shocks to dominate the macro or common factor shocks we need $\delta_{\max} > \alpha_{\max} > 1/2$.

5. Price networks with dominant units

Consider the price network (23), and assume that it contains one dominant unit and $n = N - 1$ non-dominant units. The analysis can be readily extended to networks with m dominant units (m fixed), but to simplify the exposition here we confine our analysis to networks with one dominant unit. (The derivations for the general case is provided in Appendix B). In addition, the analysis is conducted assuming a single common factor for ease of notation and it can be easily extended to allow for multiple factors without additional complexity.

Without loss of generality, suppose that the first element of $\mathbf{x}_t$, namely x_{1t}, is the dominant unit, and write (23) in the partitioned form as (setting $w_{11}=0$)

$$\begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} 0 & \rho\mathbf{w}_{12}' \\ \rho\mathbf{w}_{21} & \rho\mathbf{W}_{22} \end{pmatrix} \begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} + \begin{pmatrix} g_{1t} \\ \mathbf{g}_{2t} \end{pmatrix}, \tag{32}$$

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

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

then $\mathbf{w}_{12}'\mathbf{1}_n = 1$, and $\mathbf{1}_n - \mathbf{w}_{21} = \mathbf{W}_{22}\mathbf{1}_n$. The latter result states that the ith row sum of $\mathbf{W}_{22}$ is given by $1 - w_{i1} \le 1$, and considering that $0 \le w_{i1} < 1$, then we must have $\|\mathbf{W}_{22}\|_\infty \le 1$, which also establishes that $\varrho(\mathbf{W}_{22}) \le 1$, where $\varrho(\mathbf{A})$ denotes the spectral radius of $\mathbf{A}$. Under the assumption that $|\rho| < 1$, by Lemma A.1 in Appendix A the system of equations (32) has a unique solution given by

$$\begin{pmatrix} x_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} = \begin{pmatrix} 1 & -\rho\mathbf{w}_{12}' \\ -\rho\mathbf{w}_{21} & \mathbf{S}_{22} \end{pmatrix}^{-1} \begin{pmatrix} g_{1t} \\ \mathbf{g}_{2t} \end{pmatrix} = \mathbf{S}^{-1}(\rho)\mathbf{g}_t, \tag{33}$$

where $\mathbf{S}_{22} = \mathbf{I}_n - \rho\mathbf{W}_{22}$. In addition, since $|\rho| \, \|\mathbf{W}_{22}\|_1 < 1$, it follows from Lemma A.2 in Appendix A that $\mathbf{S}_{22}^{-1}$ has bounded row and column norms. For future reference also note that the $(1, 1)$th element of $\mathbf{S}^{-1}(\rho)$ is given by ζ_1^{-1}, where $\zeta_1 = 1 - \rho^2\mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\mathbf{w}_{21} \neq 0$.[6] Finally, to allow unit 1 to be δ -dominant we consider the following exponent formulation

$$d_1 = \sum_{i=2}^{N} w_{i1} = \kappa_1 N^{\delta_1}, \tag{34}$$

where d_1 is allowed to rise with N, with $\kappa_1 > 0$ and $0 < \delta_1 \le 1$. Recall that κ_1 is a strictly positive random variable bounded in N, and δ_j is a fixed constant that does not vary with N.

The system of equations (32) can now be solved for $\mathbf{x}_{2t}$ in terms of x_{1t}, namely (recall that by assumption $|\rho| \, \|\mathbf{W}_{22}\|_1 < 1$)

$$\mathbf{x}_{2t} = x_{1t}\rho\left(\mathbf{S}_{22}^{-1}\mathbf{w}_{21}\right) + \mathbf{S}_{22}^{-1}\mathbf{g}_{2t}, \tag{35}$$

and

$$x_{1t} = \zeta_1^{-1}\left(g_{1t} + \rho\mathbf{w}_{12}'\mathbf{S}_{22}^{-1}\mathbf{g}_{2t}\right). \tag{36}$$

Using the above in (35), we now have

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

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

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

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

but our analysis equally applies to weighted averages, $x_{N,t}^* = \sum_{i=1}^{N} \varpi_i x_{it}$, so long as the weights ϖ_i are granular in the sense that $\varpi_i = O\left(N^{-1}\right)$. Using (35) and (36) we have

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

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

$$\bar{x}_{N,t} = N^{-1}\left[-a_n + \theta_n x_{1t} - (1-\rho)\psi_n f_t - (1-\rho)\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right], \tag{37}$$

where $a_n = \mathbf{1}_n'\mathbf{S}_{22}^{-1}\mathbf{b}_2$, $\boldsymbol{\phi}_n' = \mathbf{1}_n'\mathbf{S}_{22}^{-1}$, $\theta_n = 1 + \rho\boldsymbol{\phi}_n'\mathbf{w}_{21}$, and $\psi_n = \boldsymbol{\phi}_n'\boldsymbol{\gamma}_2$. The first term of (37), $N^{-1}a_n$, is bounded in N, since $\|\mathbf{W}_{22}\|_\infty \le 1$ and $\rho\,\|\mathbf{W}_{22}\|_1 < 1$, and as a result $\mathbf{S}_{22}^{-1}$ will have bounded row and column norms by Lemma A.2 in Appendix A. The second term captures the effect of the dominant unit. The third term is due to the common factor, f_t, and the final term represents the average effects of the micro productivity shocks. $N^{-1}\boldsymbol{\phi}_n$ is the influence vector associated with the non-dominant units. It is analogous to the influence vector defined by (8) which applies to all units.

6 In deriving (36), it is required that $\zeta_1 \neq 0$. This condition is met since the $N \times N$ matrix on the right-hand-side of (33) is non-singular.

Starting with the final term of (37), we first note that

$$\underline{\sigma}^2 N^{-2}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n \leq Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \leq \bar{\sigma}^2 N^{-2}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n, \tag{38}$$

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

$$\boldsymbol{\phi}_n' = \mathbf{1}_n' + \rho\mathbf{1}_n'\mathbf{W}_{22} + \rho^2\mathbf{1}_n'\mathbf{W}_{22}^2 + \cdots,$$

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

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

and $N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\phi}_n$ is bounded from below and above by finite non-zero constants. Using this result in (38) we also have

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

which establishes that

$$Var\left(N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = \ominus\left(N^{-1}\right), \tag{39}$$

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

Next, using (36) we have

$$Cov\left(x_{1t}, N^{-1}\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) = -\left(1-\rho\right)\rho\zeta_1^{-1}N^{-1}\mathbf{w}_{12}'\mathbf{H}_{22}\mathbf{1}_n, \tag{40}$$

and

$$Cov\left(x_{1t}, f_t\right) = -\left(1-\rho\right)\left(\zeta_1^{-1}\gamma_1 + \rho\zeta_1^{-1}h_2\right)Var\left(f_t\right), \tag{41}$$

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

$$\begin{aligned} Var\left(\bar{x}_{N,t}\right) &= N^{-2}\theta_n^2 Var\left(x_{1t}\right) - 2\left(1-\rho\right)N^{-2}\theta_n Cov\left(x_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) \\ &\quad + \left(1-\rho\right)^2 N^{-2} Var\left(\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) + \left(1-\rho\right)^2 N^{-2}\chi_n Var\left(f_t\right), \end{aligned} \tag{42}$$

where

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

Also, using (36) we have

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

which is easily seen to be bounded in N.

A number of results can now be obtained from (42). First, without a common factor and a dominant unit, $Var\left(\bar{x}_{N,t}\right) = \ominus(N^{-1})$, and the effects of idiosyncratic shocks on $\bar{x}_{N,t}$ will vanish at the rate of $N^{-1/2}$, as $N \to \infty$. This rate matches the decay rate of shocks in models without a network structure, namely even if we set $\mathbf{W} = \mathbf{0}$. Therefore, for micro shocks to have macroeconomic implications there must be at least one dominant unit in the network. To see this consider now the case where there is no common factor but the network includes a dominant unit. Then using (39) and (42) we have

$$Var\left(\bar{x}_{N,t}\right) = N^{-2}\theta_n^2 Var\left(x_{1t}\right) - 2\left(1-\rho\right)N^{-2}\theta_n Cov\left(x_{1t}, \boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right) + O\left(N^{-1}\right). \tag{43}$$

Recall that $Var(x_{1t})$ is bounded in N and $\theta_n = 1 + \rho\boldsymbol{\phi}_n'\mathbf{w}_{21}$. Consider the limiting properties of $N^{-1}\theta_n$. Since

$$N^{-1} + \phi_{\min}\rho N^{-1}d_1 \leq N^{-1}\theta_n \leq N^{-1} + \phi_{\max}\rho N^{-1}d_1, \tag{44}$$

where $1 \leq \phi_{\min} \leq \phi_{\max} < K$, then the asymptotic behavior of $N^{-1}\theta_n$ depends on the way the outdegree of the dominant unit, namely d_1, varies with N. Using the exponent specification given by (25), $d_1 = \kappa_1 N^{\delta_1}$, it follows that

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

which leads to

$$N^{-1}\theta_n = \ominus(N^{\delta_1 - 1}),\ 0 < \delta_1 \leq 1. \tag{45}$$

Consider now the second term of (43), and note from (40) that

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

since $\left\|\mathbf{w}_{12}'\right\|_\infty = \left\|\mathbf{w}_{12}\right\|_1 = \sum_{i=2}^N w_{1i} = 1$, $\left\|\mathbf{S}_{22}^{-1}\right\|_\infty < K$, $\left\|\mathbf{V}_{22,\varepsilon}\right\|_\infty = \bar{\sigma}^2 < K$, and $\left\|\boldsymbol{\phi}_n\right\|_\infty = \phi_{\max} < K$. Using the above results in (43) we have

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

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

$$Var\left(\bar{x}_{N,t}\right) = \ominus(N^{2\delta_1-2}) + O\left(N^{-1}\right), \tag{46}$$

and hence

$$Var\left(\bar{x}_{N,t}\right) = \ominus(N^{2\delta_1-2}), \text{ if } \delta_1 > 1/2. \tag{47}$$

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

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

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

where $N^{-1}\rho\mathbf{1}_n'\mathbf{w}_{21} = \rho N^{-1}d_1$ represents the effects of the first-order network connections on θ_N, $N^{-1}\rho^2\mathbf{1}_n'\mathbf{W}_{22}\mathbf{w}_{21}$, the effects of the second-order network connections and so on. But in view of (44) and (45) all these higher order interconnections (individually and together) at most behave as $\ominus(N^{\delta_1-1})$.

Therefore, the rate at which unit-specific shocks influence the macro economy depends on δ_1, which measures the strength of the dominant unit. But it should be noted from (46) that to ensure a non-vanishing variance, $Var\left(\bar{x}_{N,t}\right) > 0$, as $N \to \infty$, we need a value of $\delta_1 = 1$. When $1/2 < \delta_1 < 1$, the network accentuates the diffusion of the idiosyncratic shocks across the network but does not lead to lasting impacts. No network effects of unit-specific shocks can be identified when $\delta_1 \leq 1/2$. Hence, for the dominant unit to have any impact over and above the standard rates of diversification of micro shocks on $\bar{x}_{N,t}$, we need $\delta_1 > 1/2$.[7]

Consider now networks subject to a common shock but without a dominant unit, and note that

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

and the rate of convergence of $\bar{x}_{N,t}$ is determined by the strength of the factor as given by $N^{-2}\psi_n^2$. Recall that $\psi_n = \boldsymbol{\phi}_n'\boldsymbol{\gamma}_2$ and $\varrho\left(\mathbf{W}_{22}\right) \leq 1$, we have

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

By a similar line of reasoning as before, it is then easily seen that $N^{-1}\psi_n = \ominus\left(N^{\alpha-1}\right)$, where α is the cross-section exponent of the factor loadings, γ_i, and measures the degree to which the common factor is pervasive in its effects on sector-specific productivity. Finally, suppose that the production network is subject to a common factor as well as containing a dominant unit. Then for $\delta_1 > 1/2$ and $\alpha > 1/2$ we have

$$Var\left(\bar{x}_{N,t}\right) = \ominus\left(N^{2\delta_1-2}\right) + \ominus\left(N^{2\alpha-2}\right) + \ominus\left(N^{-1}\right). \tag{48}$$

It is clear that the relative importance of the dominant unit and the common factor depends on the relative magnitudes of δ_1 and α. We need estimates of these exponents for a further understanding of the relative importance of macro and micro shocks in business cycle analysis. It is also clear that for the first two terms of (48) to dominate the third terms we must have $\delta_1 = \delta_{max} > 1/2$ and/or $\alpha = \alpha_{max} > 1/2$.

Allowing for multiple factors and multiple dominant units does not alter the main results, and the general expression in (48) will continue to apply. The following proposition summarizes the main theoretical results for the general case.

