Propagation of Shocks in Networks: Identification and Applications*

Gabriel Chodorow-Reich, Xavier Gabaix, Davide Viviano

June 3, 2026. Preliminary and Incomplete.

Abstract

This paper develops the Network GIV (granular instrumental variables), an estimation procedure for causally identifying the effects of shocks in general networks, including the relevant elasticities. Using panel data with some large units—such as countries, industries, or firms—and a parametric model of shock propagation, the method recovers idiosyncratic unit-level shocks and uses them as instruments to estimate the underlying network parameters. The algorithm also applies to panels that are short, have measurement error and even missing data. We apply the Network GIV to study the propagation of TFP shocks across sectors through customer-supplier linkages. We find large amplification: a 1% average primitive increase in industry TFP raises aggregate TFP by 5% in the long run. More broadly, the Network GIV offers a general-purpose approach for causally estimating elasticities and other structural objects, and for studying how productivity, demand, and financial shocks propagate within and across countries. The method is particularly useful in macro-finance environments, where instruments are often scarce.

*{chodorowreich,xgabaix,dviviano}@fas.harvard.edu. We thank Ralph Koijen for very useful conversations at the beginning of this project. For excellent research assistance, we thank Agustin Barboza Da Rosa, Suri Chen, Dario Iori, Zhiyu Fu, Lilian Hartmann, Pier Giuseppe Sarzotti, Constantin Schesch, Judy Yue, George Wang, and for helpful comments we thank Kirill Borusyak, Aureo de Paula, Basile Grassi, Mikkel Plagborg-Møller, and participants at seminars and conferences.

1 Introduction

A key challenge in economics is the identification of structural parameters that govern causal relationships. This challenge is exacerbated in the presence of network economies, wherein (large) units such as firms, industries or countries interact with each other in the propagation of shocks through the economy. Such propagation can be difficult to measure, due to the lack of valid instruments to identify the shocks (and their effect on the economy), as often pointed out in the macro-finance literature (e.g., Ramey (2016), Nakamura and Steinsson (2018), Chodorow-Reich (2019), Huber (2023)).

To meet this challenge, the present paper proposes a novel method to identify parameters in economic networks, the "network granular instrumental variables", or Network GIV. We build on the idea of granular instrumental variables (GIV) developed in Gabaix and Koijen (2024). There, the procedure extracted idiosyncratic shocks (via a factor model) and used them as primitive instruments to estimate for instance demand and supply elasticities in a given market. That procedure worked for one market — in the Network GIV we generalize it to N markets: we consider N units such as industries, markets or countries that interact with each other, for example through customer-supplier relationships. An idiosyncratic shock to a particular unit may then also affect other units. We identify and estimate these idiosyncratic shocks, and use their propagation through the network to estimate causal parameters of interest, e.g. elasticities. For example, we want to study the impact of a TFP shock in the steel sector on TFP in other sectors, mediated throughout its first-order effect on customers and suppliers of steel firms, and its second-order effects on the other markets.

Intuition for the procedure We observe a panel of N units over T periods (N is fixed and we will take the large T limit). We consider a model of the form

$$Y_t = B_0(\gamma)Y_t + B_1(\gamma)Y_{t-1} + \lambda\eta_t + u_t. \tag{1}$$

Here Y_t is the observed vector of outcomes of dimension N. The $N \times N$ matrices $B_0(\gamma)$ and $B_1(\gamma)$ characterize the network (for example, input-output matrices) and are known up to an unknown low-dimensional parameter vector $\gamma^\star \in \mathbb{R}^p$ that we want to identify. For

instance, γ might encode elasticities, or intensities of influence. There is also a vector of (unobserved) idiosyncratic shocks u_t, and aggregate shocks η_t, with a vector of loadings λ. The key assumption is that u_t defines idiosyncratic shocks uncorrelated between units. The lagged term B_1 and the factors $\lambda\eta_t$ are not necessary for identification but included as they are often present in applications: one can have an i.i.d. model with $B_1(\gamma) \equiv 0$, and also no factors $\lambda\eta_t \equiv 0$.

We identify the target parameters using the following idea. Suppose, to build intuition, that $B_1(\gamma) = 0$, $\lambda = 0$ and we observed the idiosyncratic shocks $u_{j,t}$. Then in the regression $Y_{it} = e_i' B_0(\gamma) Y_t + u_{i,t}$ of individual i, we could use all shocks $u_{j,t}$ for all units j except for unit i to instrument the endogenous regressor Y_t. More precisely, we could use as an instrument for Y_t that value of Y_t that would be generated by the idiosyncratic shocks of units $j \neq i$ only (i.e., the expectation of Y_t conditional on all shocks $u_{j,t}$, $j \neq i$). This procedure would recover our target parameter γ. The underlying intuition is that all shocks $u_{j,t}$ are valid instruments except for the shock of individual i.

In practice, however, we do not observe the shocks u_{jt} directly. Nevertheless, using this intuition (and extending it to more complex scenarios), we can build valid moment restrictions and an iterative estimation procedure to identify and estimate both the target parameter $\gamma^\star$ and recover the idiosyncratic shocks. Given a guess of γ, we recover the idiosyncratic shock $u_{i,t}$, and given those idiosyncratic shocks, we can estimate γ. Our estimator is the fixed point of this iterative procedure. The procedure is laid out in Algorithm 1.

Intuitively, our moment conditions find the value of γ that best explains how a sector i is impacted by idiosyncratic shocks $u_{j,t}$ to the other sectors j ($j \neq i$); this is the key Network GIV moment, associated with the outcomes in sector i. We then aggregate each of these moment restrictions for estimation.

We study the asymptotic properties of our estimators in regimes with finite N and large T, and present a simple procedure to estimate $\gamma^\star$ and the variance of the estimator. In some relevant cases (see Example 4), the standard errors on the parameters are of order $\mathcal{O}(1/\sqrt{TN})$. Intuitively, this implies that even with a relatively short time series of length T, we can obtain precise estimates if the number of units N is large. This provides some suggestive evidence that our method can be applicable to a host of recent "big data" sets with large cross-sections though short time series (e.g. Andersen et al. 2026, Borusyak and

Hull 2023).

We present the basic Algorithm 1. Then, in Appendix (A), we extend it to the relevant case of missing data, and data corrupted by measurement error. That makes the algorithm more wildly applicable.

Applications We apply our method to study TFP spillovers in the US economy: If industry i has a positive productivity shock, how does the productivity of its suppliers and consumers change? This question has central relevance to the design of innovation policy and our understanding of economic growth (Liu and Ma 2021). It also faces a classic "reflection" problem (Manski 1993) — if TFP of i and j rise, does this comovement reflect a primitive innovation in i that affects j, an innovation in j that affects i, or do both sectors respond to a common macroeconomic force —- that our procedure is well-suited to handle, provided that sufficient conditions hold for identification. We construct a panel data set of production network linkages and TFP covering 65 industries over 1964-2023 and implement our procedure under the maintained assumption that the direct spillover from two sectors is proportional to the strength of the input-output link between them.

We find that a primitive 1% increase in the (gross output) TFP of each industry leads to a long-run increase of approximately 5% in aggregate TFP. This amplification is the product of two effects. First, the propagation of the primitive industry shocks through the network of suppliers and customers learning from each other results in a total average increase of 2.7% in industry-level gross output TFP — an externality factor of 2.7. This is the number that we estimate. Second, by Hulten (1978)'s theorem, the impact on aggregate value-added TFP is multiplied by the total sales to value added ratio, which is 1.8. The product yields the total impact of a 5× impact on aggregate TFP. This finding of a high externality (about 2.7×) echoes the results from previous studies that find that the social return to R&D far exceeds the private return (e.g. Griliches 1992; Hall, Mairesse, and Mohnen 2010; Bloom, Schankerman, and Reenen 2013). Our identification, though, is very different from those studies (which rely e.g. on changes in tax subsidies).

More generally, our procedure can be used in a host of situations. For instance, in international macroeconomics, one can ask: if country j has a supply or demand shock, what happens to each country i's demand and supply in the world economy? What about exchange

rates or other relevant outcomes? Using idiosyncratic shocks to countries, and seeing their effects throughout the world economy, the Network GIV can answer these questions. In business cycle macroeconomics, one could study how demand and supply shocks propagate throughout the production network, affecting prices and quantities. In finance, one can ask what's the impact of a demand or supply shock to one corner of the stock market (as inKoijen and Yogo (2019) and Gabaix and Koijen (2023)), on the stock market, its several components, and other connected markets, such as bonds, exchange rates, et cetera? These effects could be estimated using the Network GIV.[1]

Related literature We extend greatly (moving from one to N markets) the GIV method of Gabaix and Koijen (2024), which has been used in a number of papers (Chodorow-Reich, Ghent, and Haddad (2021), Adrian et al. (2022), Camanho, Hau, and Rey (2022), Galaasen et al. (2023), Gabaix and Koijen (2023), Qian (2024), Chaudhary, Fu, and Zhou (2025)) by using idiosyncratic shocks to large firms or sectors for identification. Baumeister and Hamilton (2023) propose a maximum likelihood (MLE) approach to the GIV, without, however, allowing for network effects. Different from these references, the presence of network effects challenges the identification and estimation strategy: we show that, because individual idiosyncratic shocks of a given unit are correlated with the endogenous regressors (the other units' outcomes), moment restrictions take a new form and solve a fixed point problem. Here, we opt for transparency, using moments that are close (in a sense we will make precise in Section 2.6.3) to the (pseudo) MLE, so to preserve efficiency if certain distributional assumptions hold, without violating validity if these fail. One advantage of our method however, with respect to MLE, is its robustness to the incidental parameter problem. As we discuss in Section 2.6.3, the MLE formulation may lead to inconsistent estimates for relevant data-generating processes in the presence, for instance of time fixed effects (Neyman and Scott (1948)). As we show, in relevant data generating processes, this does not occur for our estimator because it leverages specific moment conditions obtained through independence of shocks.

We more broadly connect to a large literature on vector autoregression, see for example

1. For instance, Gabaix and Koijen (2023) use the simple GIV to measure the impact of demand shock for stocks, on the price of stock themselves. But with the Network GIV we present here, one could extend this analysis to several interconnected markets.

Baumeister and Hamilton (2024) and Rubio-Ramírez, Waggoner, and Zha (2010). When these models are estimated via (pseudo) MLE these may suffer from similar issues in the presence of incidental parameter problems, different from our proposed estimator, which overcomes this difficulty. The literature on spatial autoregressive models also has studied similar problems using likelihood or quasi-likelihood based methods where interactive fixed effects may be concentrated out see e.g., Bai 2003 and Shi and Lee 2017, overcoming incidental parameter problems. Our main contribution is an iterative IV reformulation that gives interpretable estimators, first-stage diagnostics and overidentification tests, using explicit and (in the class of macro-economic applications we consider) more interpretable moment conditions that, instead of being based on the particular choice of the quasi-likelihood, directly leverage independence of individual level shocks.