Proposition 3. *Consider the price network represented by (23), where $\mathbf{x}_t = \mathbf{p}_t - \omega_t\mathbf{1}_N$ is the log price-wage ratio. Suppose that the network contains m dominant units with degrees of dominance δ_i, $i = 1, 2, \ldots, m$ (m is finite), and is subject to r common factors with factor loadings having cross-sectional exponents α_ℓ, for $\ell = 1, 2, \ldots, r$ (r is finite). Then macro volatility, defined as the variance of the aggregate measure $\bar{x}_{N,t} = N^{-1}\mathbf{1}_N'\mathbf{x}_t$, has the following order decomposition*

$$Var\left(\bar{x}_{N,t}\right) = \ominus\left(N^{2\delta_{\max}-2}\right) + \ominus\left(N^{2\alpha_{\max}-2}\right) + \ominus\left(N^{-1}\right), \tag{49}$$

where $\delta_{\max} = \max(\delta_1, \delta_2, \ldots, \delta_N)$ and $\alpha_{\max} = \max\left(\alpha_1, \alpha_2, \ldots, \alpha_r\right)$, with the first two terms having a dominant effects if $\delta_{\max} > 1/2$ and/or $\alpha_{\max} > 1/2$.

6. Estimation and inference

In this section we consider the problem of estimating the degree of dominance of units in a given network. We consider the power law approach employed in the literature as well as a new method that we propose when the outdegrees, $\{d_1, d_2, \ldots, d_N\}$, follow the exponent specification defined by (25). It is unclear if a power law specification for the outdegrees (above a given cut-off value) is necessarily to be preferred to a specification which relates the outdegrees directly to the size of the network, N, without the need to specify a cut-off value. The exponent specification of outdegrees has the added advantage that it also allows identification of more than one dominant units in the network.

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

6.1. Power law estimators

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

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

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

A popular specification is the Pareto distribution. Its CCDF is given by

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

for some shape parameter $\beta > 0$, and the 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. The first is to run the following log–log regression (also known as Zipf regression),

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

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

$$\ln(i - 1/2) = a - \beta \ln d_{(i)}, \; i = 1, 2, \ldots, N_{\min}. \tag{50}$$

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

Another often-used estimator of β is the maximum likelihood estimator (MLE), denoted by $\hat{\beta}_{MLE}$, which is also the well-known Hill estimator (Hill et al., 1975). It can be easily verified that[9]

$$\hat{\beta}_{MLE} = \frac{N_{\min}}{\sum_{i=1}^{N_{\min}} \ln d_{(i)} - N_{\min} \ln d_{(N_{\min})}}, \tag{51}$$

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

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

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

Here $F(d)$ is the CDF of the Pareto distribution that best fits the data for $d \geq d_{\min}$. The MLE in (51) is then computed using the estimated value of $d_{\min}$. Hereafter, we refer to this estimator as the feasible maximum likelihood estimator and denote it by $\hat{\beta}_{CSN}$.[10]

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

6.2. Extremum estimators

Our proposed extremum estimator is motivated by the exponent specification of outdegrees given by (25). In line with the literature on estimation of β, we begin with the case where only a single set of observations on the outdegrees, $\{d_i\}$,

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

9 See, for example, Appendix B of Newman (2005).

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

is available, but instead of the power law specification we assume that d_i, $i = 1, 2, \ldots, N$, are generated according to the following exponent specification:

$$d_i = \kappa N^{\delta_i} \exp(\upsilon_i), \ \ i = 1, 2, \ldots, N, \tag{52}$$

where $0 \leq \delta_i \leq 1$, and $\kappa > 0$ are fixed constants. The above specification is in line with (25) in Definition 1, where we have set $\kappa_i = \kappa \exp(\upsilon_i)$, with $\{\upsilon_i, i = 1, 2, \ldots, N\}$ representing the idiosyncratic shocks to the outdegrees. We also note that since by construction $\sum_{i=1}^{N} d_i = 1$, then (see also (26))

$$\kappa \sum_{i=1}^{N} N^{\delta_i} \exp(\upsilon_i) = N. \tag{53}$$

We make the following assumptions on $\{\upsilon_i\}$.

Assumption 2. The errors $\{\upsilon_i, i = 1, 2, \ldots, N\}$ have zero means and a constant variance σ_υ^2, and there exist finite positive constants C_0 and C_1 such that for all $a > 0$,

$$\sup_i \Pr\left(|\upsilon_i| > a\right) \leq C_0 \exp\left(-C_1 a^2\right). \tag{54}$$

Assumption 3. The errors $\{\upsilon_i, i = 1, 2, \ldots, N\}$ are a stationary sequence of mixing random variables with exponential mixing coefficients given by $\eta_k = \eta_0 \varphi^k$, $\eta_0 > 0$, $0 < \varphi < 1$, and are cross-sectionally weakly correlated, namely,

$$\sup_i \sum_{j=1}^{N} \left|Cov\left(\upsilon_i, \upsilon_j\right)\right| < K. \tag{55}$$

Remark 3. Assumption 2 requires the errors to be sub-Gaussian, which is implied by condition (54).

Remark 4. Assumption 3 allows for a limited degree of dependence in the errors. The mixing condition can also be justified using the mixing results established in Jenish and Prucha (2009) for arrays of random fields. For our application this requires that there exists an ordering of the outdegrees, $\{d_i\}$, such that the cross-correlations decline sufficiently fast along that ordering.

To establish the asymptotic distribution of the extremum estimator, we require a stronger assumption than Assumption 3.

Assumption 4. Denote the ordered values of δ_i by $\delta_{(i)}$, where $\delta_{\max} = \delta_{(1)} > \delta_{(2)} \geq \cdots \geq \delta_{(N)}$. Also denote the random variable, υ_i, associated with $\delta_{(i)}$ by υ_i^*, for $i = 1, 2, \ldots, N$. There exists a finite integer $m > 0$, such that for any $a_i \in \mathbb{R}$,

$$\Pr(\cap_{i=1}^{m} \upsilon_i^* < a_i, \cap_{i=m+1}^{N} \upsilon_i^* < a_i) = \Pr(\cap_{i=1}^{m} \upsilon_i^* < a_i) \Pi_{i=m+1}^{N} \Pr(\upsilon_i^* < a_i). \tag{56}$$

Remark 5. υ_1^* is associated with $\delta_{(1)}$, υ_2^* is associated with $\delta_{(2)}$, and so on, but note that υ_i^* for $i = 1, 2, \ldots, N$ need not have the same ordering as $\delta_{(i)}$.

Remark 6. Assumption 4 allows the shocks to the first m largest outdegrees to dependend on each other, but requires the remaining outdegrees to be independently distributed. Therefore, it follows that $\sup_i \sum_{j=1}^{N} \left|cov\left(\upsilon_i, \upsilon_j\right)\right| = \sup_i \sum_{j=1}^{N} \left|cov\left(\upsilon_i^*, \upsilon_j^*\right)\right| < K$, for a finite m, and under Assumption 4 condition (55) will also be met.

Remark 7. It is worth noting that $\delta_{(1)}$ is assumed to be strictly greater than $\delta_{(2)}$, whereas $\delta_{(i)}$, for $i = 2, 3, \ldots, N$, do not need to take different values. This is a key assumption for identification of $\delta_{(1)}$ associated with the unit with the largest outdegree. The same argument also applies to $\delta_{(2)}$ and so on.

Now we are ready to introduce the extremum estimator. Taking the log transformation of (52) we obtain

$$\ln d_i = \ln \kappa + \delta_i \ln N + \upsilon_i, \ i = 1, 2, \ldots, N. \tag{57}$$

Averaging across i yields

$$N^{-1} \sum_{i=1}^{N} \ln d_i = \ln \kappa + \left(N^{-1} \sum_{i=1}^{N} \delta_i\right) \ln N + N^{-1} \sum_{i=1}^{N} \upsilon_i. \tag{58}$$

But from the summability condition (31) of Assumption 1, it follows that

$$\left(N^{-1} \sum_{i=1}^{N} \delta_i\right) \ln N \leq K \left(\frac{\ln N}{N}\right), \tag{59}$$

and hence

$$\left(N^{-1}\sum_{i=1}^{N}\delta_i\right)\ln N \to 0, \text{ as } N \to \infty. \tag{60}$$

Under Assumptions 2 and 3, the last term of (58) also tends to zero, and $\ln\kappa$ can be estimated by

$$\widehat{\ln\kappa} = N^{-1}\sum_{i=1}^{N}\ln d_i. \tag{61}$$

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

$$\hat{\delta}_{\max} = \frac{\ln d_{\max} - N^{-1}\sum_{i=1}^{N}\ln d_i}{\ln N} = \frac{N^{-1}\sum_{i=1}^{N}\ln\left(d_{\max}/d_i\right)}{\ln N}, \tag{62}$$

where $d_{\max}$ is the largest value of $d_i > 0$.

We will next establish the asymptotic properties of the extremum estimator, $\hat{\delta}_{\max}$. To this end, we make use of the following proposition that gives the tail probability bounds for the sum and deviations of the errors.

Proposition 4. *(i) Under Assumptions 2 and 4, there exist finite positive constants C_0, $\tilde{C}_1$, and C_2, that do not depend on N, such that for any $a \in \mathbb{R}^+$,*

$$\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right| > Na\right) \le C_0 m \exp\left(-\tilde{C}_1 N^2 a^2\right) + \exp\left[-C_2\frac{N^2a^2}{(N-m)}\right], \tag{63}$$

where m is the positive finite integer such that (56) holds.

(ii) Under Assumptions 2 and 3, there exist constants C_0, C_1 not depending on N, $0 < C_0, C_2 < \infty$, and $C_{1N}, C_{3N} > 0$ that are bounded in N, such that for any $a \in \mathbb{R}^+$,

$$\sup_i \Pr\left(|\upsilon_i - \bar{\upsilon}| > a\right) \le C_0\exp\left(-C_{1N}a^2\right) + C_2\exp\left[-C_{3N}a^{2/3}(N-1)^{1/3}\right], \tag{64}$$

where $\bar{\upsilon} = N^{-1}\sum_{j=1}^{N}\upsilon_j$.

See Appendix C for a proof.

Remark 8. The second term in (64) is due to $\bar{\upsilon}$ and will not be present if υ_i is Gaussian or if υ_i for $i = 1, 2, \dots, N$ is a sequence of independent sub-Gaussian processes. In the general case considered here the second term plays a crucial role in allowing the errors, υ_i, to be weakly cross-correlated.

The following theorem establishes the consistency and asymptotic distribution of $\hat{\delta}_{\max}$. Its proof is given in Appendix C.

Theorem 1. *Suppose that the outdegrees of a network follow the exponent specification given by (52). Consider the extremum estimator, $\hat{\delta}_{\max}$, defined by (62).*

(i) Under Assumptions 1–3, $\hat{\delta}_{\max}$ is a consistent estimator of $\delta_{\max}$.

(ii) Under Assumptions 1, 2 and 4, for any $a \in \mathbb{R}$,

$$\lim_{N\to\infty}\Pr\left[\frac{(\ln N)\left(\hat{\delta}_{\max} - \delta_{\max}\right)}{\sigma_\upsilon} \le a\right] \le \Pr\left(\frac{\upsilon_1^*}{\sigma_\upsilon} \le a\right), \tag{65}$$

where υ_1^* is the shock associated with the unit having the largest outdegree and $\sigma_\upsilon^2 = Var\left(\upsilon_1^*\right)$.

(iii) If $\upsilon_1^* \sim N\left(0, \sigma_\upsilon^2\right)$ and σ_υ is known, then a $100(1-p)\%$ symmetric confidence interval for $\delta_{\max}$ is given by $\left[\hat{\delta}_{\max} - c_p\frac{\sigma_\upsilon}{\ln N}, \hat{\delta}_{\max} + c_p\frac{\sigma_\upsilon}{\ln N}\right]$, where $c_p \ge \Phi^{-1}(1-p/2)$, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Remark 9. Derivation of the asymptotic distribution of the extremum estimator is based on Assumption 4 which is stronger than Assumption 3, although it is compatible with it. The results in Theorem 1 differ from those in the literature on order statistics, since the distribution of outdegrees differs across i, even if it is assumed that the shocks, υ_i, are identically distributed.

Remark 10. The above theorem can be applied sequentially to identify units associated with $\delta_{(2)} > \delta_{(3)} > \cdots > \delta_{(m)}$, for a fixed *a priori* given value of m, so long as $\delta_{(m)} > 1/2$. We conjecture that this result follows since we have shown that the most dominant unit with $\delta_{(1)}$ can be identified with probability tending to unity as $N \to \infty$, and conditional on knowing this unit the theorem can then be applied to the rest of the units, and so on. However, it is important to note that our analysis cannot distinguish between two units that are equally dominant, namely if $\delta_{(i)} = \delta_{(i-1)}$ for any $i = 1, 2, \dots, m$.

Please cite this article as: M.H. Pesaran and C.F. Yang, Econometric analysis of production networks with dominant units. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.03.014.

It can be seen from (65) that the limiting distribution of $\hat{\delta}_{\max}$ depends on the distribution of υ_1^*, i.e., the shock to the largest outdegree. If υ_1^* is normally distributed, or equivalently the largest outdegree follows a log-normal distribution, then the critical value of the standard normal distribution can be applied in constructing confidence bounds around $\delta_{\max}$, assuming σ_υ is known. The confidence bounds on $\delta_{\max}$ also shrink at the logarithmic rate of $1/(\ln N)$, which could be slow unless N is sufficiently large. Both of these shortcomings can be overcome in the panel contexts where observations on the outdegrees are available for more than one time period. In most empirical applications the focus would be on short T panels, due to data availability and also because it is unlikely that the same unit continues to be dominant over a long time period.

Specifically, consider as before the exponent specification for d_{it}:

$$d_{it} = \kappa N^{\delta_i} \exp(\upsilon_{it}),\ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{66}$$

where T is finite ($T > 1$) and N is large. Let $\bar{\upsilon}_{iT} = T^{-1}\sum_{t=1}^{T}\upsilon_{it}$, and $\bar{\upsilon}_{NT} = N^{-1}\sum_{i=1}^{N}\bar{\upsilon}_i$, and suppose that Assumptions 2–4 hold with υ_i replaced by $\bar{\upsilon}_{iT}$. Consider the estimators

$$\hat{\delta}_{i,T} = \frac{T^{-1}\sum_{t=1}^{T}\ln d_{it} - (TN)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}\ln d_{jt}}{\ln N},\ \text{ for } i = 1, 2, \ldots, N, \tag{67}$$

and note that the "panel extremum estimator" of $\delta_{\max}$ is given by $\hat{\delta}_{\max,T} = \sup_i(\hat{\delta}_{i,T})$. Using (66) we also have

$$\hat{\delta}_{i,T} - \delta_i = \bar{\delta} + \frac{\bar{\upsilon}_{iT} - \bar{\upsilon}_{NT}}{\ln N}, \tag{68}$$

where as before $\bar{\delta} = N^{-1}\sum_{j=1}^{N}\delta_j$. Consistency of $\hat{\delta}_{\max}$ now follows by setting $\xi_i = \bar{\upsilon}_{iT} - \bar{\upsilon}_{NT}$ in the proof of Theorem 1 and noting that $\bar{\upsilon}_{iT}$ continues to be sub-Gaussian so long as υ_{it} is sub-Guassian. The distributional results of Theorem 1 also follows with this difference that under independence of υ_{it} over t, the confidence band for $\delta_{\max}$ is now given by $\left[\hat{\delta}_{\max,T} - c_p\frac{\sigma_{\upsilon,T}}{\ln N}, \hat{\delta}_{\max,T} + c_p\frac{\sigma_{\upsilon,T}}{\ln N}\right]$, where $\sigma_{\upsilon,T}^2 = Var\,(\bar{\upsilon}_{iT})$, and assuming that the average shock to the largest outdegree is normally distributed. However, $\hat{\delta}_{\max}$ continues to be consistent even if the errors are non-Gaussian.

To estimate $\sigma_{\upsilon,T}^2$ we assume that $\upsilon_{it} \sim IID(0, \sigma_\upsilon^2)$ and note that in this case $\sigma_{\upsilon,T}^2 = T^{-1}\sigma_\upsilon^2$, and σ_υ^2 can be estimated by (for $T > 1$)

$$\hat{\sigma}_\upsilon^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\upsilon}_{it}^2}{N\,(T-1)}, \tag{69}$$

where $\hat{\upsilon}_{it} = \ln d_{it} - \widehat{\ln\kappa} - \hat{\delta}_i \ln N$, and

$$\widehat{\ln\kappa} = (TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\ln d_{it}. \tag{70}$$

It is now easily seen that under our assumptions, $\hat{\sigma}_\upsilon^2$ is an asymptotically unbiased estimator of σ_υ^2 for any fixed $T > 1$. To see this note that

$$\begin{aligned}\hat{\upsilon}_{it} &= \ln d_{it} - \widehat{\ln\kappa} - \hat{\delta}_i \ln N \\ &= -\left(\widehat{\ln\kappa} - \ln\kappa\right) - \left(\hat{\delta}_i - \delta_i\right)\ln N + \upsilon_{it}.\end{aligned} \tag{71}$$

Now using (70) and (68) yields $\hat{\upsilon}_{it} = -2\bar{\delta}\ln N + \upsilon_{it} - \bar{\upsilon}_{iT}$, and in view of Assumptions 1 and 4 it follows that

$$\hat{\upsilon}_{it} = \upsilon_{it} - \bar{\upsilon}_{iT} + o(1),$$

with υ_{it} being cross-sectionally weakly dependent. Using this result in (69) and taking expectations, we have (for a fixed $T > 1$)

$$E\left(\hat{\sigma}_\upsilon^2\right) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}E\,(\upsilon_{it} - \bar{\upsilon}_{iT})^2}{N\,(T-1)} + o(1) = \sigma_\upsilon^2 + o(1),$$

which establishes the desired result.

A test of the null hypothesis that $\delta_{\max} = \delta_{\max}^0$ can now be based on the statistic

$$\mathfrak{D} = \frac{(\ln N)\left(\hat{\delta}_{\max,T} - \delta_{\max}^0\right)}{\hat{\sigma}_\upsilon\left(\frac{1}{T} - \frac{1}{TN}\right)^{1/2}}, \tag{72}$$

which will be normally distributed if $\bar{\upsilon}_{1T}^* \sim N\left(0, \sigma_\upsilon^2\right)$, where $\bar{\upsilon}_{1T}^*$ is the time average of the shocks to the most dominant unit, the unit with the largest outdegree. But one would expect that the normality assumption becomes less critical for large (but finite) values of T. This is because for T sufficiently large, $\sqrt{T}\bar{\upsilon}_{1T}^*$ is likely to be approximately normally distributed under some standard regularity conditions on υ_{1t}^*, for $t = 1, 2, \ldots, T$. However, it is important that T is not too large relative N, otherwise the distribution of $\hat{\delta}_{\max,T}$ could depend on the nuisance parameter, $\bar{\delta}$. To avoid such a possibility we must have

$$\bar{\delta}\,(\ln N)\sqrt{T}=\left(\sum_{i=1}^{N}\delta_i\right)\frac{(\ln N)\sqrt{T}}{N}\to 0,\text{ as } N\to\infty. \tag{73}$$

Under the summability condition (31) of Assumption 1, this requires that $\frac{(\ln N)\sqrt{T}}{N}\to 0$, as N and $T\to\infty$, jointly. A full treatment of the case where both N and $T\to\infty$ is outside the scope of the present paper and is not considered as empirically relevant since in most network applications T is likely to be small relative to N.

6.3. Comparison of power law and extremum estimators

The exponent specification has the advantage that it is closely related to (25) in that $\kappa_i=\kappa\exp(\upsilon_i)>0$, and is in line with the production network model derived from a set of underlying economic relations. Nonetheless, in practice it is difficult to know if the true data generating process follows the exponent or a power law specification. But it turns out that $1/\hat{\delta}_{\max}$ is a consistent estimator of β, the shape parameter of the Pareto distribution, even under the Pareto distribution.

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

$$\begin{aligned} f(d_i) &\propto d_i^{-1-\beta},\text{ if } d_i\geq d_{\min},\\ &\propto \psi(d_i),\text{ if } d_i<d_{\min}, \end{aligned} \tag{74}$$

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

Using (62), the extremum estimator, $\hat{\delta}_{\max}$, can be rewritten as

$$\hat{\delta}_{\max}=\frac{z_{\max}-N^{-1}\sum_{i=1}^{N}z_i}{\ln N}, \tag{75}$$

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 as before. Since $d_{\min}$ is a given constant and by assumption d_i are independently distributed, it then follows that for $z_i\geq 0$, z_i are independent draws from an exponential distribution with parameter β, namely

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

with $E\left(z\,|z\geq 0\right)=1/\beta$, and $Var\left(z\,|z\geq 0\right)=1/\beta^2$, for $\beta>0$.[11] We also assume that $E\left(z_i\,|z_i<0\right)$ exists, which is a mild moment condition imposed on $\psi(d_i)$ for $\ln\left(d_i/d_{\min}\right)<0$. The following theorem summarizes the consistency property of $\hat{\delta}_{\max}$ as an estimator of $1/\beta$.

Theorem 2. *Suppose that the outdegrees d_i, for $i=1,2,\dots,N$, are independent draws from the Pareto tail distribution given by (74) 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. It then follows that*

$$\lim_{N\to\infty}E\left(\hat{\delta}_{\max}\right)=1/\beta,\text{ and } Var\left(\hat{\delta}_{\max}\right)=O\left[\frac{1}{(\ln N)^2}\right], \tag{76}$$

where $\hat{\delta}_{\max}$ is defined by (75).

A proof is provided in Appendix C.

Remark 11. The convergence of $\hat{\delta}_{\max}$ to $\delta=1/\beta$, is at the rate of $1/(\ln N)$ which is rather slow. But it is obtained without making any assumptions about $d_{\min}$ and/or the shape of $\psi(d)$, the non-Pareto part of the distribution.

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

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

7. Monte Carlo experiments

In this section, we investigate the small sample properties of the proposed extremum estimator for balanced panels using Monte Carlo (MC) techniques, and compare its performance with that of the power law method.[12]

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

$$\ln d_{it} = \ln \kappa + \delta_i \ln N + \upsilon_{it},\ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{77}$$

where υ_{it} are weakly cross-sectionally dependent and generated as

$$\upsilon_{it} = \psi \sum_{j=1}^{N} w_{\upsilon,ij}\upsilon_{jt} + \varepsilon_{it}, \tag{78}$$

or in stacked form,

$$\boldsymbol{\upsilon}_t = \psi \mathbf{W}_\upsilon \boldsymbol{\upsilon}_t + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\upsilon}_t = (\upsilon_{1t}, \upsilon_{2t}, \ldots, \upsilon_{Nt})'$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$, $\varepsilon_{it} \sim IIDN(0, 1)$, and $\mathbf{W}_\upsilon = \left(w_{\upsilon,ij}\right)_{N\times N}$ is the 1-ahead-and-1-behind circular weights matrix:

$$\mathbf{W}_\upsilon = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \ldots & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \ldots & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{1}{2} & 0 & \ldots & \frac{1}{2} & 0 \end{pmatrix}. \tag{79}$$

The strength of cross-sectional dependence is set to $\psi = 0.5$.[13] To ensure that d_{it} add up to N across i for each t, κ is set to

$$\kappa = \frac{\exp\left(-\frac{\sigma_\upsilon^2}{2}\right)}{N^{-1}\sum_{i=1}^{N} N^{\delta_i}} > 0, \tag{80}$$

where $\sigma_\upsilon^2 = Var\left(\upsilon_{it}\right)$, which equals the diagonal element of $Var\left(\boldsymbol{\upsilon}_t\right) = \mathbf{RR}'$, where $\mathbf{R} = \left(\mathbf{I}_N - \psi \mathbf{W}_\upsilon\right)^{-1}$.

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

$$f(y_{it}) = \begin{cases} C_t (y_{it}/y_{\min,t})^{-(\beta+1)}, & \text{for } y_{it} \geq y_{\min,t} \\ C_t e^{-(\beta+1)(y_{it}/y_{\min,t}-1)}, & \text{for } y_{it} < y_{\min,t} \end{cases}, \tag{81}$$

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

$$C_t = \left[\frac{y_{\min,t}\left(e^{\beta+1} - 1\right)}{\beta + 1} + \frac{y_{\min,t}}{\beta}\right]^{-1}. \tag{82}$$

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

$$\Pr\left(d_{it} \geq d_{\min,t}\right) = \Pr\left(y_{it} \geq y_{\min,t}\right) = \frac{1}{\beta}\left(\frac{e^{\beta+1} - 1}{\beta + 1} + \frac{1}{\beta}\right)^{-1}, \tag{83}$$

which is time-invariant and depends only on the value of β.[14] The inverse transformation sampling method is used to generate y_{it} such that its distribution satisfies (81). To this end we first generate u_{it} as $IIDU[0, 1]$, $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$, and set

$$u_{\min,t} = C_t\left(\frac{y_{\min,t}}{\beta + 1}\right)\left(e^{\beta+1} - 1\right), \tag{84}$$

[12] Small sample properties of the extremum estimator for unbalanced panels are investigated in the online supplement.

[13] We have considered various intensities of cross-sectional dependence: $\psi = 0.2$, 0.5 and 0.75, as well as independent errors. A full set of results are presented in the online supplement.