Much of the recent empirical literature has turned to external instruments for identification (Mertens and Ravn 2013; Plagborg-Møller and Wolf 2022). While appealing in settings where they can be adequately defended, in many cases no obvious or systematic excluded instruments exist. Our approach generates instruments for panel VARs internally by recovering large idiosyncratic shocks using the known network structure; this gives it a potentially very large domain of applicability. Identification from idiosyncratic shocks also contrasts with a large body of work in empirical macroeconomics and related fields based on Bartik-style exposure designs (Borusyak and Hull 2023), where identification comes from heterogeneous exposure to common shocks. Our method complements and differs from this approach, as it does not require existence of a common exogenous shock (and in fact it removes the factors from the estimation); instead it leverages independence of unit-specific individual shocks. In spirit, it considers a recentered instrument as in Borusyak and Hull 2023 but different from them the recentering is a function of the parameters, therefore leading to an iterative estimation algorithm.

We build on the idea that the economy is "granular" in nature— due to incompressible "grains" of volatility coming from idiosyncratic shocks to firms or industries (Gabaix (2011); Acemoglu et al. (2012), Giovanni and Levchenko (2012), Giovanni, Levchenko, and Méjean (2014), Baqaee and Farhi (2019), Carvalho and Grassi (2019), and Gaubert and Itskhoki (2021)). This is also confirmed in multi-sector models in the tradition of Long and Plosser (1983), e.g. Atalay (2017), Brownlees and Mesters (2020), and De Graeve and Schneider

(2023).

We link to a large literature on networks. Manski (1993)'s influential work distinguished between different types of peer effects and pointed out challenges for identification in network setups with correlated effects (see Kline and Tamer (2020) for a review). Different from standard peer effects models as in Manski (1993) (and motivated by macro-economic applications), we identify the key parameters through idiosyncratic shocks, leveraging the fact that idiosyncratic shocks are uncorrelated between units. Identification of peer effects has then been extensively studied, see Bramoullé, Djebbari, and Fortin (2009), De Giorgi, Pellizzari, and Redaelli (2010), Lee (2007), among others. These methods consider a linear in means model (without a factor structure), and leverage identification by using either sufficient variation in the rows of the adjacency matrix together with sufficient variation of auxiliary observable covariates in the outcome equation, or variation in the size of non-overlapping groups. Using related identification strategies, Blume et al. (2015) consider settings where the adjacency matrix is only partially observed and Blume et al. (2011) propose (spatial) models of the adjacency matrix. Our approach complements these existing methods in general setups by parametrizing a single network as a function of a low dimensional set of parameters and leveraging valid moment restrictions that use uncorrelated idiosyncratic shocks, and do not require variation in additional auxiliary covariates.[2]Battaglini, Patacchini, and Rainone (2022) and Breza et al. (2020) propose Bayesian procedures to recover social networks, different from our approach here. De Paula, Rasul, and Souza (2019) present general identification results for a linear in means model leveraging variation in auxiliary covariates. Here, instead, we leverage assumptions on a low dimensional parametrization of the network structure and independence of the shocks. These assumptions allow us to identify effects through simple moment conditions that leverage exogeneity of the other units' idiosyncratic shocks. Our moment conditions can be easily extended to more complex models relevant in macroeconomic applications where a factor structure occurs. Finally, in recent work, Griffith and Peng (2023) study identification of the network assuming sparsity restrictions (e.g., only few countries may affect other countries). Although such restrictions can be plausible in micro-economics applications, these typically fail in macroeconomics where we observe

2. Relatedly, Manresa (2013) also considers settings with unknown adjacency matrix, but assumes no endogenous peer effects (i.e., the outcome does not depend on other units' outcomes), different from our framework.

dense networks. Here, we do not require such restrictions on the network, motivated by application in macroeconomics.

The rest of the paper is organized as follows. In Section 2, we present the basic econometric methodology. We present the main Algorithm 1, and a variety of extensions, e.g. to autocorrelations, heteroskedasticity, non-linearities, and missing data. Section 3 is our empirical application to the propagation of TFP shocks. Section 4 concludes. Most formal proofs are in the online appendices.

Notations We call $e_i = (0, \dots, 0, 1, 0, \dots, 0)'$ the unit vector with 1 at position i, 0 otherwise; $\iota = (1, \dots, 1)$ a vector of ones; $\mathbb{I}$ the identity matrix — all of the appropriate dimension; when needed, we use $\mathbb{I}_N$ for the identity matrix of dimension N. For given random variables $X_{i,t}$, we denote $X_t = [X_{1,t}, \cdots, X_{N,t}]'$ the vector of such variables. We define $\mathbb{E}[\cdot], \mathbb{V}(\cdot)$ the expectation and variance operator, respectively. We call $V^y = \mathbb{E}[y_t y_t']$ the variance of a random variable y_t, which is mean 0 and vector valued, and $W^y = (V^y)^{-1}$ the inverse variance. For two vectors, $x \odot y$ denote the component-wise product. We use the lag operator L, $Lx_t = x_{t-1}$.

2 Econometric Methodology: Main Ideas

Consider a panel with N units indexed by i and T periods indexed by t. Define $Y_{i,t} \in \mathbb{R}$ the observed outcome of interest. The key goal is to estimate a *low-dimensional* parameter $\gamma^\star$. The low dimensionality of $\gamma^\star$ embeds a large class of macro-economic applications, including our leading application in the following section. We will consider the case of a fixed N, but a large T.

2.1 Some examples

We study networks of the form (1). We start by presenting some examples, which will be useful to keep in mind when interpreting the general econometric framework.

Example 1 (Influence matrix linear in factors)**.** Consider the following model:

$$Y_t = \gamma \bar{B} Y_t + u_t, \tag{2}$$

where the matrix $\bar{B}$ is known and $\gamma \in \mathbb{R}$ is an unknown parameter, so that $B(\gamma) = \gamma \bar{B}$. For instance, $\bar{B}_{ij}$ could be a known "network strength" of the flows of information from j to i (e.g., how frequently i hears j's recommendations), and we are looking to estimate the "reactivity coefficient" γ. This model is similar to Bramoullé, Djebbari, and Fortin (2009), who study a model of the form $Y_t = \gamma \bar{B} Y_t + X_t \beta + u_t$, where X_{it} are observed covariates. But while Bramoullé, Djebbari, and Fortin (2009) needs to use covariates X_{it}, we do not need any covariates. Instead, we use the impact of idiosyncratic shocks u_{jt} in our instrumental variable method, and particularly of large "influencers" j, that have a large average influence $\frac{1}{N}\sum_i \bar{B}_{ij}^2$. This example generalizes to settings where

$$Y_t = \sum_{k=1}^{K} \gamma_k \bar{B}^{(k)} Y_t + u_t, \tag{3}$$

where the matrices $\bar{B}^{(k)}$ are observed by the econometrician and γ_k denotes a relative weight on the matrix $\bar{B}^{(k)}$, so that $B(\gamma) = \sum_{k=1}^{K} \gamma_k \bar{B}^{(k)}$. This would be the case if there are possibly different channels (matrices $\bar{B}_{ij}^{(k)}$), i.e. via tweets, or emails, or Facebook messages, and each could have a different influence γ_k.

Example 2 (Supply and Demand)**.** Take a supply and demand model:

$$q_{it} = \phi_i^d p_t + \eta_{1t} + u_{it} \tag{4}$$

$$p_t = \psi q_{St} + \eta_{2t} \tag{5}$$

Firm $i \in \{1, \ldots, n\}$ has log demand q_{it} reacting to the log price p_t with demand elasticity ϕ_i^d; and the price p_t that depends on the aggregate demand $q_{St} = \sum_{i=1}^{n} S_i q_{it}$ (where S_i are size shares) with inverse supply elasticity ψ. We assume that u_t is uncorrelated with η_t. This fits in the framework, with $Y_t = (q_{1t}, \ldots, q_{nt}, p_t)'$, $\gamma = (\phi^d, \psi)$, $B = \begin{pmatrix} 0 & \phi^d \\ \psi S' & 0 \end{pmatrix}$.[3] This shows how the framework generalizes that of Gabaix and Koijen (2024), easily allowing for heterogeneous elasticities of demand. One could specialize it to $\phi_i^d = \varphi^d \iota$, for a uniform elasticity φ^d, or to a parametric model $\phi_i^d = \dot{\varphi} X_i$, where $\dot{\varphi}$ is a vector of parameters to estimate, and each X_i represents firm i's characteristics, which are given. Appendix C

3. Here, the price's idiosyncratic shock is just 0, as η_{2t} could be correlated with η_{1t}.

details how the general procedure applies to this example.

Example 3 (Macroeconomic production network)**.** A macroeconomic production network will feature n sectors with price vector p_t and output vector q_t, and a link of the type (1), where $Y_t = (\pi_t, \Delta\hat{y}_t) \in \mathbb{R}^{2n}$ records the change in price and quantities. Importantly, the matrix $B(\gamma)$ typically depends non-linearly on the parameter γ, e.g. via mixtures of elasticities and speeds of adjustment. Section K develops such a model, see in particular Proposition8.

2.2 Most basic version of the procedure

In this section we present the basic model to convey the main intuition.

Assumption 1 (Basic model)**.** *For all $t \in \{1, \ldots, T\}$,*

$$Y_t = B(\gamma^\star)Y_t + u_t, \quad \mathbb{E}[u_{i,t}|\{u_{j,t}\}_{j\neq i}] = 0, \tag{6}$$

where $\gamma^\star \in \mathbb{R}^p$ is an unknown vector of parameters in a space $U \subseteq \mathbb{R}^p$, $B : U \to \mathbb{R}^{N\times N}$ is a known, continuously differentiable function and $A(\gamma^\star) = (\mathbb{I}_N - B(\gamma^\star))^{-1}$ exists.

Assumption 1 states that the outcome of each unit depends on weighted combinations of other units' outcomes and shocks. The dependence is captured through the parameter $\gamma^\star$. We assume that we know the functional form $B(\gamma^\star)$, but do not know $\gamma^\star$. The key identification assumption is that the shocks $u_{i,t}$ are mean independent (uncorrelated) across units i. The intuition is the following: a shock $u_{i,t}$ will affect all the units j at time t. Hence, by tracing out the impact of $u_{i,t}$ to units $j \neq i$, we are able to identify $\gamma^\star$.

To derive our results for inference, we will make the simplifying assumption that shocks are independent and stationary over time, although we could directly relax this assumption by assuming forms of mixing and covariance stationarity (omitted for brevity).