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

and then generate y_{it} as

$$y_{it}=\begin{cases}-\dfrac{y_{\min,t}}{\beta+1}\ln\left[1-\dfrac{(\beta+1)\,u_{it}}{C_t e^{\beta+1}y_{\min,t}}\right], & \text{if } u_{it}<u_{\min,t}\\ \left[\dfrac{\beta\left(u_{\min,t}-u_{it}\right)+C_t y_{\min,t}}{C_t y_{\min,t}^{\beta+1}}\right]^{-1/\beta}, & \text{if } u_{it}\geq u_{\min,t}\end{cases}. \tag{85}$$

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

Exponent DGP: Observations on d_{it} are generated according to the exponent specification, (77), with a finite number of dominant units and a large number of non-dominant units. Specifically

- A.1. One dominant unit: $\delta_{\max}=\delta_{(1)}>0$, and $\delta_{(i)}=0$ for $i=2,3,\ldots,N$. In particular, we consider four cases: (i) $\delta_{\max}=1$; (ii) $\delta_{\max}=0.95$; (iii) $\delta_{\max}=0.85$; and (iv) $\delta_{\max}=0.75$.
- A.2. Two dominant units: $\delta_{\max}=\delta_{(1)}>0$, $\delta_{(2)}>0$, and $\delta_{(i)}=0$ for $i=3,4,\ldots,N$. We consider $\delta_{(1)}=0.95$ and $\delta_{(2)}=0.85$..[15]

We consider all combinations of $N=100$, 300, 500, and 1000, and $T=1,2,6,10$, and 20, and also provide simulation results for a very large data set with $N=450{,}000$, which can arise when using inter-firm level sales data.[16] We focus on the 5 largest estimates of δ, which we denote by $\hat{\delta}_{\max,T}=\hat{\delta}_{(1),T}>\hat{\delta}_{(2),T}>\cdots>\hat{\delta}_{(5),T}$, computed according to (67). When $T>1$, the test statistic is computed following (72), where $\hat{\sigma}_v^2$ is given by (69).

Pareto DGP: Observations on d_{it} are generated according to the mixture Pareto distribution, (81), described above and we consider Experiments B.1: $\beta=1.0$, and B.2: $\beta=1.3$.[17] The values of $y_{\min,t}$ are set as $y_{\min,t}=y_{\min}=15$. The sample sizes are combinations of $N=100$, 300, 500, 1000, and 450,000, and $T=1$ and 2. We assess the performance of the Gabaix–Ibragimov estimator ($\hat{\beta}_{GI}$) given by (50) for different given cut-off values, $d_{\min,t}$, the maximum likelihood estimator ($\hat{\beta}_{MLE}$) given by (51) for different $d_{\min,t}$, and the CSN estimator ($\hat{\beta}_{CSN}$) which estimates β jointly with the cut-off value.

As shown in Theorem 2, the inverse of the extremum estimator, $1/\hat{\delta}_{\max}$, is a consistent estimator of β, and analogously one can consider the inverse of β as an estimator $\delta_{\max}$.[18] It is therefore of interest to see how the extremum estimator performs under the Pareto DGP, and conversely how the power law estimators perform under the Exponent DGP. To investigate the robustness of the alternative estimators of β to the choice of the underlying distribution, we conduct two sets of misspecification experiments where we compare the small sample properties of the four estimators of β, namely $\hat{\beta}_{GI}$, $\hat{\beta}_{MLE}$, $\hat{\beta}_{CSN}$, and $\hat{\beta}_{\max}=1/\hat{\delta}_{\max,T}$, when the underlying DGP is Pareto, and conversely when Exponent DGP is used. We consider the values of $\beta=1$, and 1.3 under Pareto DGP, and $\delta_{\max}=1$ and $1/1.3\approx 0.77$ under Exponent DGP. We focus on small values of $T=1$ and 2, for all combinations of $N=100$, 300, 500, 1000, and 450,000.

All experiments are carried out with 2,000 replications.[19]

MC results The estimation results under Exponent DGP for Experiment A are summarized in Table 1, and focus on the extremum estimator of $\delta_{\max}=\delta_{(1)}$ and $\delta_{(2)}$ when applicable. For each experiment we report bias ($\times$100), root mean squared error (RMSE$\times$100), as well as size ($\times$100) and power ($\times$100) for the estimators under consideration. We first note that the bias and RMSE of the extremum estimator decline as N and/or T rises. The bias and RMSE reduction is particularly pronounced as T is increased. This is in line with the theoretical derivations which establish that along the cross-sectional dimension the rate of convergence is of order $1/(\ln N)$, as compared to $T^{-1/2}$ along the time dimension. We also note that the empirical sizes of the tests based on $\hat{\delta}_{\max,T}$ and $\hat{\delta}_{(2),T}$ are close to the assumed 5% nominal size in most cases. It is particularly satisfying to note that the extremum estimator has satisfactory performance even when N approaches 450,000. The slow rate of convergence along the cross section dimension is, however, important for the power of the test. For example, in the case of Experiment A.1, the power of detecting the strongly dominant unit (against the alternative that $\delta_{\max}=0.8$) is around 17.05% 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 41% for $N=100$ to 99.9% for $N=450{,}000$, more than twice the values obtained for $T=2$.

We also consider the frequency with which the dominant unit is correctly selected under Exponent DGP for Experiment A.1. The results are summarized in Table 2, and show that the dominant unit is almost always correctly selected, especially when $T>2$. The frequency of correct selection can be low in the case where $\delta_{\max}=0.75$ and $T=1$, but it increases substantially when $T\geq 2$ even if $N=100$.

Tables 3 and 4 summarize the results for the first set of misspecification experiments where the data are generated from the Pareto tail distribution given by (74). For different values of β, the extremum estimator demonstrates robustness

---

15 We have also considered other values of $\delta_{(1)}$ and $\delta_{(2)}$ for Experiment A.2, including $\delta_{(1)}=0.95$, $\delta_{(2)}=0.75$; and $\delta_{(1)}=0.85$, $\delta_{(2)}=0.75$. These results are provided in the online supplement.

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

17 We have also considered $\beta=1.1$ and 1.2. The results are given in the online supplement.

18 See also the discussions at the end of Section 4 on the relationship between $\delta_{\max}$ and β.

19 We have also investigated the small sample properties of the extremum estimator of $\delta_{\max}$ for exponentially decaying δ_i's and for unbalanced panels. The results are provided in the online supplement.

Please cite this article as: M.H. Pesaran and C.F. Yang, Econometric analysis of production networks with dominant units. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.03.014.

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

	Bias($\times$100)					RMSE($\times$100)					Size($\times$100)					Power($\times$100)				
$T\backslash N$	100	300	500	1,000	450,000	100	300	500	1,000	450,000	100	300	500	1,000	450,000	100	300	500	1,000	450,000
	Experiment A.1(i): $\delta_{\max}=1$																			
1	0.41	−0.03	−0.30	−0.13	−0.09	23.89	20.82	19.36	17.49	9.44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	−1.26	−0.09	−0.57	−0.30	0.03	18.95	15.58	13.97	12.70	6.75	4.90	5.20	5.05	4.75	5.80	17.05	25.00	28.25	33.95	83.85
6	−1.12	−0.28	−0.44	−0.14	0.02	11.29	9.04	8.07	7.39	3.89	6.10	5.75	5.30	5.05	4.55	41.00	59.00	67.90	77.55	99.90
10	−0.91	−0.29	−0.31	−0.19	0.01	8.63	6.81	6.31	5.70	2.96	5.20	5.00	4.65	5.35	4.45	60.80	81.45	87.95	93.55	100.00
20	−1.00	−0.23	−0.25	−0.23	−0.04	5.96	4.89	4.45	4.09	2.12	4.75	5.00	4.90	6.15	4.80	88.70	98.15	99.20	99.60	100.00
	Experiment A.1(ii): $\delta_{\max}=0.95$																			
1	1.19	0.29	−0.09	−0.01	−0.09	23.12	20.39	19.06	17.28	9.44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	−1.15	−0.07	−0.56	−0.29	0.03	18.83	15.54	13.97	12.70	6.75	4.65	5.15	5.05	4.75	5.80	17.10	25.00	28.25	33.95	83.85
6	−1.07	−0.26	−0.43	−0.14	0.02	11.29	9.04	8.07	7.39	3.89	6.05	5.80	5.25	5.05	4.55	41.00	59.10	68.05	77.55	99.90
10	−0.86	−0.27	−0.30	−0.19	0.01	8.62	6.81	6.31	5.70	2.96	5.20	5.00	4.70	5.35	4.45	60.95	81.65	88.05	93.55	100.00
20	−0.95	−0.22	−0.24	−0.23	−0.04	5.95	4.89	4.45	4.09	2.12	4.80	5.00	4.90	6.15	4.80	88.90	98.15	99.20	99.60	100.00
	Experiment A.1(iii): $\delta_{\max}=0.85$																			
1	3.46	1.49	0.73	0.48	−0.09	21.40	19.06	18.04	16.57	9.44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	−0.82	0.02	−0.54	−0.28	0.03	18.44	15.42	13.96	12.69	6.75	4.05	5.05	5.00	4.75	5.80	17.20	25.20	28.35	33.90	83.85
6	−0.97	−0.23	−0.41	−0.13	0.02	11.28	9.04	8.07	7.39	3.89	6.05	5.80	5.25	5.05	4.55	41.35	59.25	68.05	77.55	99.90
10	−0.76	−0.24	−0.28	−0.18	0.01	8.61	6.81	6.30	5.70	2.96	5.25	4.95	4.70	5.35	4.45	61.55	81.90	88.15	93.55	100.00
20	−0.85	−0.18	−0.22	−0.22	−0.04	5.94	4.89	4.45	4.09	2.12	4.50	4.95	4.85	6.05	4.80	89.15	98.15	99.20	99.60	100.00
	Experiment A.1(iv): $\delta_{\max}=0.75$																			
1	6.80	3.74	2.57	1.69	−0.08	20.03	17.37	16.38	15.26	9.43	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	−0.09	0.23	−0.44	−0.23	0.03	17.57	15.07	13.79	12.61	6.75	2.55	4.35	4.50	4.60	5.80	17.30	25.35	28.25	33.85	83.85
6	−0.87	−0.20	−0.39	−0.12	0.02	11.27	9.03	8.07	7.39	3.89	6.00	5.85	5.20	5.10	4.55	41.85	59.35	68.20	77.55	99.90
10	−0.66	−0.20	−0.26	−0.17	0.01	8.61	6.81	6.30	5.70	2.96	5.25	4.90	4.70	5.35	4.45	62.15	82.05	88.15	93.60	100.00
20	−0.75	−0.15	−0.20	−0.21	−0.04	5.92	4.89	4.45	4.09	2.12	4.40	4.95	4.85	6.05	4.80	89.60	98.20	99.20	99.60	100.00
	Experiment A.2: $\delta_{(1)}=0.95$, $\delta_{(2)}=0.85$																			
	$\delta_{(1)}=0.95$																			
1	5.41	3.95	3.34	2.86	0.63	23.21	19.89	18.41	16.57	9.04	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	1.71	1.95	1.28	1.15	0.27	17.57	14.67	13.18	11.99	6.58	3.30	4.25	3.75	3.50	5.15	19.30	28.10	31.60	37.05	85.30
6	−0.79	−0.01	−0.16	0.08	0.03	10.69	8.71	7.71	7.19	3.88	4.70	5.10	4.05	4.70	4.45	41.20	60.55	69.15	78.65	99.90
10	−1.22	−0.32	−0.30	−0.19	0.01	8.37	6.62	6.15	5.62	2.96	4.50	4.10	4.15	5.15	4.45	59.80	82.15	88.20	94.05	100.00
20	−1.68	−0.45	−0.39	−0.31	−0.04	6.02	4.86	4.43	4.09	2.12	4.70	4.95	4.70	6.00	4.80	86.55	98.15	99.40	99.60	100.00
	$\delta_{(2)}=0.85$																			
1	−3.44	−3.19	−3.21	−2.80	−0.89	19.47	17.64	16.82	15.86	9.23	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	−5.34	−3.12	−2.56	−1.71	−0.25	17.60	14.68	13.18	11.95	6.58	3.75	4.95	3.50	4.20	4.45	9.60	17.75	21.80	28.45	84.15
6	−3.07	−1.10	−0.82	−0.39	−0.01	10.68	8.69	7.74	7.18	3.82	4.40	4.20	4.05	5.10	4.55	33.35	57.05	66.10	76.20	99.95
10	−2.29	−0.62	−0.54	−0.32	0.00	8.31	6.53	6.22	5.65	2.96	4.30	4.00	4.80	4.85	4.55	55.35	81.80	88.35	93.35	100.00
20	−1.82	−0.48	−0.41	−0.33	−0.02	6.18	4.78	4.43	4.04	2.07	5.25	4.30	4.50	4.40	5.30	86.25	98.25	99.50	99.85	100.00

Notes: The Data Generating Process (DGP) is given by the exponent specification (77), where the errors are generated by (78) with $\psi = 0.5$. For Experiment A.1, there is one dominant unit and the rest of the units are non-dominant: $\delta_{\max} = \delta_{(1)} > 0$, and $\delta_{(i)} = 0$ for $i = 2, 3, \ldots, N$. For Experiment A.2, there are two weakly dominant units and the rest are non-dominant: $\delta_{\max} = \delta_{(1)} = 0.95$, $\delta_{(2)} = 0.85$, and $\delta_{(i)} = 0$ for $i = 3, 4, \ldots, N$. $\delta_{(i)}$ denotes the ith largest δ, i.e., $\delta_{\max} = \delta_{(1)} > \delta_{(2)} \geq \delta_{(3)} \geq \ldots$, which are estimated by (67). The test statistic is computed using (72), for $T \geq 2$. N/A indicates that the standard error cannot be computed when $T = 1$. The nominal size of the test is 5%. The power is computed at $\delta = \delta_{\max} - 0.2$ for Experiment A.1, and at $\delta = \delta_{(i)} - 0.2$, where $i = 1, 2$, for Experiment A.2. The number of replications is 2,000.

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 6.3. The extremum estimator, $\hat{\beta}_{\max} = 1/\hat{\delta}_{\max,T}$, performs well particularly when $\beta = 1$, even when N is relatively small. For example, under Pareto DGP with $\beta = 1$ (Experiment B.1), for $N = 300$ and $T = 2$, $\hat{\beta}_{\max} = 1.01\ (0.05)$, while $\hat{\beta}_{GI} = 1.04\ (0.19)$ and $\hat{\beta}_{MLE} = 1.05\ (0.14)$, assuming a 10% cut-off value.[20] It is also worth noting that the Gabaix–Ibragimov estimator ($\hat{\beta}_{GI}$) and the ML estimator ($\hat{\beta}_{MLE}$) are quite sensitive to the choice of the cut-off values.[21] The feasible MLE, $\hat{\beta}_{CSN}$, performs better, but it is important to note that the validity of the feasible MLE procedure critically depends on how close the assumed specification of the distribution of d_{it} below $d_{\min,t}$ is to the true underlying distribution.

Consider now the case where the outdegrees are generated according to the exponent specification, (77). In this misspecified case the Pareto estimators, $\hat{\beta}_{GI}$, $\hat{\beta}_{MLE}$, and $\hat{\beta}_{CSN}$, all show significant biases (see Tables 5 and 6). For instance, when $\beta = 1$, $N = 300$ and $T = 2$, and the outdegrees are generated according to the Exponent DGP, the bias of $\hat{\beta}_{GI}$ ranges

20 Figures in brackets are standard errors.

21 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).

Table 2
Frequencies with which the dominant unit is correctly selected, under Exponent DGP for Experiment A.1.

	Empirical frequency (percent)				
$T\backslash N$	100	300	500	1,000	450,000
	A.1(i): $\delta_{\max} = 1$				
1	87.15	94.95	97.05	98.25	100.00
2	99.15	99.80	100.00	100.00	100.00
6	100.00	100.00	100.00	100.00	100.00
10	100.00	100.00	100.00	100.00	100.00
20	100.00	100.00	100.00	100.00	100.00
	A.1(ii): $\delta_{\max} = 0.95$				
1	83.40	92.70	95.20	97.15	100.00
2	98.65	99.70	100.00	100.00	100.00
6	100.00	100.00	100.00	100.00	100.00
10	100.00	100.00	100.00	100.00	100.00
20	100.00	100.00	100.00	100.00	100.00
	A.1(iii): $\delta_{\max} = 0.85$				
1	73.65	83.90	88.00	92.70	100.00
2	96.35	99.25	99.85	99.90	100.00
6	100.00	100.00	100.00	100.00	100.00
10	100.00	100.00	100.00	100.00	100.00
20	100.00	100.00	100.00	100.00	100.00
	A.1(iv): $\delta_{\max} = 0.75$				
1	61.20	71.30	75.00	81.50	99.90
2	90.30	96.55	98.35	99.20	100.00
6	100.00	100.00	100.00	100.00	100.00
10	100.00	100.00	100.00	100.00	100.00
20	100.00	100.00	100.00	100.00	100.00

Notes: This table complements Table 1 and reports the frequencies with which the dominant unit is correctly selected across 2,000 replications. In Experiment A.1, there is one dominant unit and the rest of the units are non-dominant, namely, $\delta_{\max} = \delta_{(1)} > 0$, and $\delta_{(i)} = 0$ for $i = 2, 3, \ldots, N$. See also the notes to Table 1.

from 0.10 to 0.35, for the cut-off values 10% to 30%. Also, the bias of $\hat{\beta}_{GI}$ increases rapidly with N. The ML type estimators exhibit similar biases.

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

8. Dominant units in US production networks

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

We begin with a re-examination of the data set used by Acemoglu et al. (2012) so that we have a direct comparison of the estimates of β (or its inverse) based on the shape of the power law, and the extremum estimator which is given by $\hat{\delta}_{\max,T} = \sup_i \left(\hat{\delta}_{i,T}\right)$, and $\hat{\delta}_{i,T}$ is computed using (67). The Acemoglu et al. (2012) data set is based on the US input–output accounts data over the period 1972–2002 compiled by the Bureau of Economic Analysis (BEA) every five years. We first confirmed that we can replicate their estimates of β, which we denote by $\hat{\beta}_{GI}$ assuming a 20% cut-off value (the percentage above which the degree sequences are assumed to follow the Pareto distribution). The estimates $\hat{\delta}_{\max}$ and the inverse of $\hat{\beta}$ for the years 1972, 1977, 1982, 1987, 1992, 1997 and 2002 are given in Tables 7 and 8. For the inverse of $\hat{\beta}$, Tables 7 and 8 report estimates based on the first-order and second-order interconnections, respectively.[22] We estimate β by the three approaches discussed above, namely Gabaix–Ibragimov estimator ($\hat{\beta}_{GI}$) given by (50), the MLE ($\hat{\beta}_{MLE}$) given by (51),

[22] 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 $\mathbf{d} = (d_1, d_2, \ldots, d_N)'$ is the vector of first-order degrees and $\mathbf{w}_{\cdot j}$ is the jth column of $\mathbf{W}$.

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

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

Notes: The DGP follows the Pareto tail distribution given by (74) with $\beta = 1$. $d_{\min,t}$ denotes the assumed lower bound for the Pareto distribution. The cut-off value refers to the percentage of the largest observations (sorted in descending order) that are assumed to follow the Pareto distribution. The infeasible cut-off value is computed by (83) assuming the true value of $d_{\min,t}$ is known. All estimates are averaged across 2,000 replications. Standard errors are in parentheses. $\hat{\beta}_{GI}$ is the Gabaix–Ibragimov estimator obtained by running the log–log regression, (50). $\widehat{\beta}_{MLE}$ is computed by (51). $\hat{\beta}_{CSN}$ is calculated by applying the joint MLE procedure described in Clauset et al. (2009). $\hat{\delta}_{\max,T} = \sup_i \left(\hat{\delta}_{i,T}\right)$, where $\hat{\delta}_{i,T}$ is computed using (67). The standard error for the inverse of $\hat{\delta}_{\max,T}$ is computed by the delta method. (N/A) indicates that the standard error of $\hat{\delta}_{\max,T}$ cannot be computed when $T = 1$.

and the feasible MLE ($\hat{\beta}_{CSN}$). For the Gabaix–Ibragimov regression and MLE, we give estimates for the cut-off values of 10%, 20%, and 30%. For the feasible MLE, we present both the estimates of β and the estimated cut-off values.[23]

The results in Tables 7 and 8 show that the yearly estimates of $\delta_{\max}$ are clustered within the narrow range of 0.77 to 0.82, covering a relatively long period of 30 years. We cannot provide standard errors for such yearly estimates, but given the small over-time variations in these estimates we can confidently conclude that there is a high degree of sectoral pervasiveness in the US economy, although these estimates do not support the presence of a strongly dominant unit which requires $\hat{\delta}_{\max}$ to be close to unity. In contrast, the estimates of $\delta_{\max}$ based on the inverse of $\hat{\beta}$ differ considerably depending on the estimation methods, the choice of the cut-off value, and whether the first- or second-order interconnections are considered. For example, for 1972, the estimates based on the power law, inverse of $\hat{\beta}_{GI}$, range from 0.694 when the cut-off value is 10% and the first-order interconnections are used, and rise to 1.035 when the second-order interconnections are used with a 30% cut-off value. The estimates of δ based on the inverses of $\hat{\beta}_{GI}$ and $\hat{\beta}_{MLE}$, rise with the choice of cut-off values and with the order of interconnections, whilst our estimator does not require making such choices. Recall that $\delta_{\max}$ provides an exact measure of the rate at which the variance of aggregate output responds to sectoral shocks, whilst β characterizes a lower bound if the first-order interconnections are used. A 20% cut-off value, which is assumed

23 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.

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

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

Notes: The DGP follows the Pareto tail distribution given by (74) with $\beta = 1.3$. See also the notes to Table 3.

by Acemoglu et al. (2012) seems reasonable, considering the closeness between the estimates of $\hat{\delta}_{\max}$ and the inverse of $\widehat{\beta}_{GI}$, and given its similarity to the estimated cut-off values by the feasible MLE. Nevertheless, the estimated cut-off value based on the first-order interconnections for the year 1992 is only 9.5%, which is markedly lower than that for the other years. Similar issues arise when the second-order interconnections are used. The differences between $\hat{\delta}_{\max}$ and inverse of $\widehat{\beta}_{GI}$ also vary across the years. For example, using the second-order interconnections and a cut-off value of 20%, $\hat{\delta}_{\max}$ and inverse of $\widehat{\beta}_{GI}$ are reasonably close for the years 1992, 1997 and 2002, but diverge for the earlier years of 1972, 1977 and 1982.

The data set provided by Acemoglu et al. (2012) does not give the identities of the sectors, which is fine if one is only interested in β or $\delta_{\max}$. But, as noted earlier, our estimation approach also allows us to identify the sectors with the highest degrees of pervasiveness in the production network. With this in mind, we compiled our own **W** matrices from the input–output tables downloaded from the BEA website.[24] The **W** matrices for different years were computed from commodity-by-commodity direct requirements tables at the most 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.[25] 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.[26]

The top five largest estimates of δ, denoted by $\hat{\delta}_{\max} = \hat{\delta}_{(1)} > \hat{\delta}_{(2)} > \cdots > \hat{\delta}_{(5)}$, for each of the years 1972 to 2007 are given in Table 9. The identities of the associated sectors are given in Table 10. We note that both the degrees of dominance and the identities of the pervasive sectors in the US economy are relatively stable over the years. Consistent

24 The input–output accounts data are available at http://www.bea.gov/industry/io_annual.htm.

25 As in Acemoglu et al. (2012), the terms sector and commodity are used interchangeably to convey the same meaning.

26 The Commodity-by-Commodity Direct Requirements (**DR**) table is derived by: $\mathbf{DR} = (\mathbf{TR} - \mathbf{I})\,(\mathbf{TR})^{-1}$, where **I** is an identity matrix, and **TR** denotes the Commodity-by-Commodity Total Requirements table. Then **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**.

Please cite this article as: M.H. Pesaran and C.F. Yang, Econometric analysis of production networks with dominant units. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.03.014.

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

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

Notes: The DGP is given by the exponent specification, (77). There is one strong dominant unit and the rest are non-dominant: $\delta_{\max} = \delta_{(1)} = 1$, with $\delta_{(i)} = 0$ for $i = 2, 3, \ldots, N$, where $\delta_{(i)}$ denotes the ith largest δ. The true value of β is $\beta = 1$. See also the notes to Table 3 for other details.

with the results in Table 7, no sector is strongly dominant. The highest estimate of $\delta_{\max}$ is 0.82, for the year 1992, with an average estimate of around 0.78 over the sample. The wholesale trade sector turns out to be the most dominant sector for all the years with the exception of 2002. In this year the management of companies and enterprises is the most dominant sector with the wholesale trade coming second. It seems reasonable that wholesale trade plays the dominant role given the importance of transportation linking up the different sectors of the economy, providing intermediate goods and transporting final goods to retail sectors.[27]

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 cannot be computed using outdegrees for one single year.[28] Accordingly, to provide more reliable estimates of $\delta_{(1)}, \delta_{(2)}, \ldots, \delta_{(5)}$ and the associated sectoral identities, we also consider pooled estimates. However, there have been major changes in the BEA industry classifications over the years, with the input–output tables for the period 1972–1992 being based on the Standard Industrial Classification (SIC) system, while starting from 1997 they are based on the North American Industry Classification System (NAICS). Accordingly, we compute panel estimates of δ for the two sub-samples separately.[29] The results are summarized in Table 11, which also gives standard errors in parentheses. It is interesting that despite changes to the sectoral classifications, the wholesale trade sector is identified as the most dominant sector in both sub-samples, with $\hat{\delta}_{\max,T} = 0.762$ (0.036) for the first sub-sample (1972–1992), and $\hat{\delta}_{\max,T} = 0.716$ (0.045) for the second sub-sample (1997–2007). The two panel estimates are quite close and identify wholesale trade as the most dominant sector in the US economy. Turning to the estimates of $\delta_{(2)}, \delta_{(3)}, \ldots, \delta_{(5)}$, we find that these estimates are also remarkably similar across the two sub-samples, ranging from 0.667 to 0.605 in the first sub-sample, and 0.683 to 0.595 in the second sub-sample. What has changed is the identity of the sectors across the two sub-samples. For example, the second most dominant sector has been blast furnaces and steel mills over the first sub-sample (1972–1992), whilst it is management companies and enterprises over the second sub-sample (1997–2007).