Assumption 2. *The u_t are independent and stationary over time t.*

Given the model $Y_t = B(\gamma^\star)Y_t + u_t$, we cannot regress directly $Y_{i,t}$ onto the other outcomes, because $E[e_i^\top B(\gamma^\star)Y_t|u_{i,t}] \neq 0$. We therefore need to instrument the endogenous regressor $e_i^\top B(\gamma^\star)Y_t$. To do so, we use a simple idea: we instrument the endogenous

regressor $e_i^\top B(\gamma^\star)Y_t$ using the unobserved shocks of all units except unit i, i.e., $\{u_{j,t}\}_{j\neq i}$. Following Chamberlain (1987), the construction of the optimal (efficient) instrument is a function of $E[e_i^\top B(\gamma^\star)Y_t|\{u_{j,t}\}_{j\neq i}]$ (as well as possible heteroskedastic components). However, without distributional assumptions and arbitrary heteroskedasticity, this construction is infeasible because the optimal instrument depends on the unknown vector of parameters $\gamma^\star$. We therefore propose an intuitive approximation to the optimal instrument that retains validity of the resulting moment conditions without requiring distributional assumptions on the unobserved shocks. Our idea proceeds as follows.

Suppose we have a guess γ, and we imagine that the discrepancy $\Delta\gamma := \gamma^\star - \gamma$ is small. As a first step, we use a Taylor expansion to write (where $B_\gamma(\gamma^\star)$ denotes the gradient of $B(\gamma)$ evaluated at the true parameter $\gamma^\star$)

$$u_t = (I - B(\gamma^\star))\, Y_t = (I - B(\gamma + \Delta\gamma))\, Y_t = (I - B(\gamma))\, Y_t - \Delta\gamma \cdot B_\gamma\,(\gamma)\, Y_t + O\left(\|\Delta\gamma\|^2\right)$$

Therefore defining the residual

$$\tilde{v}_t(\gamma) := (I - B\,(\gamma))\, Y_t \tag{7}$$

we have

$$\tilde{v}_t(\gamma) = \Delta\gamma \cdot B_\gamma(\gamma)Y_t + u_t + O(||\Delta\gamma||^2).$$

Hence, pre-multiplying by e_i the left and right hand side, we obtain

$$\tilde{v}_{it}(\gamma) = \Delta\gamma \cdot e_i^\top B_\gamma\,(\gamma)\, Y_t + u_{it} + O\left(\|\Delta\gamma\|^2\right). \tag{8}$$

The obstacle to running OLS and estimating $\Delta\gamma$ is that regressor $e_i^\top B_\gamma\,(\gamma)\, Y_t$ is typically correlated with u_{it}. So, as our second step, we replace $e_i^\top B_\gamma\,(\gamma)\, Y_t$ in the regression above with an instrument $z_{i,t}(\gamma)$ that approximates $E[e_i^\top B_\gamma\,(\gamma)\, Y_t|\{u_{j,t}\}_{j\neq i}]$. Such an instrument guarantees under independence of shocks across units that $E[z_{i,t}(\gamma)\tilde{v}_{it}(\gamma)] = 0$ when $\gamma = \gamma^\star$ If we can construct such an instrument, it follows that we can directly estimate $\Delta\gamma$ by OLS in a simple (and intuitive) regression:

$$\tilde{v}_{it}(\gamma) = \Delta\gamma \cdot z_{it}(\gamma) + \varepsilon_{i,t}, \quad E[\varepsilon_{i,t}|z_{i,t}(\gamma^\star)] = 0. \tag{9}$$

We construct the instrument $z_{i,t}(\gamma)$ as follows. Recall that $Y_t = A(\gamma^\star)u_t$. Therefore, we have $e_i^\top B_\gamma(\gamma) Y_t = e_i^\top B_\gamma(\gamma) A(\gamma^\star)u_t$. Hence, we set

$$z_{i,t}(\gamma) \coloneqq e_i^\top B_\gamma(\gamma) A(\gamma) u_t^{-i}(\gamma), \qquad u_t^{-i}(\gamma) \coloneqq (I - e_i e_i^\top)\tilde{v}_t(\gamma) \tag{10}$$

so $z_{i,t}$ is the "granular" instrument using only shocks to entities different from i – the GIV. It is strong when some units j are large enough to affect the outcome Y_{it} (this is intuitive, and will be verified analytically below).

We therefore obtain a simple and recursive estimation procedure where we first form a guess γ and then we update such a guess over each iteration by taking $\gamma + \Delta\gamma$ in the subsequent iteration. At convergence, the estimator provides us with a valid moment restriction for $\gamma^\star$.

This leads to Algorithm 1, which, for subsequent use, we present in a fairly general form (with factors and lags, which we discuss below) .

Algorithm 1 One-step estimation of $\gamma^\star$ via the Network GIV

Require: $\hat{\gamma}$, a guess about the value of $\gamma^* \in \mathbb{R}^p$; and r, the number of factors.

1: Define $\tilde{v}_t(\hat{\gamma}) \coloneqq Y_t - B_0(\hat{\gamma})Y_t - B_1(\hat{\gamma})Y_{t-1}$. Below, we remove the explicit dependences on $\hat{\gamma}$, but everything is evaluated using $\hat{\gamma}$.

2: Run PCA with heteroskedasticity using Algorithm 5 passing as the argument $\tilde{v}_t$ and as the number of factors r. The PCA algorithm iterates to estimate the loadings $\tilde{\lambda}$. Alternatively, to control for time fixed effects, apply PCA to the cross-sectionally demeaned residuals $\breve{v}_t$, as detailed in Section B.3.

3: For each i, estimate η_t^{-i} as the slope in the regression: $\tilde{v}_{jt} = \tilde{\lambda}_j \eta_t^{-i} + \tilde{u}_{jt}^{-i}$, where we only use the units $j \neq i$. We collect the residuals $\tilde{u}_{jt}^{-i}$. We also set $\tilde{u}_{it}^{-i} = 0$.

4: For each i, define $\tilde{y}_{0|-i} = 0$, and construct recursively the instruments $\tilde{y}_{t|-i}$, which are the values of y_t that are due to the idiosyncratic shocks from units $\neq i$:

$$\tilde{y}_{t|-i} \coloneqq (I - B_0)^{-1} \left(\tilde{u}_t^{-i} + B_1 \tilde{y}_{t-1|-i} \right). \tag{11}$$

5: For each i, t, construct the instrument $z_{i,t} \in \mathbb{R}^p$. Its $k-$th component is:

$$z_{i,t}^{(k)} \coloneqq e_i^\top \left(\frac{\partial B_0}{\partial \gamma_k} \tilde{y}_{t|-i} + \frac{\partial B_1}{\partial \gamma_k} \tilde{y}_{t-1|-i} \right) \tag{12}$$

6: Estimate the updated $\hat{\gamma}'$ as follows:

1. For all i, regress η_t^{-i} on $z_{i,t}$, and collect the residuals $\eta_t^{-i,\perp}$: $\eta_t^{-i} = \beta z_{i,t} + \eta_t^{-i,\perp}$.
2. Run the linear regression of $\tilde{v}_{it}$ on $z_{i,t}$, controlling for $\eta_t^{-i,\perp}$:

$$\tilde{v}_{it} = \Delta\gamma \cdot z_{i,t} + b_i \eta_t^{-i,\perp} + \varepsilon_{it} \tag{13}$$

   and estimate the slope vector $\Delta\gamma$. For efficiency, use GLS weights $\sigma_{u_i}^{-2}$.
3. Update $\hat{\gamma}$ as $\hat{\gamma}' \coloneqq \hat{\gamma} + \texttt{step}\, \Delta\gamma$ with $\texttt{step} = 1$ (or e.g. 0.3 initially).

return $\hat{\gamma}'$

We state two propositions giving theoretical guarantees for the procedure.

Proposition 1. *Assume that u_{it} are uncorrelated with u_{jt} for $j \neq i$ (but could have heteroskedastic variance or be dependent over time periods). At the correct $\gamma^\star$, we have $\mathbb{E}\left[\tilde{v}_{it}(\gamma^\star) z_{it}(\gamma^\star)\right] = 0$, where $\tilde{v}_t(\gamma) = Y_t - B(\gamma) Y_t$ and $z_{it}(\gamma)$ is the instrument in (10).*

The advantage of the proposed procedure is its simplicity (it only requires running a sequence of linear regressions), and the moment conditions to identify the target parameter of interest are defined solely based on the assumption that shocks are independent across units (therefore without imposing e.g., distributional conditions on the shocks).

How many parameters can we estimate? We observe that in principle (i.e., with large T), we can estimate a large number of parameters:

$$\text{Maximum \# of parameters } \dim \gamma \text{ that can be estimated: } p^{\max} = \frac{1}{2}N\left(N+1\right) - \left(r+1\right)N \sim \frac{1}{2}N^2 \tag{14}$$

with a model with $r = \dim \eta_t$ factors. The reason is because the covariance matrix of the shocks provide us with several moment restrictions (one moment restriction for each pair of residuals). Even $N = 50$ units, (14) instructs us that we could in principle estimate over 1,200 parameters, at least in the large T limit. Our empirical example will have $N = 65$ units.[4]

2.3 Enrichments

In this section we present a series of extensions of our baseline model, deferring more formal details to the appendices.

2.3.1 Adding common factors

Suppose next that $Y_t = B(\gamma)Y_t + \lambda\eta_t + u_t$, where λ indicates some (unobserved loadings) and η_t indicates unobserved factors. We define $\tilde{v}_t$ as above in (7). Whenever $\gamma = \gamma^\star$, we should have $\tilde{v}_t = u_t + \lambda\eta_t$.

As the first step, we run iteratively, for different units i, a version of heteroskedastic robust PCA to estimate a factor model for $\tilde{v}_{jt}$ for $j \neq i$: $\tilde{v}_{jt} = u_{jt} + \lambda_j\eta_t$ (see Section B.2). We collect the estimated factors η_t^{-i} and the residuals $\tilde{u}_{jt}^{-i}$, obtained from such factor models using all units except information from unit i (setting $\tilde{u}_{it}^{-i} \coloneqq 0$)

In addition, with factors, (8) becomes:

$$\tilde{v}_{it} = \Delta\gamma \cdot e_i^\top B_\gamma\left(\gamma\right) Y_t + u_{it} + \lambda_i\eta_t + O\left(\left\|\Delta\gamma\right\|^2\right) \tag{15}$$

Our idea proceeds as follows. We form the instrument $z_{it}(\gamma) \coloneqq e_i^\top B\left(\gamma\right) A\left(\gamma\right) \tilde{u}_t^{-i}\left(\gamma\right)$. Now,

4. Indeed, we have the variance matrix of Y_t, which is symmetric and $N \times N$, so has $d_{VY} = \frac{1}{2}N\left(N+1\right)$ degrees of freedom. We have also one (in the homoskedastic case) free parameters for σ_u^2, or N in the heteroskedastic case. So, we can estimate up to $p^{\max} = \frac{1}{2}N\left(N+1\right) - N$ parameters, if we have a very long time series. When we add r factors, as we will do later, this removes rN degrees of freedom, so that $p^{\max} = \frac{1}{2}N\left(N+1\right) - \left(r+1\right)N$.

we also need to control for the residuals η_t^{-i}. We first residualize, i.e. estimate β in $\eta_t^{-i} = \beta z_{i,t}(\hat{\gamma}) + \eta_t^{-i,\perp}$ and collect the residuals $\eta_t^{-i,\perp}$. Then, we simply run (13), which is the basic OLS regression, but controlling for $\eta_t^{-i,\perp}$. This is done in Algorithm 1.

2.3.2 Adding lagged values of Y_t

With a term $B_1(\gamma^\star)Y_{t-1}$ in the process (1), we naturally define $\tilde{v}_t$ as in Step 1 of Algorithm 1, and follow verbatim our procedure now also controlling for the lagged outcomes.

2.3.3 Additional covariates

In the presence of additional known covariates X_{it}, so that the model is

$$Y_t = B_0(\gamma^\star)Y_t + B_1(\gamma^\star)Y_{t-1} + X_t b_X + \lambda\eta_t + u_t \tag{16}$$

we use a natural extension of the basic procedure. Given γ, we estimate b_X by OLS of $Y_t - B_0(\gamma)Y_t - B_1(\gamma)Y_{t-1}$ on X_t; and given b_X, we proceed as in the rest of the paper, replacing $v_t \coloneqq Y_t - B_0(\gamma)Y_t$ by $v_t \coloneqq Y_t - B_0(\gamma)Y_t - X_t b_X$, in Step 1 of Algorithm 1 and its variants.[5]

2.3.4 Adding equilibrium relations

One can easily add equilibrium relations (e.g. supply equal demand) in the system. They don't need to be explicitly "substituted out", so that the system is clean. Appendix Cdetails this.

5. There can be in principle another instrument, namely (imagining that we know, or have already estimated b_X, for simplicity)

$$z_{it}^X \coloneqq e_i' B_\gamma A X_t^{-i} b_X, \tag{17}$$

If we think that the value of X_{jt} a truly exogenous (i.e. X_{jt} is uncorrelated with all u_{it}, at least when $j \neq i$), we get an extra instrument. Indeed regression (9), i.e. (13) in the Algorithm, we could have

$$\tilde{v}_{it} = \Delta\gamma \cdot z_{i,t} + \Delta\gamma^X \cdot z_{it}^X + b_i \eta_t^{-i,\perp} + \varepsilon_{it} \tag{18}$$

Hence, we have two instruments: $z_{i,t}$, using idiosyncratic shocks, and z_{it}^X, using covariates. Each of them gives a valid estimate of $\Delta\gamma$. At convergence, both should be 0.

2.3.5 Heteroskedastic residuals

If the u_{it} are heteroskedastic with variance $\sigma_{u_i}^2$, it is preferable to weight each regression by $\sigma_{u_i}^{-2}$, as in GLS.[6]

2.3.6 Missing data, and measurement error

The framework also handles missing data, and data measured with error. Section A develops this. We propose a "first pass" Algorithm 2 that is very simple to use, and while not completely rigorous, is a useful approach to construct a first guess for the value of the parameters. We also propose a rigorous Algorithm 3 for that leverages a Bayesian model on the measurement error distribution. In simulations reported in the appendix, for a calibration of our empirical illustration, we find that both algorithms exhibit competitive performance.

2.3.7 Nonlinear model

The basic model is linear in u_t, Y_t, though nonlinear in parameter γ. Appendix D.5 generalizes the procedure to non-linear specifications, e.g. $Y_t = \mathcal{A}(u_t + \lambda\eta_t; \gamma)$ where $\mathcal{A}(\cdot;\gamma) : \mathbb{R}^N \to \mathbb{R}^N$ is a general smooth function, that *could be nonlinear*. The logical is similar. This one, one doesn't need to loglinearize, one can in principle work with a nonlinear model.

2.4 Formal statements

We present more formal econometric results. We define

$$G^{(k)} = \frac{\partial B(\gamma)}{\partial \gamma_k}\Big|_{\gamma=\gamma^\star} A(\gamma^\star), \quad \Gamma_{i,j}^{(k)} = G_{i,j}^{(k)} 1_{i\neq j}. \tag{19}$$

so that Γ gives the marginal network impact "removing self-impacts", as $\Gamma_{ii} = 0$, for a marginal change in γ. We also define $u_t(\gamma) \coloneqq (I - B(\gamma)) y_t$. Then, in our basic Algorithm,

$$z_{it}^{(k)} = e_i^\top \Gamma^{(k)} u_t(\gamma) \tag{20}$$

6. The identifying moment (22) becomes $\frac{1}{T}\sum_t u_t^\top(\gamma) W^u \Gamma^{(k)}(\gamma) u_t(\gamma) = 0$ with $W^u = Diag\left(\sigma_{u_i}^{-2}\right)$.

Define unit i's empirical moment restriction

$$\bar{H}_i(\gamma) = \frac{1}{T}\sum_t u_{it}(\gamma)\, z_{it}(\gamma) \tag{21}$$

The moment for $\Delta\gamma = 0$ is $\sum_i \bar{H}_i(\gamma) = 0$, i.e. more compactly:

$$\frac{1}{T}\sum_t u_t^\top(\gamma)\,\Gamma^{(k)}(\gamma)\,u_t(\gamma) = 0, \quad k \in 1, \cdots, p \tag{22}$$

We define $\mathbb{E}[u_{i,t}^2] = \sigma_{u_i}^2$. Algorithm 1 uses a regression on z_{it}. Hence, its variance should be non-degenerate. We next derive the variance of z_{it}, and assume that it is full rank.

Lemma 1. (Variance of z_{it}) *Suppose that the u_{it} are uncorrelated across periods, and covariance-stationary over time. Define $\sigma_u^2 = \text{Diag}\left(\sigma_{u_i}^2\right)$. The following holds:*

$$\sum_i \mathbb{E}\left[z_{it}(\gamma^\star) z_{it}(\gamma^\star)^\top\right] = V^z, \tag{23}$$

where the $p \times p$ matrix V^z is, for k, k' in $\{1, \ldots, p\}$:

$$V^z_{p_1,p_2} = \sum_{i,j} \Gamma_{ij}^{(p_1)} \Gamma_{ij}^{(p_2)} \sigma_{u_j}^2 \tag{24}$$

Proof. See Appendix D.1.2. □

We will make the following assumption.

Assumption 3 (Moment restriction)**.** *Assume that for each i, $H_{i,t}(\gamma)$ is continuously differentiable almost surely, $\mathbb{E}\left[\sup_\gamma ||H_{i,t}(\gamma)||\right], \mathbb{E}\left[\sup_\gamma \left|\left|\frac{\partial H_{i,t}(\gamma)}{\partial \gamma}\right|\right|\right] < \infty, \mathbb{E}\left[\sup_\gamma \left|\left|\frac{\partial^2 H_{i,t}(\gamma)}{\partial \gamma^{(p_1)} \partial \gamma^{(p_2)}}\right|\right|\right] < \infty$, for all $p_1, p_2 \in \{1, \cdots, p\}$, $\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N H_{i,t}(\gamma^\star)\right)\left(\frac{1}{N}\sum_{i=1}^N H_{i,t}(\gamma^\star)\right)^\top\right]$ is positive definite, and $\frac{1}{N}\sum_{i=1}^N H_{i,t}(\gamma^\star)$ has finite third moment. Suppose in addition that $-\sum_{i=1}^N \frac{\partial \mathbb{E}[\bar{H}_i(\gamma)]}{\partial \gamma}\Big|_{\gamma=\gamma^\star}$ is invertible and there exists a unique $\gamma^\star \in \mathcal{S}$, such that $\sum_{i=1}^N \mathbb{E}[\bar{H}_i(\gamma^\star)] = \mathbf{0}$, and $\gamma^\star$ is in the interior of $\mathcal{S}$.*

Assumption states uniqueness of the solution and invertibility of the derivative of the score function. Uniqueness is often not guaranteed for arbitrary parameter space, e.g., identification often requires sign restrictions on the parametersDe Paula, Rasul, and Souza

(2019). Here, we assume that the parameter is (locally) identified given the set $\mathcal{S}$. The remaining conditions guarantee that the central limit theorem applies to our setting.

Proposition 2. (Estimator variance) *Suppose that Assumptions 1,2, and 3 hold. Then as* $T \to \infty$,

$$\sqrt{T} V(\gamma^\star)^{-1/2} \Big(\hat{\gamma}^\star - \gamma^\star \Big) \to^d \mathcal{N}(0, \mathbb{I}) \tag{25}$$

for some positive definite matrix $V(\gamma^\star)$. *In addition, if* u_{it} *are homoskedastic across units,*

$$V(\gamma^\star) = W(\gamma^\star)^{-1}, \qquad W_{p_1,p_2}(\gamma^\star) = \sum_{i,j} \left(\Gamma_{i,j}^{(p_1)} \Gamma_{i,j}^{(p_2)} + \Gamma_{i,j}^{(p_1)} \Gamma_{j,i}^{(p_2)} \right) \tag{26}$$

where Γ *is defined in* (19).[7]

Proposition 2 follows directly from standard arguments for extremum estimators (Newey and McFadden 1994). The exact expression for the variance is derived in Appendix D.1.3. Note that Proposition 2 considers settings where T grows, and it holds for any (finite) N. When N grows, the convergence rate also depends on how $V(\gamma^\star)$ converges to zero as a function of N (see e.g., Example 4, here omitted for brevity).

The expression for the variance provides us with a concrete description of the (local) identification conditions required by our model, which therefore impose restrictions (as for the OLS formulation) on the trace of the matrix obtained from $\Gamma\Gamma^\top$.[8] We see this as a simple to verify condition as it only depends on the $B(\gamma)$ matrix and not on the variance of the shocks.[9]

In the following lines we illustrate that under gaussianity and homoskedasticity of the shocks, our estimator is not only consistent but also efficient (note that similar arguments also hold with heteroskedastic shocks, by using a two-step generalized least squares estimator as opposed to the simple OLS estimator proposed here). This efficiency property, together

7. With heteroskedasticity across units, we get the immediate generalization (e.g. via dimensional analysis): $W_{p_1,p_2}(\gamma^\star) = \sum_{i,j} \left(\Gamma_{i,j}^{(p_1)} \Gamma_{i,j}^{(p_2)} \frac{\sigma_{u_j}^2}{\sigma_{u_i}^2} + \Gamma_{i,j}^{(p_1)} \Gamma_{j,i}^{(p_2)} \right)$

8. We only consider growing T in Proposition 2 for expositional convenience (whereas the proposition is valid for any finite N). In the presence of growing N, we should explicitly consider a positive lower bound on the matrix $\mathbb{E}\Big[\Big(\frac{1}{N} \sum_{i=1}^N H_{i,t}(\gamma^\star) \Big) \Big(\frac{1}{N} \sum_{i=1}^N H_{i,t}(\gamma^\star) \Big)^\top \Big]$ that may shrink with N and adjust the rate of convergence accordingly based on this lower bound. Here, we directly rescale by the variance $V(\gamma^\star)$ in the left-hand side of Equation (25) without being explicit about the rate of convergence in N when N grows.