27 The wholesale trade sector is also found to be dominant in other economies. Dungey and Volkov (2018) apply our extremum estimator to 49 OECD countries and find that in over half of them wholesale trade is in fact the dominant sector.

28 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 take as given.

29 The estimates are computed with unbalanced panels. See Section S2 of the online supplement for an extension of the extremum estimator to unbalanced panels.

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

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

Notes: The DGP is given by the exponent specification, (77). There is one strong dominant unit and the rest of the units are non-dominant: $\delta_{\max} = 1/1.3 = 0.77$, with $\delta_{(i)} = 0$ for $i = 2, 3, \ldots, N$, where $\delta_{(i)}$ denotes the ith largest δ. The true value of β is $\beta = 1.3$. See also the notes to Table 5.

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

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

Notes: Estimates are obtained using the data sets provided by Acemoglu et al. (2012), which are based on the US input–output account data by the Bureau of Economic Analysis (BEA). N is the total number of sectors in a given year and the standard errors are in parentheses. $\hat{\delta}_{\max}$ is the extremum estimator given by (62). The first-order degree sequence is used in the estimation of the shape parameter of the power law, β. $\hat{\beta}_{GI}$ is obtained by the log–log regression with (Gabaix and Ibragimov, 2011) correction using the OLS regression defined by (50). $\hat{\beta}_{MLE}$ is the maximum likelihood estimate (MLE) of β computed by (51). A 10% cut-off value, for example, means that the Pareto tail is taken to be the top 10% of all sectors in terms of outdegrees in each year. Acemoglu et al. (2012) report $\hat{\beta}_{GI}$ estimates only based on a 20% cut-off point. $\hat{\beta}_{CSN}$ is the feasible MLE proposed by Clauset et al. (2009) and its estimated cut-off values are reported in the last column of the table.
[a]From the year 1997 and thereafter, the BEA input–output tables are based on the North American Industry Classification System (NAICS), while for the earlier years they are based on the Standard Industrial Classification (SIC) system.

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

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

Notes: This table differs from Table 7 in that the second-order degree sequence is used to produce the estimates of β . The results of $\widehat{\delta}_{\max}$ in the third column are the same as those in Table 7 and are reproduced here for the convenience of readers. See also the notes to Table 7 for further details.

Table 9
Yearly estimates of the degree of dominance, δ, for the top five pervasive sectors, using US input–output tables (our data).

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

Notes: Estimates are obtained using the input–output accounts data downloaded from the Bureau of Economic Analysis (BEA) website. The table reports the five largest yearly estimates of δ, computed using (67), denoted by $\hat{\delta}_{(1)} = \hat{\delta}_{\max}, \hat{\delta}_{(2)}, \ldots, \hat{\delta}_{(5)}$. 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.

9. Concluding remarks

This paper extends the production network considered by Acemoglu et al. (2012) and derives a dual price network, which allows us to obtain exact conditions under which sectoral shocks can have aggregate effects. The paper presents a simple nonparametric estimator of the degree of pervasiveness of sectoral shocks that compares favorably with the parametric estimators based on Pareto distribution fitted to the outdegrees. The proposed extremum estimator is simple to implement and is applicable not only to the pure cross section models where the Pareto shape parameter is estimated, but also extends readily to short T panels. The paper also develops a simple test of the degree of pervasiveness of the most dominant units in the network, which is shown to have satisfactory size and power properties when N is large, even if T is quite small. The production and price networks considered in this paper are static, but the proposed statistical framework can be extended to allow for dynamics, along similar lines as in Pesaran and Chudik (2014) who consider aggregation of large dynamic panels.

Our empirical application to US input–output tables suggests some evidence of sector-specific shock propagation, but such effects do not seem sufficiently strong and long-lasting, and are likely to be dominated by common technological effects. Similar empirical evidence are also provided by Foerster et al. (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

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

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

Notes: This table complements Table 9 and reports the identities of those sectors corresponding to the five largest estimates of δ (in descending order) for each year.

[a]From the year 1997 and thereafter, the BEA input–output tables are based on the North American Industry Classification System (NAICS), while for the previous years they are based on the Standard Industrial Classification (SIC) system.

linkages.[30] 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.

Appendix A. Lemmas

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

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

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

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

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

Notes: The pooled estimates for the years 1972, 1977, 1982, 1987 and 1992 are based on US input–output data using the Bureau of Economic Analysis (BEA) industry codes, which are in turn based on the Standard Industrial Classification (SIC). For the years 1997, 2002 and 2007, the sectoral classifications are based on the BEA industry codes, which are based on the North American Industry Classification System (NAICS). The table gives the five largest panel estimates of δ together with the identities of the associated sectors. The estimates are denoted by $\hat{\delta}_{(1),T} = \hat{\delta}_{\max,T}, \hat{\delta}_{(2),T}, \ldots, \hat{\delta}_{(5),T}$, and the standard errors are given in parentheses. 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.

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

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

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

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

Proof. See Pesaran (2015, p.756). □

Appendix B. Multiple dominant units

This appendix extends the analysis of Section 5 to the scenario where there is more than one dominant unit in the network. Specifically, we assume that the first m units are dominant with degrees of dominance $\{\delta_1, \delta_2, \ldots, \delta_m\}$, and the rest n units are non-dominant, with $\delta_i = 0$, for $i = m+1, m+2, \ldots, m+n$, and let $N = m+n$. Consider now the following partitioned version of model (23),

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

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

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

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

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

and rearranging yields

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

where

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

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

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

$$\bar{\mathbf{x}}_{Nt}=N^{-1}\left(\mathbf{1}_m'\mathbf{x}_{1t}+\mathbf{1}_n'\mathbf{x}_{2t}\right). \tag{B.3}$$

Using (B.1) in (B.3) gives

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

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

$$\bar{\mathbf{x}}_{Nt}=N^{-1}\left[-a_n+\boldsymbol{\theta}_n'\mathbf{x}_{1t}-(1-\rho)\,\psi_nf_t-(1-\rho)\,\boldsymbol{\phi}_n'\boldsymbol{\varepsilon}_{2t}\right],$$

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

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

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

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

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

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

Turning to the individual terms of $Var(\bar{\mathbf{x}}_{Nt})$, which is given by

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

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

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

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

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

and

$$N^{-1}+\phi_{\min}N^{-1}\sum_{j=1}^n w_{ji,21}\le N^{-1}\theta_{i,n}\le N^{-1}+\phi_{\max}N^{-1}\sum_{j=1}^n w_{ji,21}. \tag{B.4}$$

Also note that $\mathbf{w}_{\cdot i,21}'\mathbf{1}_n=\ominus(N^{\delta_i})$, which immediately follows that

$$\mathbf{w}_{\cdot i,21}'\mathbf{1}_n+\mathbf{w}_{\cdot i,11}'\mathbf{1}_m=d_i=\kappa_iN^{\delta_i},$$

with m being fixed. Therefore, by (B.4) it follows that $N^{-1}\theta_{i,n}=\ominus(N^{\delta_i-1})$, for $i=1,2,\dots,m$, and then $N^{-2}\boldsymbol{\theta}_n'\boldsymbol{\theta}_n=\ominus(N^{2\delta_{\max}-2})$, where $0<\delta_{\max}=\max\left(\delta_1,\delta_2,\dots,\delta_m\right)\le 1$. Further notice that

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

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

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

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

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

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

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

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

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

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

Finally, if the network is subject to both dominant units and a common factor, using $N^{-1}\psi_n=\ominus(N^{\alpha-1})$ and similar arguments as before leads to

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

which is a direct extension of (48) to the multiple-dominant-units network. It is easily seen that when there are multiple factors and multiple dominant units, (49) in Proposition 3 readily follows.

Appendix C. Mathematical proofs

C.1. Proof of Proposition 4

We begin by proving (i). By the triangle inequality,

$$\left|\sum_{i=1}^{N}\upsilon_i\right|=\left|\sum_{i=1}^{N}\upsilon_i^*\right|\leq\left|\sum_{i=1}^{m}\upsilon_i^*\right|+\left|\sum_{i=m+1}^{N}\upsilon_i^*\right|,$$

and it follows that for any $a_N\in\mathbb{R}^+$,

$$\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right|>a_N\right)\leq\Pr\left(\left|\sum_{i=1}^{m}\upsilon_i^*\right|+\left|\sum_{i=m+1}^{N}\upsilon_i^*\right|>a_N\right).$$

By Lemma A11 in the supplement to Chudik et al. (2018), there exists a constant π in the range $0<\pi<1$, such that

$$\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right|>a_N\right)\leq\Pr\left(\left|\sum_{i=1}^{m}\upsilon_i^*\right|>\pi a_N\right)+\Pr\left(\left|\sum_{i=m+1}^{N}\upsilon_i^*\right|>(1-\pi)\,a_N\right). \tag{C.1}$$

Consider the first term on the right-hand side of (C.1). Applying Lemma A11 of Chudik et al. (2018) again gives

$$\Pr\left(\left|\sum_{i=1}^{m}\upsilon_i^*\right|>\pi a_N\right)\leq\sum_{i=1}^{m}\Pr\left(\left|\upsilon_i^*\right|>\pi_i\pi a_N\right),$$

where π_i are any constants satisfying $0<\pi_i<1$ and $\sum_{i=1}^{m}\pi_i=1$. By the sub-Gaussian condition (54) of Assumption 2, we have

$$\Pr\left(\left|\upsilon_i^*\right|>\pi_i\pi a_N\right)\leq C_0\exp\left(-C_1\pi_i^2\pi^2a_N^2\right),$$

for $i=1,2,\ldots,m$, and hence

$$\Pr\left(\left|\sum_{i=1}^{m}\upsilon_i^*\right|>\pi a_N\right)\leq C_0\sum_{i=1}^{m}\exp\left(-C_1\pi_i^2\pi^2a_N^2\right)\leq C_0m\exp\left(-C_1\pi_{\min}^2\pi^2a_N^2\right), \tag{C.2}$$

where $\pi_{\min}=\min\left(\pi_1,\pi_2,\ldots,\pi_m\right)$. Consider now the second term on the right-hand side of (C.1). Note that $\{\upsilon_i\}$, $i=m+1,m+2,\ldots,N$, are independently distributed under Assumption 4; they have zero means, variance σ_υ^2, and are sub-Gaussian under Assumption 2. Therefore, Lemma A3 of Chudik et al. (2018) is applicable,[31] and we have

$$\Pr\left[\left|\sum_{i=m+1}^{N}\upsilon_i^*\right|>(1-\pi)\,a_N\right]\leq\exp\left[-\frac{c^2\left(1-\pi\right)^2a_N^2}{2\sigma_\upsilon^2\left(N-m\right)}\right], \tag{C.3}$$

where c is any constant in the range $0<c<1$. Overall, using (C.2) and (C.3) in (C.1) and letting $a_N=Na$, $a>0$, we obtain

$$\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right|>Na\right)\leq C_0m\exp\left(-C_1\pi_{\min}^2\pi^2N^2a^2\right)+\exp\left[-\frac{c^2\left(1-\pi\right)^2N^2a^2}{2\sigma_\upsilon^2\left(N-m\right)}\right].$$

[31] Lemma A3 of Chudik et al. (2018) provides a more general result on the tail bound for martingale difference sequence.

Setting $\tilde{C}_1 = C_1\pi_{\min}^2\pi^2$, $C_2 = \frac{c^2(1-\pi)^2}{2\sigma_\upsilon^2}$ and recalling that m is finite leads to

$$\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right| > Na\right) \leq C_0 m \exp\left(-\tilde{C}_1 N^2 a^2\right) + \exp\left[-C_2\frac{N^2a^2}{(N-m)}\right],$$

as required.