9. The lack of dependence of the estimator's variance on σ_u^2 is because σ_u^2 determines the variance of both the endogenous regressor and the instrument.

with its robustness under misspecification of the likelihood function, motivates our proposed procedure.

Corollary 1. *Suppose that the conditions in Proposition 2 hold and, in addition, u_{it} are jointly Gaussian and homoskedastic across units i and time periods t. Then Network GIV's variance achieves the Cramer-Rao lower bound.*

Corollary 1 states that under Gaussianity of the shocks, the proposed procedure is also efficient. The proof is derived in Appendix D.2.

We conclude this discussion with a comparison of the variance of our estimator with the "naive" variance computed in the last iteration of the OLS regression.

The "nominal" standard variance returned by the OLS software (applying the usual OLS formula in (13)) in the homoskedastic case is given by $V^{\gamma,OLS} = (V^z)^{-1}$. Therefore, as we may expect, the OLS regression underestimates the variance by an additional component $\text{tr}\left(\Gamma^{(p_1)}\Gamma^{(p_2)}\right)$, in the homoskedastic and Gaussian case. Interestingly, that additional term is very small compared to the first term in some leading examples, see the Example 4 below. So, the OLS standard error is a good (and slightly conservative in that case) approximation of the true standard error.

2.5 Examples illustrating the precision of the estimator

We now develop some examples that build the intuition for the precision of the estimator.

Example 4 (Intuition for the standard error in the estimation of γ: A simple example)**.** Consider Example 1 with $\bar{B}_{i,i} = 0$ and $\bar{B}_{i,j} = c_{i,j}S_j$ for constants $c_{i,j}$, and $\sum_j S_j = 1$. That is, different countries may have different "spheres of influences", parametrized by $c_{i,j}$, and S_j (some of which may also be unknown). Suppose that $\gamma^\star = 0$, and known S, c. In this case, under homoskedasticity, we obtain that the gradient of the moment equation, and in particular $\sum_{i=1}^N \frac{\partial \mathbb{E}[\bar{H}_i(\gamma)]}{\partial \gamma}\Big|_{\gamma=\gamma^\star} / \sigma_u^2 = \sum_i \sum_{j\neq i} c_{i,j}^2 S_j^2 + c_{i,j}c_{j,i}S_iS_j$. Suppose now that $c_{i,j}$ is iid across i, j with quadratic mean $v_c = \left(\mathbb{E}[c_{i,j}^2]\right)^{1/2}$ independent of (i, j). We can write, by taking expectations over c, the expression W in (27) reads as $(N-1)v_c^2h^2 + v_c^2$ where $h^2 = \sum_j S_j^2$ denotes the Herfindahl of country sizes. This implies that we expect the gradient to be of order $Nv_c^2h^2$ for large N. Suppose we are in the "granular" case where there are a few large

units, so that h remains of order unity even as N is large (for instance, a fixed $S_j > 0$ for a given/small number of units, even as N become large).[10] Then the variance $V(\gamma^\star)$, that is obtained through the standard Sandwich formula (and whose properties therefore depend on the gradient of the moment equation) is of order $\mathcal{O}(1/N)$, and the estimator's convergence rate is $1/\sqrt{TN}$, and more precisely:[11]

$$Stderr\left(\gamma\right) \sim \frac{1}{\sqrt{NT}v_c h} \tag{27}$$

This hints at the possibility that even with a not relatively large T, but a large cross-section N, one might reach a good precision.[12]

We can generalize the above example to a generic network.

Example 5 (Intuition for the standard error in the estimation of γ: Generalization of Example 4)**.** Consider now a generic network, with marginal impact G_{ij} defined in (19) and one-dimensional γ. To define a "Herfindahl" as in Example 4, we set

$$I_j \coloneqq \sqrt{\frac{1}{N-1}\sum_{i\neq j} G_{ij}^2}, \qquad \bar{v}_c \coloneqq \sum_j I_j, \qquad \bar{S}_j \coloneqq \frac{I_j}{\bar{v}_c}, \qquad h \coloneqq \sqrt{\sum_j \bar{S}_j^2} \tag{28}$$

so that I_j is mean influence of entity j, $\bar{S}_j$ is the relative size (in term of influence), so that $\sum_j \bar{S}_j = 1$, and $h^2 = \sum_j \bar{S}_j^2$ is the Herfindahl; so that the typical influence is $I_j = \bar{v}_c \bar{S}_j$: $\bar{v}_c$ gives the scale of the influence, controlling for normalized size S_j of the origin. Indeed, in the Example 4 above, this gives (for large N, so we can use $\frac{1}{N-1}\sum_{i\neq j} c_{ij}^2 \to v_c^2$) $I_j = v_c S_j$, $\bar{v}_c = v_c$, and $\bar{S}_j = S_j$. We take the case $\sum_{i\neq j} G_{ij}^2 \gg \sum_{i\neq j} G_{ij}G_{ji}$, which was true in Example 4. Then, the standard error of the estimator is still (27), but with those definitions of the Herfindahl h and average impact $\bar{v}_c$.[13]

Likewise, this would generate to each component of a multidimensional γ, provided the influence of the components of γ are sufficiently independent (i.e. the $V(\gamma^\star)$ matrix in (27)

10. Or, h could decay very mildly with N, as in the case of a fat-tailed distribution of sizes (Gabaix (2011)).

11. Hence, if $S_j = 1/N$ for all j, the variance $V(\gamma^\star)$ is of order $\mathcal{O}(1)$, and the estimator convergence rate much slower, as $1/\sqrt{T}$.

12. Suppose that $G_{ij} = S_j$ with $\sum_j S_j = 1$, so that S_j are influence shares. Call $h^2 = \sum_j S_j^2$ the Herfindahl. Then, $\sum_{i,j} \Gamma_{ij}^2 = (N-1)h^2$, while $\sum_{i,j} \Gamma_{ij}\Gamma_{ji} = \sum_{i\neq j} S_j S_i = 1 - h^2$. With large N and a fat-tailed distribution keeping $Nh^2 \to \infty$ for large N, the first term dominates the second one.

13. Indeed, the variance W in (26) is $W = \sum_{i\neq j} G_{ij}^2 = (N-1)\,\bar{v}_c^2 h^2$, where $h^2 = \sum_j \bar{S}_j^2$ is the Herfindahl.

is sufficiently close to diagonal).

In applications, we recommend using those expressions (28) to evaluate the degree of granularity of the network, and (27) to have a rough sense of the standard error that can be expected, e.g. before doing a thorough investigation.

Results with factor models The appendix (Section D.3) provides propositions for the more complex case with factor models and regressors and discusses inference in those scenarios.

2.6 Discussion

In this section, we discuss how to use the algorithm in practice, and also its robustness.

2.6.1 Overidentifying restrictions

The model has many overidentifying restrictions: each individual entity idiosyncratic shock leads to a different moment. This implies that we could use such restrictions to enrich the model when the model poorly approximates the data, following in spirit Gabaix and Koijen (2024), section III.C.

For example, one simple way is to decompose $z_{i,t}$ into subparts, e.g. $z_{i,t}^{(1)}$ which uses idiosyncratic shocks to the odd firms (say, heavy industry firms), and $z_{i,t}^{(2)}$ for even firms (say, light industry firms). Then, at convergence, we should have that regressing the generalization of (13):

$$\tilde{v}_{it}\left(\hat{\gamma}\right) = \Delta\gamma^{(1)} \cdot z_{i,t}^{(1)}(\hat{\gamma}) + \Delta\gamma^{(2)} \cdot z_{i,t}^{(2)}(\hat{\gamma}) + b_i \eta_t^{-i,\perp} + \varepsilon_{it}$$

should give both $\Delta\gamma^{(1)} = \Delta\gamma^{(2)} = 0$. Testing whether these coefficients are zero allows us to test whether we should consider heterogeneity in firms' characteristics in the construction of the parameters γ. Following similar approaches, one can see how to progressively enrich a model.

2.6.2 Threats to identification

The main threat to identification is that the model is misspecified. In particular, that the factor model is wrong — so that two entities covary not because of the network, but because

of an extra common shock that is not captured by the factor model.

One remedy is to extend the factor model progressively, by seeing that the results are not materially sensitive to adding more factors; or by using a formal test of the number of factors.[14]

Another diagnostic is to test overidentifying restrictions, as in Section (2.6.1). One can also narratively check the idiosyncratic shocks, esp. when one has high frequency data, e.g. for oil supply shocks, or TFP shocks.

2.6.3 Connection to MLE

It is important to compare our procedure to MLE for two reasons: (i) MLE is efficient under correct model specification; but (ii) MLE can be sensitive to misspecification of the model as discussed below. In Section D.2 of the online appendix we show that (pseudo) MLE that assumes homoskedastic and temporally uncorrelated shocks is equivalent to our procedure in the absence of unobserved factors and loadings. This result is suggestive that, despite the higher interpretability of our chosen procedure, our procedure retains efficiency when the true model does not have factors and unobservables are homoskedastic.

This comparison suggests that it is possible (and sometimes desirable) to use MLE in a Network GIV by being explicit on the exogeneity assumptions of the model. For instance Baumeister and Hamilton (2023) propose a decomposition of the MLE with emphasis on the"granular IV" terms.

MLE has several advantages: it is a well-understood, widely-used procedure; it is easy to flexibly enrich, e.g. via measurement error; it comes with various tests.

However, we can think of three main advantages of our procedure for applications. The first one is interpretability and simplicity in the estimation, as the procedure is explicit on which exogeneity assumption (through shocks) drives identification. The second one is robustness, as our procedure requires weaker conditions on factor models compared to MLE. The third one is flexibility: as we better understand what drives identification, we can modify our procedure appropriately to incorporate information from the exogenous components only. To be more precise:

14. When there are transient factors (e.g. Covid), for which a new factor structure cannot be realistically estimated, then one can just not use those observations in the sample – not use them to construct the instruments, or in the regression (13).

1. It is more robust when we have factors, and in particular, different from MLE, it is robust to the incidental parameter problem with e.g., time fixed effects. That is, if one interprets factors as fixed effects (which therefore do not require assumptions on the correlation of factors with covariates), in some common simple yet intuitive data generating processes, the MLE score function does not identify the parameter of interest. Therefore, to achieve identification, MLE must posit assumptions on the correlation between factors and covariates (interpreting factors as correlated random effects). This is undesirable when such correlation is unknown (e.g., when unobserved economic shocks impact observed prices) whereas our method does not require assumptions on the correlation between factors and observable covariates. In particular, different from MLE, our identification strategy leverages independence of shocks to build moment restrictions that depend on the parameter of interest γ, whereas MLE builds moment restrictions that lack identification of γ. We provide formal details in Appendix D.2.[15]

2. It is easier to interpret what drives identification: in the main case, this is idiosyncratic shocks to all (and in particular, large units). This follows in spirit the discussion inAndrews et al. (2025), who highlight the relevance of using different sources of exogenous variation for identification of models. Our approach allows us to be explicit about the particular source of exogenous variation we leverage for identification: independence between unit-specific unobservable shocks. But one can easily refine the identification to use only some of the units. For instance, we might select "well-measured" countries to form z_{it} (just using the u_{jt} we trust), and we still have a valid instrument. For instance, in the simplest case, we form

$$z_{i,t} = e_i^\top B_\gamma A \sum_j v_{jt} 1_{j \in \mathcal{N}_{it}}$$

15. Relatedly, our method leverages a key intuition: we can write the variance-covariance matrix of the observed outcomes, appropriately weighted by $I - B_0$ as the sum of a low rank matrix and a diagonal matrix. Instead of estimating jointly both the low-rank and diagonal matrix (which would amount to estimating a number of parameters equal to the overall sample size under arbitrary heteroskedasticity of the shocks), our method focuses on the off-diagonal entries to construct valid moment conditions. This turns a potentially high-dimensional number of moment conditions into an exactly identified system, that can be easily estimated in practice.

where $\mathcal{N}_{it}$ is a set of shocks we "trust", e.g. coming from countries in which we have very good data, or historical confirmation that we had indeed an idiosyncratic shock.[16] Likewise, we saw in Section 2.6.1 how to choose different "subsets of shocks" for the instruments.

3. Related to interpretability, one can interpret the instruments through a first stage and second stage (see the empirical application), which naturally provide with robustness checks.

In sum, our procedure is arguably more transparent, parsimonious and robust (for more complex models) than standard MLE procedures (while retaining efficiency guarantees in simple models).

2.6.4 Invariance by representation

One may wonder: does it matter which representation one uses? e.g. in (5), perhaps we can put supply on the left-hand side and price on the right-hand side, as $q_{St} = \frac{p_t - \eta_{2t}}{\psi}$? perhaps some recombination of the variables changes the result? The representation actually does not matter, as the next proposition shows.

Proposition 3. *The procedure outlined above is representation-invariant, in the following sense. Suppose that we have a representation* $Y_t = B(\gamma) Y_t + \lambda \eta_t + u_t$*; and another representation, using* $\tilde{Y}_t \coloneqq \Omega Y_t$*, where* Ω *is an* $N \times N$ *matrix, which leads to* $\tilde{Y}_t = \tilde{B}(\gamma) \tilde{Y}_t + \tilde{\lambda} \eta_t + u_t$ *with the same substantial outcomes (so* $I - B(\gamma) = \Omega^{-1} \left(I - \tilde{B}(\gamma) \right) \Omega$*). Then, the procedure, applied to* (Y_t, B) *or* $\left(\tilde{Y}_t, \tilde{B} \right)$*, yields at convergence the same estimate of* γ*.*

The conclusion is that one can write the system as one pleases (with some variable on the left vs right-hand side, recombining or not variables), to guide intuition, and perhaps to accelerate convergence. The result will be the same.

16. This is analogous to the "narratively checked" shocks in Gabaix and Koijen (2024).

3 Application: Propagation of TFP in the US

3.1 Setup

We present an empirical application: how does a productivity shock in one sector affect its "neighbors" (e.g. customers or suppliers)? This question has central importance to our understanding of economic growth and to optimal innovation policy. It also faces a "reflection" problem (Manski 1993) that previous literature has addressed by arguing for specific excluded instruments in the determination of research and development and TFP (e.g., tax changes in Bloom, Schankerman, and Reenen 2013) or timing assumptions that rule out contemporaneous spillovers (e.g. Liu and Ma 2021). The Network GIV offers an alternative, general approach to measuring the strength of such spillovers given a network. Specifically, we assume that TFP spillovers between two sectors depend on the strength of the economic link between the two sectors, multiplied by a coefficient γ to be estimated (and depending on the direction of the trade).

To make this idea operational, we call Y_{it} the TFP growth of sector i. We model as usual:

$$Y_t = \sum_{h=0}^{2} B_h Y_{t-h} + \lambda \eta_t + u_t. \tag{29}$$

We call x_i the *value* (i.e. price times quantity, in constant dollars) of the output of sector i, and $x_{ij} = x_{i \leftarrow j}$ the value of inputs from sector j used by sector i. We define the input-output matrix:

$$\Omega_{ij} = \frac{x_{ij}}{x_i}, \tag{30}$$

where entry i, j contains how much of j is used by i, as a share of i's total sales. We also define the "up" matrix,

$$\Omega^{\text{up}}_{ji} = \frac{x_{ij}}{x_j}$$

i.e., the fraction of sector j's output absorbed by downstream sector i.

We parametrize the influence matrix B_h at lag $h = 0, 1, 2$ as:

$$B_h(\gamma) = \gamma^{(h)}_{\text{down}} \Omega + \gamma^{(h)}_{\text{up}} \Omega^{\text{up}}, \tag{31}$$

The scalars $\gamma_{\text{down}}^{(h)}, \gamma_{\text{up}}^{(h)}$ are weights on two different channels: Ω captures the influence of the supplier's TFP and Ω^{up} the influence of the customer's TFP. Economically, these "influences" may be microfounded via "learning by closeness", like in modern growth theory (Romer 1986; Aghion and Howitt 1992; Jones 1995; Acemoglu 2009).[17]

3.2 Impact multipliers

We next define some metrics to evaluate the propagation of shocks.

3.2.1 Long run impact matrix

We construct the long run impact matrix as:

$$\mathcal{I} \coloneqq \bar{\mathcal{I}}(1) = (I - B_0 - B_1 - B_2)^{-1} \tag{32}$$

For instance, the long run impact of a TFP shock y_j in industry j on the TFP of industry i is $\mathcal{I}_{ij} y_j$.[18] The total impact of a 1% shock to all industries on industry i is $\sum_j \mathcal{I}_{ij}$ times 1%.

3.2.2 TFP multipliers

To give an average aggregate impact, we weight spillovers by economic importance, using S_i to denote the size of the sector i. In practice we equate S_i with the sales of sector i (as per Hulten's theorem); but, depending on the context, one might wish to use value added, or even just "1" to report equal-weighted averages.

The *aggregate multiplier* is defined as follows: if all sectors have a 1% primitive productivity shock, then by how much does total productivity increase (in percent)? It is:

$$M^{\text{tot}} = \frac{\sum_{i,j} S_i \mathcal{I}_{ij}}{\sum_i S_i} \tag{34}$$

17. Modeling spillovers as determined by the strength of input-output linkages distinguishes our approach from work that looks at ex ante technology similarity as measured by patent class overlap (Jaffe 1986; Bloom, Schankerman, and Reenen 2013; Liu and Ma 2021). Of course, these channels complement each other. We also focus on TFP, a broader measure of technological progress than patents.

18. Indeed, we have (omitting aggregate factors η_t for simplicity) $Y_t = \bar{\mathcal{I}}(L)\, u_t$, where L is lag operator ($Lx_t = x_{t-1}$) and $\bar{\mathcal{I}}(L)$ is the impact matrix (with dimension $N \times N$):

$$\bar{\mathcal{I}}(L) \coloneqq \left(I - B_0 - B_1 L - B_2 L^2\right)^{-1} \tag{33}$$

So, the long run impact on cumulative productivity is $\frac{d}{du_t} \sum_{h=0}^{\infty} Y_{t+h} =: \mathcal{I}$.

The *self TFP multiplier* is defined as follows: if a sector has a 1% primitive productivity shock, then by how much does its productivity increase after the shock propagates through the network? It is:

$$M^{\text{self}} = \frac{\sum_i S_i \mathcal{I}_{ii}}{\sum_i S_i} \tag{35}$$

The aggregate network amplification is $\frac{M^{\text{tot}}}{M^{\text{self}}}$.

3.2.3 Hulten effect

The effect on net TFP (i.e. on GDP, keeping constant factor inputs) also involves a Hulten (1978) effect:[19] Given TFP growth $\hat{Y}_i$ in sector i, whose sales are S_i, the change in aggregate TFP $\hat{y}$ is

$$\hat{y} = \sum_i \frac{\text{Sales}_i}{\text{GDP}} \hat{Y}_i = \frac{\sum_i S_i}{\text{GDP}} \frac{\sum_i S_i \hat{Y}_i}{\sum_i S_i} = M^{\text{Hulten}} \frac{\sum_i S_i \hat{Y}_i}{\sum_i S_i} \tag{36}$$

with

$$M^{\text{Hulten}} = \frac{\sum_i \text{Sales}_i}{\text{GDP}} \tag{37}$$

In practice, $M^{\text{Hulten}} \simeq 1.8$. Hence, if all firms have a gross productivity growth of 1%, aggregate net TFP growth is 1.8%.[20]

3.2.4 How much of the action comes from upstream vs downstream effects?

How much are the effects due to upstream vs downstream learning? To do that, we use the Shapley value on the total multiplier M^{tot} . Recall that for a function $f(x, y)$, the Shapley value provides an answer to "how much of the value of $f(x, y)$ is due to x, controlling for the value of $f(0, 0)$":[21]

$$S^{x,\text{Shapley}} = \frac{1}{2}\left(f(x, y) - f(0, y) + f(x, 0) - f(0, 0)\right). \tag{38}$$

19. See Gabaix (2011), Appendix B, for a pedagogical derivation of Hulten's theorem, which showed its usefulness in thinking about networks. Baqaee and Farhi provided many enrichments (Baqaee and Farhi 2019; 2020).

20. Why? Because the production of each firm goes up by 1% of sales, and that 1% of sales can go into final demand (say, consumption): if the sum of sales is 1.8 times GDP (which is net consumption in an economy without capital), this is a 1.8% increase in consumption, i.e. in GDP.

21. For instance, if $f(x, y) = g(x) + h(y)$, the Shapley value "coming from x" is $S^{x,\text{Shapley}} = g(x) - g(0)$, reassuringly.

We apply the Shapley value to the log multiplier, $\ln M(\gamma_{\text{down}}, \gamma_{\text{up}})$ (we drop the "tot" superscript for simplicity). Hence, we get for the multiplier coming from each linkage direction (and using $M(0,0) = 1$):

$$M_{\text{down}} = \exp\left[\frac{1}{2}\left(\ln M(\gamma_{\text{down}}, \gamma_{\text{up}}) - \ln M(0, \gamma_{\text{up}}) + \ln M\ (\gamma_{\text{down}}, 0)\right)\right]$$

$$M_{\text{up}} = \exp\left[\frac{1}{2}\left(\ln M(\gamma_{\text{down}}, \gamma_{\text{up}}) - \ln M(\gamma_{\text{down}}, 0) + \ln M(0, \gamma_{\text{up}})\right)\right]$$

By construction

$$M = M_{\text{down}} M_{\text{up}} \tag{39}$$

3.3 Data

Our data set consists of an annual panel of 65 industries over the period 1964-2023. The main data source is Bureau of Economic Analysis (2023). Data on productivity, industry gross output value and final "consumption" expenditure (which also includes investment and exports) come directly from these accounts. We demean the productivity series before the estimation.