We now turn to proving (ii). Let $\xi_i = \upsilon_i - \bar{\upsilon}$ and note that

$$\xi_i = (1-N^{-1})\upsilon_i - N^{-1}\sum_{j\neq i}^{N}\upsilon_j.$$

By the triangle inequality $|\xi_i| \leq (1-N^{-1})\,|\upsilon_i| + N^{-1}\left|\sum_{j\neq i}\upsilon_j\right|$, and hence for any $a \in \mathbb{R}^+$,

$$\Pr\left(|\xi_i| > a\right) \leq \Pr\left[(1-N^{-1})\,|\upsilon_i| + N^{-1}\left|\sum_{j\neq i}^{N}\upsilon_j\right| > a\right].$$

By Lemma A11 in the supplement to Chudik et al. (2018), there exists a constant π in the range $0 < \pi < 1$, such that

$$\Pr\left(|\xi_i| > a\right) \leq \Pr\left[(1-N^{-1})\,|\upsilon_i| > (1-\pi)a\right] + \Pr\left(N^{-1}\left|\sum_{j\neq i}^{N}\upsilon_j\right| > \pi a\right). \tag{C.4}$$

But by the sub-Gaussian condition (54) of Assumption 2, we have

$$\Pr\left[(1-N^{-1})\,|\upsilon_i| > (1-\pi)a\right] \leq C_0 \exp\left[-C_1\frac{N^2(1-\pi)^2a^2}{(N-1)^2}\right].$$

Consider now the second term of (C.4) and note that

$$\Pr\left(N^{-1}\left|\sum_{j\neq i}^{N}\upsilon_j\right| > \pi a\right) = \Pr\left(\left|\sum_{j\neq i}^{N}\upsilon_j\right| > N\pi a\right).$$

Under Assumption 3, Theorem 3.5 of White and Wooldridge (1991) can be applied to obtain (for choice of $p = 2$ and $\lambda = 1$ in their notations)

$$\begin{aligned}\Pr\left(\left|\sum_{j\neq i}^{N}\upsilon_j\right| > N\pi a\right) &\leq C_2\exp\left\{-C_3\left[N\pi a\,(N-1)^{-1/2}\right]^{2/3}\right\}\\ &= C_2\exp\left[-C_3\left(\frac{\pi N}{N-1}\right)^{2/3}a^{2/3}(N-1)^{1/3}\right],\end{aligned}$$

for some finite positive constants C_2 and C_3. Hence

$$\Pr\left(|\xi_i| > a\right) \leq C_0\exp\left[-C_1\frac{N^2(1-\pi)^2a^2}{(N-1)^2}\right] + C_2\exp\left[-C_3\left(\frac{\pi N}{N-1}\right)^{2/3}a^{2/3}(N-1)^{1/3}\right].$$

Setting $C_{1N} = \frac{C_1(1-\pi)^2N^2}{(N-1)^2}$, $C_{3N} = C_3\left(\frac{\pi N}{N-1}\right)^{2/3}$, and noting that C_{1N} and C_{3N} are positive and bounded in N, we have

$$\Pr\left(|\xi_i| > a\right) \leq C_0\exp\left(-C_{1N}a^2\right) + C_2\exp\left[-C_{3N}a^{2/3}(N-1)^{1/3}\right],$$

which is the desired result.

C.2. Proof of Theorem 1

(i) Consistency of $\hat{\delta}_{\max}$. The extremum estimator of $\delta_{\max}$ can be rewritten as $\hat{\delta}_{\max} = \sup_i\left(\hat{\delta}_i\right)$, where $\hat{\delta}_i$ is defined by

$$\hat{\delta}_i = \frac{\ln d_i - N^{-1}\sum_{j=1}^{N}\ln d_j}{\ln N}. \tag{C.5}$$

Substituting (57) into (C.5) we obtain

$$\hat{\delta}_i - \delta_i = \bar{\delta} + \frac{\xi_i}{\ln N}, \tag{C.6}$$

where $\xi_i = \upsilon_i - \bar{\upsilon}$, $\bar{\upsilon} = N^{-1}\sum_{j=1}^{N}\upsilon_j$, and under Assumption 1 we have $\bar{\delta} = N^{-1}\sum_{j=1}^{N}\delta_j = O(N^{-1})$. For any $\epsilon > 0$,

$$\Pr\left(\left|\hat{\delta}_{\max} - \delta_{\max}\right| > \epsilon\right) = \Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) + \Pr\left(\hat{\delta}_{\max} - \delta_{\max} \leq -\epsilon\right). \tag{C.7}$$

Consider the first term on the right-hand side of (C.7), and note that

$$\begin{aligned}\Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) &= \Pr\left[\sup_i\left(\hat{\delta}_i\right) > \epsilon + \delta_{\max}\right] = \Pr\left[\cup_{i=1}^{N}\left(\hat{\delta}_i > \epsilon + \delta_{\max}\right)\right] \\ &= \Pr\left[\cup_{i=1}^{N}\left(\hat{\delta}_i - \delta_i > \epsilon + a_i\right)\right] \leq \sum_{i=1}^{N}\Pr\left(\hat{\delta}_i - \delta_i > \epsilon + a_i\right),\end{aligned}$$

where $a_i = \delta_{\max} - \delta_i \geq 0$. Using (C.6) we obtain

$$\Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) \leq \sum_{i=1}^{N}\Pr\left(\bar{\delta} + \frac{\xi_i}{\ln N} > \epsilon + a_i\right) = \sum_{i=1}^{N}\Pr\left(\xi_i > q_{iN}\right), \tag{C.8}$$

where $q_{iN} = \left(\epsilon - \bar{\delta} + a_i\right)(\ln N)$. Since $a_i \geq 0$ for all i, and $\bar{\delta} \leq K/N$, for some fixed $K > 0$, we then have $q_{iN} \geq \epsilon\ln N - K\left(\frac{\ln N}{N}\right)$, and hence

$$\Pr\left(\xi_i > q_{iN}\right) \leq \Pr\left[\xi_i > \epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right].$$

Using this result in (C.8) we now obtain

$$\Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) \leq N\sup_i\Pr\left[\xi_i > \epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right], \tag{C.9}$$

where for any choice of ϵ and K, there is some N_0, such that for all $N > N_0$, $\epsilon\ln N - K\left(\frac{\ln N}{N}\right) > 0$. Applying Proposition 4(ii) to ξ_i leads to

$$\begin{aligned}\Pr\left[\xi_i > \epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right] \leq\ & C_0\exp\left\{-C_{1N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^2\right\} \\ &+ C_2\exp\left\{-C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3}\right\}.\end{aligned}$$

Substituting this result in (C.9) now yields

$$\begin{aligned}\Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) \leq\ & C_0\exp\left\{\ln N - C_{1N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^2\right\} \\ &+ C_2\exp\left\{\ln N - C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3}\right\}.\end{aligned}$$

The first exponential term tends to zero since $(\ln N)/N = o(1)$, $C_{1N}\epsilon^2 > 0$, C_{1N} is bounded in N, and as a result $\ln N - C_{1N}\epsilon^2(\ln N)^2 \to -\infty$, as $N \to \infty$. Similarly, the second exponential term also approaches zero for ϵ sufficiently small. To see this, note that

$$\begin{aligned}&\ln N - C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3} \\ &= -C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3}\left\{1 - \frac{\ln N}{C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3}}\right\}.\end{aligned}$$

But (recalling that C_{3N} is bounded in N)

$$\frac{\ln N}{C_{3N}\left[\epsilon\ln N - K\left(\frac{\ln N}{N}\right)\right]^{2/3}(N-1)^{1/3}} = O\left[\left(\frac{\ln N}{N}\right)^{1/3}\right] = o(1),$$

and hence $\Pr\left(\hat{\delta}_{\max} - \delta_{\max} > \epsilon\right) \to 0$ as $N \to \infty$. Consider now the second term on the right-hand side of (C.7),

$$\begin{aligned}\Pr\left(\hat{\delta}_{\max} - \delta_{\max} \leq -\epsilon\right) &= \Pr\left[\sup_i\left(\hat{\delta}_i\right) \leq -\epsilon + \delta_{\max}\right] = \Pr\left[\cap_{i=1}^{N}\left(\hat{\delta}_i \leq -\epsilon + \delta_{\max}\right)\right] \\ &= \Pr\left[\cap_{i=1}^{N}\left(\hat{\delta}_i - \delta_i \leq -\epsilon + a_i\right)\right].\end{aligned}$$

Using (C.6) again gives

$$\Pr\left(\hat{\delta}_{\max}-\delta_{\max}\leq-\epsilon\right)=\Pr\left[\cap_{i=1}^{N}\left(\frac{\xi_i}{\ln N}\leq-\epsilon-\bar{\delta}+a_i\right)\right].$$

Consider the ordered values of δ_i, namely $\delta_{(1)}>\delta_{(2)}\geq\cdots\geq\delta_N\geq0$, and denote the ξ_i associated with this ordering by ξ_i^*. Specifically, ξ_1^* is associated with $\delta_{(1)}$, ξ_2^* is associated with $\delta_{(2)}$ and so on. Note that the probability of intersection of events is invariant to the reordering. Therefore, we have

$$\Pr\left(\hat{\delta}_{\max}-\delta_{\max}\leq-\epsilon\right)=\Pr\left[\cap_{i=1}^{N}\left(\frac{\xi_i^*}{\ln N}\leq-\epsilon-\bar{\delta}+a_{(i)}\right)\right],$$

where $a_{(1)}=\delta_{\max}-\delta_{(1)}=0$, and $a_{(i)}>0$ for $i>1$. Denote the events $\{\frac{\xi_i^*}{\ln N}\leq-\epsilon-\bar{\delta}+a_{(i)},\ i=1,2,\ldots,N\}$ by $\left\{\mathcal{A}_i^*\right\}$, and suppose that $\Pr\left(\mathcal{A}_1^*\right)>0$. The case where $\Pr\left(\mathcal{A}_1^*\right)=0$ can be ruled out, since in that case we must have $\Pr\left(\hat{\delta}_{\max}-\delta_{\max}\leq-\epsilon\right)=0$, and $\Pr\left(\hat{\delta}_{\max}-\delta_{\max}>\epsilon\right)\to0$ as $N\to\infty$, and consistency of $\hat{\delta}_{\max}$ follows trivially. But under $\Pr\left(\mathcal{A}_1^*\right)>0$, the conditional probability $\Pr\left(\cap_{i=2}^{N}\mathcal{A}_i^*\left|\mathcal{A}_1^*\right.\right)$ exists and since $0\leq\Pr\left(\cap_{i=2}^{N}\mathcal{A}_i^*\left|\mathcal{A}_1^*\right.\right)\leq1$, it follows that

$$\Pr\left[\cap_{i=1}^{N}\left(\frac{\xi_i^*}{\ln N}\leq-\epsilon-\bar{\delta}+a_{(i)}\right)\right]=\Pr\left(\mathcal{A}_1^*\right)\times\Pr\left(\cap_{i=2}^{N}\mathcal{A}_i^*\left|\mathcal{A}_1^*\right.\right)\leq\Pr\left(\mathcal{A}_1^*\right).$$

Consider now $\Pr\left(\mathcal{A}_1^*\right)$ and note that $\Pr\left(\mathcal{A}_1^*\right)=\Pr\left[\xi_1^*\leq-\left(\epsilon+\bar{\delta}\right)\ln N\right]$, where $\epsilon+\bar{\delta}>0$. Now using result (ii) in Proposition 4 we have

$$\begin{aligned}\Pr\left(\hat{\delta}_{\max}-\delta_{\max}\leq-\epsilon\right)\leq{}&C_0\exp\left[-C_{1N}\left(\epsilon+\bar{\delta}\right)^2(\ln N)^2\right]\\&+C_2\exp\left[-C_{3N}\left(\epsilon+\bar{\delta}\right)^{2/3}(\ln N)^{2/3}(N-1)^{1/3}\right].\end{aligned}$$

Recalling that $\bar{\delta}\ln N=O\left(N^{-1}\ln N\right)=o(1)$, C_{1N} and C_{3N} are positive and bounded in N, it is easily seen that both exponential terms of the above tend to zero for any $\epsilon>0$. Thus, overall $\Pr\left(\left|\hat{\delta}_{\max}-\delta_{\max}\right|>\epsilon\right)\to0$ as $N\to\infty$, and this completes the proof of part (i) of the theorem.