We construct the input-output matrix Ω as follows. The "Use" table U reports the dollar amount of each commodity used by each industry. For years starting in 2007, Bureau of Economic Analysis 2023 also provides an import use table U^M; for years before 2007, we provisionally fix the import share of commodity c used in industry j to its 2007 level and then scale up or down all imports of c to match aggregate imports of c as reported in U. To convert use of commodity c into use of output of industry i, we assume that the share of commodity c that j purchases from i is equal to the share of total commodity output of c produced by industry i. We normalize the make table M so that entry i, c gives the fraction of commodity c produced by i. We then define the domestic input-output matrix $\Omega = M\left(U - U^M\right)$. We fix Ω at its 2007 value for our baseline estimation and report robustness for other years; in practice the input-output network is sufficiently stable that the year matters little to the results. In general, one could use the input-output matrix from a previous year.

Table 1: TFP propagation: estimated parameters

Model	$\gamma_{\text{down}}^{(0)}$	$\gamma_{\text{up}}^{(0)}$	$\gamma_{\text{down}}^{(1)}$	$\gamma_{\text{up}}^{(1)}$	$\gamma_{\text{down}}^{(2)}$	$\gamma_{\text{up}}^{(2)}$
0 Lag	0.04	0.74				
	(0.10)	(0.10)				
1 Lag	0.10	0.68	0.15	0.15		
	(0.10)	(0.11)	(0.10)	(0.11)		
2 Lags	0.09	0.69	0.18	0.15	0.11	0.24
	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.11)

This table reports estimates of the TFP spillover coefficients for three models with progressive number of lags. The model is Equation 29 with three common factors and with B matrices in Equation 31. The estimation uses heteroskedastic PCA as in Algorithm 6. Standard errors are in parentheses.

3.4 Empirical results

3.4.1 Estimated model

We estimate the model with 2 lags, and 3 factors $\eta_t \in \mathbb{R}^3$. Adding additional lags or factors does not make a big difference. Table 1 shows the results for different model specifications. In the first we hold all lags muted, in the second we introduce the first lag and the third is the full specification with two lags. We estimate the supplier and customer coefficients γ_{down} and γ_{up}. Using the estimated parameters, we then build B_0, B_1 and B_2 to obtain the long run impact matrix using the closed form expression evaluated at $L = 1$ and the impulse responses by iterating the implied VAR(2) recursion horizon-by-horizon, as in Section 3.2. Multipliers are then obtained using the sales weights S_i.

3.4.2 Implied long-run multipliers

In the full specification with 2 lags and 3 factors, the multipliers are:

$$M^{\text{tot}} = 2.7 \text{ (SE 0.83)}, \qquad M^{\text{self}} = 1.2 \text{ (SE 0.05)} \tag{40}$$

implying an aggregate network amplification factor $M^{\text{tot}}/M^{\text{self}} = 2.25$. The upstream and downstream decomposition of M^{tot} yields:

$$M_{\text{up}} = 2.0 \text{ (SE 0.39)}, \qquad M_{\text{down}} = 1.3 \text{ (SE 0.21)}$$

so, most ($70\% = \frac{\ln M_{\text{up}}}{\ln M^{\text{tot}}}$) of the TFP influence is upstream, i.e. from customers to suppliers ($M_{\text{up}} > M_{\text{down}}$).[22] Empirically, we find that the Hulten multiplier (37) is:

$$M^{\text{Hulten}} = 1.8$$

So, the total multiplier is

$$M^{\text{Full}} = M^{\text{Hulten}} M^{\text{tot}} = 1.8 \times 2.7 = 4.85 \simeq 5 \text{ (SE 1.5)}.$$

Hence, if primitively, industries experience individually a 1% TFP rise, then after propagation through the production network, the TFP of the average industry (on gross output) rises by 2.7%, and the aggregate economy-wide TFP (on net output) rises by 5% (the latter effect being due to the traditional Hulten effect). This five-fold amplification of TFP shocks is a long-run effect.

This is admittedly a surprisingly large effect. One part is "mechanical", the Hulten multiplier of 1.8, and it works under all frictionless models. Surprisingly, this is the first time one sees it, but it is not surprising to trained macroeconomists.The new part is the total multiplier of $M^{\text{tot}} = 2.7$. The "learning from one's neighbors" effect has long been qualitatively found (at least since Marshall 1920). Here, we have a quantification from arguably plausible shocks. It complements other estimates showing that the social return to R&D is much higher than the private return (e.g. Griliches 1992; Hall, Mairesse, and Mohnen 2010; Bloom, Schankerman, and Reenen 2013; Jones and Summers 2022).

The order of magnitude is reasonable. For instance, Bloom, Schankerman, and Reenen 2013 find that the public returns to R&D are about 2.6 times larger than the private returns, using as instruments changes in federal and state tax incentives for R&D.

The advantage of the method here is that it can fairly easily encompass the whole economy.

One could refine the analysis in a number of interesting ways: e.g. industries that are younger or with younger firms might get more easily influenced. Physically closer industries influence each other more.[23]Also, one could microfound the slow dynamics in a more refined

22. The difference is significant at the 5% level: $\ln \frac{M_{\text{up}}}{M_{\text{down}}} = 0.41$ (SE 0.17).

23. That could be modeled respectively as $B_{ij}(\gamma) = (\gamma' X_i)\,\Omega_{ij}$, with X_i a vector of characteristics such as

way, perhaps using the slow adjustment of prices as in Minton and Wheaton (2023); we do not do that, because there are clearly other temporal frictions, so an agnostic method seems a good first pass. We do wish to note a limitation of our data: if within an industry, a firm's TFP increase influences its competitors, this is not fully detected by our data, as we have only industry-level data, rather than data on firms within an industry. We do observe some own-industry spillovers (via the $M^{\text{self}} = 1.2$). But to fully complete the firms-to-other firms in the same industry spillovers, we'd need at least some firm data. This is all doable in principle,[24] but we leave this extension to other future studies.

3.4.3 Transition and impulse responses

Those multipliers show the long run propagation of TFP shocks. We next delineate how fast they propagate, with the impulse responses in Figure 1.[25] The propagation is initially very fast (indeed, over half the effect happens in the same year), and then is slower (the effect reaches 90% of its long term value after 4 years).

3.4.4 Validation

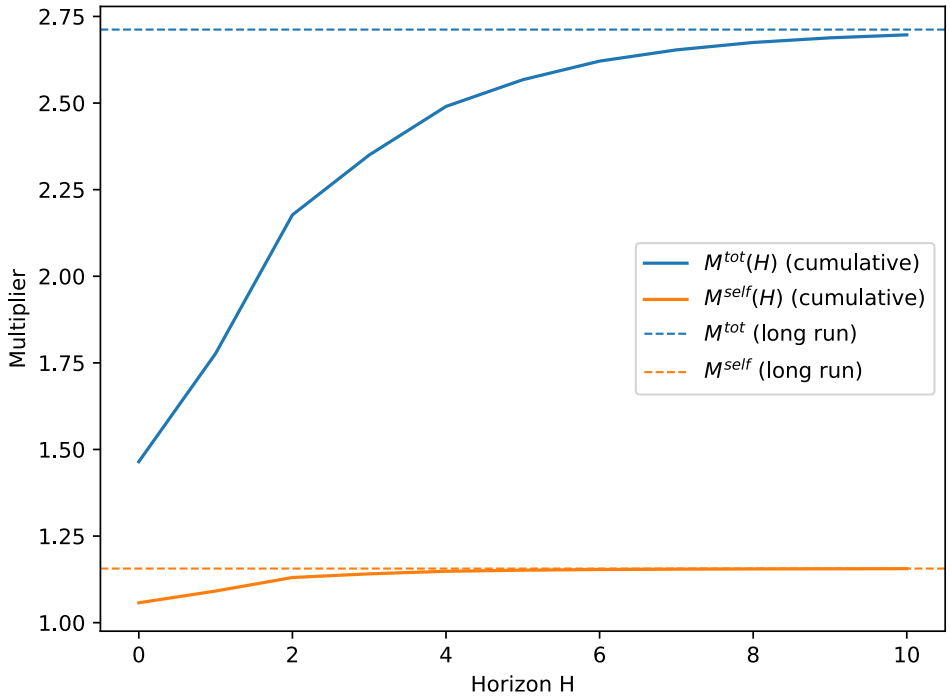
The online appendix (Section E) contains complements, including simulations, and further diagnostics. We highlight two results here. Figure 2 shows the scatter plot of fitted values using only the idiosyncratic shocks in other sectors and TFP in sector i. The linear fit is close to one, confirming the algorithm converges appropriately. Figure 3 shows a negative relationship between output price inflation and the fitted values using only the idiosyncratic shocks. This relationship is not mechanical. In fact, a concern that upstream propagation of TFP reflects higher factor utilization given higher customer demand would predict a positive price response. A negative price response is instead consistent with TFP spillovers reflecting genuine technology improvements, which exert deflationary pressure at the industry level.

age; and $B_{ij}(\gamma) = \gamma P_{ij} \Omega_{ij}$ where P_{ij} is an index of proximity.

24. Having data on some large firms in one of a few industries would in principle be enough, provided we assume that the within-industry multiplier is the same across industries.

25. Appendix E explains the construction.

Figure 1: TFP propagation: cumulative impulse response



This figure shows the cumulative TFP impulse response multipliers over horizon H in the two-lag model. $M^{\text{tot}}(H)$ is the aggregate multiplier and $M^{\text{self}}(H)$ is the own-sector multiplier. The long-run multipliers M^{tot} and M^{self} are also shown. The time unit is the year.

4 Conclusion

We proposed a method to identify causal parameters in fairly general economies that have a network structure: the network granular instrumental variable, or Network GIV. The general idea, and method, could be applied to a host of other setups, for instance for the world economy, aggregate financial markets with many assets, and more localized industries with imperfect competition. In all those cases, idiosyncratic shocks in one sector imply reactions of all the other sectors, in a way specified by the theory; if that theory is low-dimensional enough, estimation of the parameters is typically possible. In addition, there is generally a clear intuition for the way each parameter is estimated. As we saw, the procedure admits many variants that are empirically relevant, including autocorrelated shocks, lead-lag structures, measurement error, or even missing data. We hope that the Network GIV method will help economists explore a large number of causal relations in old and novel datasets.

References

Acemoglu, Daron. 2009. *Introduction to modern economic growth.* Princeton University Press, January.

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The network origins of aggregate fluctuations." *Econometrica* 80 (5): 1977–2016.