(ii) Asymptotic distribution of $\delta_{\max}$. Consider now part (ii) of the Theorem and note that for any $a\in\mathbb{R}$,

$$\Pr\left[\frac{(\ln N)\left(\hat{\delta}_{\max}-\delta_{(1)}\right)}{\sigma_\upsilon}\leq a\right]=\Pr\left\{\cap_{i=1}^{N}\left[\frac{(\ln N)\left(\hat{\delta}_i-\delta_{(1)}\right)}{\sigma_\upsilon}\leq a\right]\right\}.$$

Let $\mathfrak{D}_N=(\ln N)\left(\hat{\delta}_{\max}-\delta_{(1)}\right)/\sigma_\upsilon$. Using (C.5) we have

$$\Pr\left(\mathfrak{D}_N\leq a\right)=\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i-\bar{\upsilon}}{\sigma_\upsilon}\leq b_{iN}\right)\right],$$

where

$$b_{iN}=a+\frac{(\ln N)\left(\delta_{(1)}-\delta_i-\bar{\delta}\right)}{\sigma_\upsilon}$$

Equivalently, we can also write

$$\Pr\left(\mathfrak{D}_N\leq a\right)=\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*-\bar{\upsilon}}{\sigma_\upsilon}\leq b_{iN}^*\right)\right],$$

where υ_i^* is defined in Assumption 4 and

$$b_{iN}^*=a+\frac{(\ln N)\left(\delta_{(1)}-\delta_{(i)}-\bar{\delta}\right)}{\sigma_\upsilon}.$$

Furthermore,

$$\begin{aligned}\Pr\left(\mathfrak{D}_N\leq a\right)={}&\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*-\bar{\upsilon}}{\sigma_\upsilon}\leq b_{iN}^*\right)\left|\bar{\upsilon}\leq\frac{\ln N}{\sqrt{N}}\right.\right]\times\left[1-\Pr\left(\bar{\upsilon}>\frac{\ln N}{\sqrt{N}}\right)\right]\\&+\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*-\bar{\upsilon}}{\sigma_\upsilon}\leq b_{iN}^*\right)\left|\bar{\upsilon}>\frac{\ln N}{\sqrt{N}}\right.\right]\times\Pr\left(\bar{\upsilon}>\frac{\ln N}{\sqrt{N}}\right).\end{aligned}$$

Note that

$$\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*-\bar{\upsilon}}{\sigma_\upsilon}\leq b_{iN}^*\right)\left|\bar{\upsilon}\leq\frac{\ln N}{\sqrt{N}}\right.\right]\leq\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\leq b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)\right],$$

and under Assumption 4 we have

$$\Pr\left[\cap_{i=1}^{N}\left(\frac{\upsilon_i^*-\bar{\upsilon}}{\sigma_\upsilon}\le b_{iN}^*\right)\middle|\bar{\upsilon}\le\frac{\ln N}{\sqrt{N}}\right]$$
$$\le \Pr\left[\cap_{i=1}^{m}\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)\right]\Pi_{i=m+1}^{N}\Pr\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right).$$

In addition,

$$\Pr\left(\bar{\upsilon}>\frac{\ln N}{\sqrt{N}}\right)\le\Pr\left(|\bar{\upsilon}|>\frac{\ln N}{\sqrt{N}}\right)=\Pr\left(\left|\sum_{i=1}^{N}\upsilon_i\right|>N\frac{\ln N}{\sqrt{N}}\right).$$

Applying Proposition 4(i) yields

$$\Pr\left(\bar{\upsilon}>\frac{\ln N}{\sqrt{N}}\right)=O\left\{\exp\left[-C_3\left(\frac{N}{N-m}\right)(\ln N)^2\right]\right\},$$

for some positive finite constant C_3. Thus, overall we have

$$\Pr\left(\mathfrak{D}_N\le a\right)\le\Pr\left[\cap_{i=1}^{m}\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)\right]\Pi_{i=m+1}^{N}\Pr\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)$$
$$+O\left\{\exp\left[-C_3\left(\frac{N}{N-m}\right)(\ln N)^2\right]\right\}.$$

Now consider the limit of the above probability distribution as $N\to\infty$. Since m is assumed to be finite, the last term approaches zero as $N\to\infty$. Note that

$$b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\to a,\ \text{for } i=1,$$
$$b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\to+\infty,\ \text{for } i>1.$$

Also recall that under Assumption 1, $\ln N\bar{\delta}=O\left(N^{-1}\ln N\right)=o(1)$, and $\delta_{(1)}-\delta_{(i)}>0$, for $i>1$ under Assumption 4. Therefore, it follows that

$$\lim_{N\to\infty}\Pr\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)=1,\text{ for } i=m+1,m+2,\ldots,N,$$

and

$$\lim_{N\to\infty}\Pr\left[\cap_{i=1}^{m}\left(\frac{\upsilon_i^*}{\sigma_\upsilon}\le b_{iN}^*+\sigma_\upsilon^{-1}\frac{\ln N}{\sqrt{N}}\right)\right]=\Pr\left(\frac{\upsilon_1^*}{\sigma_\upsilon}\le a,\frac{\upsilon_2^*}{\sigma_\upsilon}<\infty,\ldots,\frac{\upsilon_m^*}{\sigma_\upsilon}<\infty\right)$$
$$=\Pr\left(\frac{\upsilon_1^*}{\sigma_\upsilon}\le a\right),$$

which is the marginal distribution of the shock to the dominant unit. Hence, overall we conclude that for any $a\in\mathbb{R}$,

$$\lim_{N\to\infty}\Pr\left(\mathfrak{D}_N\le a\right)\le\Pr\left(\frac{\upsilon_1^*}{\sigma_\upsilon}\le a\right).$$

(iii) Confidence interval for $\delta_{\max}$. Let $F(\mathfrak{D}\le a)$ denote the limiting distribution of $\mathfrak{D}_N$. In the case where $\upsilon_1^*\sim N(0,\sigma_\upsilon^2)$, the result in (ii) implies that for any $a\in\mathbb{R}$ and as $N\to\infty$ we have

$$F(\mathfrak{D}\le a)\le\Phi\left(a\right). \tag{C.10}$$

In constructing a symmetric confidence bounds for the upper tail we need to find $c_p>0$ such that $F(\mathfrak{D}>c_p)\le p/2$, and for the lower tail, $F(\mathfrak{D}<-c_p)\le p/2$. Consider the upper tail and note that $1-F(\mathfrak{D}\le c_p)\le p/2$, or $F(\mathfrak{D}\le c_p)\ge 1-p/2$. Using (C.10) we now have

$$1-p/2\le F(\mathfrak{D}\le c_p)\le\Phi\left(c_p\right),$$

and hence $\Phi\left(c_p\right)\ge 1-p/2$, which yields $c_p\ge\Phi^{-1}(1-p/2)$, since $\Phi\left(c_p\right)$ is non-decreasing in c_p. Similarly, for the lower tail (using (C.10)) we also have

$$F(\mathfrak{D}<-c_p)\le\Phi\left(-c_p\right)\le p/2.$$

But $\Phi\left(-c_p\right)\le p/2$ can also be written as $1-\Phi\left(c_p\right)\le p/2$, or $\Phi\left(c_p\right)\ge 1-p/2$, which gives the same range of values for c_p as obtained for the upper tail. Therefore, under log-normality of the largest outdegree and for N sufficiently large, setting $c_p \geq \Phi^{-1}(1-p/2)$ will ensure that

$$\lim_{N\to\infty} \Pr\left(|\mathfrak{D}_N| > c_p\right) \leq F(\mathfrak{D} > c_p) + F(\mathfrak{D} < -c_p) \leq p.$$

C.3. Proof of Theorem 2

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

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

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

$$E\left(z_{(i)}\right) = (1/\beta)\left(\sum_{j=1}^{N-i+1}\frac{1}{j}\right),\ Var(z_{(i)}) = (1/\beta)^2\left(\sum_{j=1}^{N-i+1}\frac{1}{j^2}\right),\ \text{for } i=1,2,\ldots,N. \tag{C.11}$$

Noting that $\hat{\delta}_{\max}$ is given by (75), and making use of the above results we now have

$$E\left(\hat{\delta}_{\max}\right) = \frac{E\left(z_{\max}\right) - E\left(\bar{z}_N\right)}{\ln N} = \frac{(1/\beta)\sum_{j=1}^{N} j^{-1} - E\left(\bar{z}_N\right)}{\ln N}, \tag{C.12}$$

$$\begin{aligned} Var\left(\hat{\delta}_{\max}\right) &= \frac{Var\left(z_{\max}\right) + N^{-2}\sum_{i=1}^{N} Var(z_i) - 2N^{-1}\sum_{i=1}^{N} Cov(z_{\max}, z_{(i)})}{(\ln N)^2} \\ &= \frac{Var\left(z_{\max}\right) + N^{-2}\sum_{i=1}^{N} Var(z_i) - 2N^{-1}\sum_{i=1}^{N} Var(z_{(i)})}{(\ln N)^2}. \end{aligned} \tag{C.13}$$

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

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

and hence

$$\lim_{N\to\infty}\frac{\sum_{j=1}^{N} j^{-1}}{\ln N} = 1. \tag{C.14}$$

Using (C.11) and (C.14) in (C.12) we now have $\lim_{N\to\infty} E\left(\hat{\delta}_{\max}\right) = 1/\beta$.

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

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

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

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

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

Appendix D. Supplementary Monte Carlo results

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.03.014.

References

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

Acemoglu, D., Autor, D., Dorn, D., Hanson, G.H., Price, B., 2016b. Import competition and the great US employment sag of the 2000s. J. Labor Econ. 34, S141–S198.

Acemoglu, D., Carvalho, V.M., Ozdaglar, A., Tahbaz-Salehi, A., 2012. The network origins of aggregate fluctuations. Econometrica 80, 1977–2016.
Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2016c. Networks, shocks, and systemic risk. In: Bramoulle, Y., Galeotti, A., Rogers, B. (Eds.), Oxford Handbook of the Economics of Networks. Oxford University Press, Oxford, pp. 569–609 (Chapter 21).
Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics, Vol. 54. SIAM-Society for Industrial and Applied Mathematics.
Bai, J., Li, K., 2013. Spatial panel data models with common shocks. Available at SSRN: https://urldefense.proofpoint.com/v2/url?u=http-3A__dx.doi.org_10.2139_ssrn.2373628&d=DwIGAg&c=HPMtquzZjKY31rtkyGRFnQ&r=ZqJKyRxNFVewc1f5pnDzha2M9xbG2qEx2dYeneWbhC0&m=4rq7zVb-IU53nyJhJdIJEZWEafE2Kz-TqcFHINilnYs&s=oD5wKjHkHIiUqPta7YoWAzSIeQWi2CKb7Fp22thbAoo&e=.
Bailey, N., Kapetanios, G., Pesaran, M.H., 2016. Exponent of cross-sectional dependence: Estimation and inference. J. Appl. Econometrics 31, 929–960.
Bonar, D.D., Khoury, Jr., M.J., Khoury, M., 2006. Real Infinite Series. Mathematical Association of America.
Carvalho, V.M., 2014. From micro to macro via production networks. J. Econ. Perspect. 28, 23–47.
Carvalho, V.M., Nirei, M., Saito, Y., Tahbaz-Salehi, A., 2016. Supply chain disruptions: Evidence from the Great East Japan Earthquake. Available at SSRN: https://urldefense.proofpoint.com/v2/url?u=http-3A__dx.doi.org_10.2139_ssrn.2893221&d=DwIGAg&c=HPMtquzZjKY31rtkyGRFnQ&r=ZqJKyRxNFVewc1f5pnDzha2M9xbG2qEx2dYeneWbhC0&m=4rq7zVb-IU53nyJhJdIJEZWEafE2Kz-TqcFHINilnYs&s=zJb8nVDnVEk_T_tYFjlz1diwDrgu8n6dfGJowy3oHfs&e=.
Chudik, A., Kapetanios, G., Pesaran, M.H., 2018. A one-covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. Econometrica 86 (4), 1479–1512.
Chudik, A., Pesaran, M.H., 2013. Econometric analysis of high dimensional VARs featuring a dominant unit. Econometric Rev. 32, 592–649.
Chudik, A., Pesaran, M.H., Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. Econom. J. 14, C45–C90.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Rev. 51, 661–703.
Dungey, M., Volkov, V., 2018. R&D and wholesale trade are critical to the economy: Identifying dominant sectors from economic networks. Econom. Lett. 162, 81–85.
Eeckhout, J., 2004. Gibrat's law for (all) cities. Amer. Econ. Rev. 94, 1429–1451.
Foerster, A.T., Sarte, P.-D.G., Watson, M.W., 2011. Sectoral versus aggregate shocks: A structural factor analysis of industrial production. J. Polit. Econ. 119, 1–38.
Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79, 733–772.
Gabaix, X., Ibragimov, R., 2011. Rank- 1/2: A simple way to improve the OLS estimation of tail exponents. J. Bus. Econom. Statist. 29, 24–39.
Hill, B.M., et al., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.
Horvath, M., 1998. Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks. Rev. Econ. Dyn. 1, 781–808.
Horvath, M., 2000. Sectoral shocks and aggregate fluctuations. J. Monetary Econ. 45, 69–106.
Jenish, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. J. Econometrics 150 (1), 86–98.
Long, J.B., Plosser, C.I., 1983. Real business cycles. J. Polit. Econ. 91, 39–69.
Newman, M.E., 2005. Power laws, Pareto distributions and Zipf's law. Contemp. Phys. 46, 323–351.
Pesaran, M.H., 2015. Time Series and Panel Data Econometrics. Oxford University Press, Oxford.
Pesaran, M.H., Chudik, A., 2014. Aggregation in large dynamic panels. J. Econometrics 178, 273–285.
Siavash, S.S., 2018. Dominant U.S. manufacturing sectors: A factor model analysis. Unpublished manuscript. Available at https://urldefense.proofpoint.com/v2/url?u=https-3A__sites.google.com_site_sorooshsoofisiavash_research&d=DwIGAg&c=HPMtquzZjKY31rtkyGRFnQ&r=ZqJKyRxNFVewc1f5pnDzha2M9xbG2qEx2dYeneWbhC0&m=4rq7zVb-IU53nyJhJdIJEZWEafE2Kz-TqcFHINilnYs&s=WcqsrLTd0cxlnzAhOSXJw-bfFKghbk1Ltf0ShIkqKiQ&e=.
Stewart, W.J., 2009. Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling. Princeton University Press, Princeton, NJ.
White, H., Wooldridge, J., 1991. Some results on sieve estimation with dependent observations. In: Barnett, W.A., Powell, J., Tauchen, G.E. (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, New York, pp. 459–493.
Yang, C.F., 2020. Common factors and spatial dependence: An application to US house prices. Econometric Rev. http://dx.doi.org/10.1080/07474938.2020.1741785, (forthcoming).