Adrian, Tobias, Federico Grinberg, Nellie Liang, Sheheryar Malik, and Jie Yu. 2022. "The term structure of growth-at-risk." *American Economic Journal: Macroeconomics* 14 (3): 283–323.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2): 323–351.

Andersen, Asger Lau, Kilian Huber, Niels Johannesen, Ludwig Straub, and Emil Toft Vestergaard. 2026. "Disaggregated Economic Accounts." *Quarterly Journal of Economics* 141 (2): 1005–1075.

Andrews, Isaiah, Nano Barahona, Matthew Gentzkow, Ashesh Rambachan, and Jesse M. Shapiro. 2025. "Structural Estimation Under Misspecification: Theory and Implications for Practice." *The Quarterly Journal of Economics* 140 (3): 1801–1855. https://doi.org/10.1093/qje/qjaf018.

Atalay, Enghin. 2017. "How important are sectoral shocks?" *American Economic Journal: Macroeconomics* 9 (4): 254–80.

Bai, Jushan. 2003. "Inferential theory for factor models of large dimensions." *Econometrica* 71 (1): 135–171.

Baqaee, David Rezza, and Emmanuel Farhi. 2019. "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem." *Econometrica* 87 (4): 1155–1203.

Baqaee, David Rezza, and Emmanuel Farhi. 2020. "Productivity and misallocation in general equilibrium." *The Quarterly Journal of Economics* 135 (1): 105–163.

Battaglini, Marco, Eleonora Patacchini, and Edoardo Rainone. 2022. "Endogenous social interactions with unobserved networks." *The Review of Economic Studies* 89 (4): 1694–1747.

Baumeister, Christiane, and James D Hamilton. 2023. *A Full-Information Approach to Granular Instrumental Variables.* Technical report.

———. 2024. "Advances in using vector autoregressions to estimate structural magnitudes." *Econometric Theory* 40 (3): 472–510.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81 (4): 1347–1393.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Yannis M Ioannides. 2011. "Identification of social interactions." In *Handbook of social economics,* 1:853–964. Elsevier.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Rajshri Jayaraman. 2015. "Linear social interactions models." *Journal of Political Economy* 123 (2): 444–496.

Borusyak, Kirill, and Peter Hull. 2023. "Nonrandom Exposure to Exogenous Shocks." *Econometrica* 91 (6): 2155–2185.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics* 150 (1): 41–55.

Breza, Emily, Arun G Chandrasekhar, Tyler H McCormick, and Mengjie Pan. 2020. "Using aggregated relational data to feasibly identify network structure without network data." *American Economic Review* 110 (8): 2454–2484.

Brownlees, Christian, and Geert Mesters. 2020. "Detecting granular time series in large panels." *Journal of Econometrics.*

Bureau of Economic Analysis. 2023. *Industry Economic Accounts.*

Camanho, Nelson, Harald Hau, and Hélene Rey. 2022. "Global portfolio rebalancing and exchange rates." *The Review of Financial Studies* 35 (11): 5228–5274.

Carvalho, Vasco M, and Basile Grassi. 2019. "Large firm dynamics and the business cycle." *American Economic Review* 109 (4): 1375–1425.

Chamberlain, Gary. 1987. "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of econometrics* 34 (3): 305–334.

Chaudhary, Manav, Julie Zhiyu Fu, and Haonan Zhou. 2025. "Anatomy of the Treasury Market: Who Moves Yields?" *Available at SSRN 5021055.*

Chodorow-Reich, Gabriel. 2019. "Geographic cross-sectional fiscal spending multipliers: What have we learned?" *American Economic Journal: Economic Policy* 11 (2): 1–34.

Chodorow-Reich, Gabriel, Andra Ghent, and Valentin Haddad. 2021. "Asset insulators." *The Review of Financial Studies* 34 (3): 1509–1539.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of social interactions through partially overlapping peer groups." *American Economic Journal: Applied Economics* 2 (2): 241–275.

De Graeve, Ferre, and Jan David Schneider. 2023. "Identifying sectoral shocks and their role in business cycles." *Journal of Monetary Economics* 140:124–141.

De Paula, Áureo, Imran Rasul, and Pedro Souza. 2019. "Identifying network ties from panel data: Theory and an application to tax competition." *arXiv preprint arXiv:1910.07452.*

Gabaix, Xavier. 2011. "The granular origins of aggregate fluctuations." *Econometrica* 79 (3): 733–772.

Gabaix, Xavier, and Ralph SJ Koijen. 2023. "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis." *Available at SSRN 3686935.*

Gabaix, Xavier, and Ralph SJ Koijen. 2024. "Granular instrumental variables." *Journal of Political Economy* 132 (7): 2274–2303.

Galaasen, Sigurd, Rustam Jamilov, Ragnar Juelsrud, Hélène Rey, et al. 2023. *Granular credit risk.* Technical report. NBER Working Paper No. 27994.

Gaubert, Cecile, and Oleg Itskhoki. 2021. "Granular comparative advantage." *Journal of Political Economy* 129 (3): 871–939.

Giovanni, Julian di, and Andrei A Levchenko. 2012. "Country size, international trade, and aggregate fluctuations in granular economies." *Journal of Political Economy* 120 (6): 1083–1132.

Giovanni, Julian di, Andrei A Levchenko, and Isabelle Méjean. 2014. "Firms, destinations, and aggregate fluctuations." *Econometrica* 82 (4): 1303–1340.

Griffith, Alan, and Sida Peng. 2023. "Identification of Network Structure in the Presence of Latent, Unobserved Factors: A New Result using Turan's Theorem." *Unpublished manuscript.*

Griliches, Zvi. 1992. "The Search for R&D Spillovers." *The Scandinavian Journal of Economics,* S29–S47.

Hall, Bronwyn H., Jacques Mairesse, and Pierre Mohnen. 2010. "Chapter 24 - Measuring the Returns to R&D." In *Handbook of the Economics of Innovation, Volume 2,* 2:1033–1082. Handbook of the Economics of Innovation. North-Holland.

Huber, Kilian. 2023. "Estimating general equilibrium spillovers of large-scale shocks." *The Review of Financial Studies* 36, no. 4 (August): 1548–1584.

Hulten, Charles R. 1978. "Growth accounting with intermediate inputs." *The Review of Economic Studies* 45 (3): 511–518.

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *The American Economic Review* 76 (5): 984–1001.

Jones, Benjamin F., and Lawrence H. Summers. 2022. "A Calculation of the Social Returns to Innovation." In *Innovation and Public Policy.* University of Chicago Press, March.

Jones, Charles I. 1995. "R & D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–784.

Kline, Brendan, and Elie Tamer. 2020. "Econometric analysis of models with social interactions." In *The Econometric Analysis of Network Data,* 149–181. Elsevier.

Koijen, Ralph SJ, and Motohiro Yogo. 2019. "A demand system approach to asset pricing." *Journal of Political Economy* 127 (4): 1475–1515.

Lee, Lung-Fei. 2007. "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects." *Journal of econometrics* 140 (2): 333–374.

Liu, Ernest, and Song Ma. 2021. *Innovation Networks and R&D Allocation.* Working Paper, Working Paper Series 29607. National Bureau of Economic Research, December. https://doi.org/10.3386/w29607.

Long, John B, and Charles I Plosser. 1983. "Real business cycles." *Journal of Political Economy* 91 (1): 39–69.

Manresa, Elena. 2013. "Estimating the structure of social interactions using panel data." *Unpublished Manuscript. CEMFI, Madrid* 1.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies* 60 (3): 531–542.

Marshall, Alfred. 1920. *Principles of Economics.* 8th ed. London: Macmillan.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103 (4): 1212–47. https://doi.org/10.1257/aer.103.4.1212. http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.4.1212.

Minton, Robert, and Brian Wheaton. 2023. "Delayed Inflation in Supply Chains: Theory and Evidence." *Available at SSRN 4470302.*

Nakamura, Emi, and Jon Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32, no. 3 (August): 59–86.

Newey, Whitney K, and Daniel McFadden. 1994. "Large sample estimation and hypothesis testing." *Handbook of econometrics* 4:2111–2245.

Neyman, Jerzy, and Elizabeth L. Scott. 1948. "Consistent estimates based on partially consistent observations." *Econometrica* 16 (1): 1–32.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2022. "Instrumental Variable Identification of Dynamic Variance Decompositions." *Journal of Political Economy* 130 (8): 2164–2202. https://doi.org/10.1086/720141. eprint: https://doi.org/10.1086/720141. https://doi.org/10.1086/720141.

Qian, Eric. 2024. *Heterogeneity-robust granular instruments.* arXiv: 2304.01273 [econ.EM]. https://arxiv.org/abs/2304.01273.

Ramey, Valerie A. 2016. "Macroeconomic shocks and their propagation." In *Handbook of Macroeconomics,* 2:71–162. Elsevier.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1002–1037.

Rubio-Ramírez, Juan F., Daniel F. Waggoner, and Tao Zha. 2010. "Structural vector autoregressions: theory of identification and algorithms for inference." *The Review of Economic Studies* 77 (2): 665–696.

Shi, Wei, and Lung-fei Lee. 2017. "Spatial dynamic panel data models with interactive fixed effects." *Journal of Econometrics* 197 (2): 323–347. https://doi.org/10.1016/j.jeconom.2016.12.001.

Zhang, Anru R, T Tony Cai, and Yihong Wu. 2022. "Heteroskedastic PCA: Algorithm, optimality, and applications." *The Annals of Statistics* 50 (1): 53–80.

A Estimation with Missing Outcomes and Measurement Error

This section extends the Network GIV procedure of Algorithm 1 to the case in which outcomes are observed with missing cells and, possibly, with measurement error. We propose two algorithms: a very simple heuristic one, a more sophisticated (that explicitly models measurement error and missing outcomes through a Bayesian model). Section (F.1) shows some Monte Carlo simulations, illustrating for that simulation at least, the simple Algorithm 2 and the sophisticated Algorithm 3 have quite similar performance. Hence, we recommend using the simple Algorithm 2 first, and then (e.g. to refine the estimates) the more sophisticated Algorithm 3.

A.1 Setup

Consider the model

$$Y_t^\star = \sum_{\ell=0}^{L} B_\ell(\gamma^\star) Y_{t-\ell}^\star + \Lambda \eta_t + u_t, \tag{A.1}$$

$$Y_t = S_t(Y_t^\star + \varepsilon_t), \qquad S_t = \mathrm{diag}(s_{1t}, \ldots, s_{Nt}), s_{it} \in \{0,1\} \tag{A.2}$$

for $t = 1, \ldots, T$, with observed data Y_t; $s_{it} \in \{0,1\}$ indicates which unit-period cells are observed, and ε_{it} is measurement error noise. We condition throughout on the realized masks $S_1, \ldots, S_T$. The maintained exogeneity condition is that, conditional on the masks and the model covariates, the current structural shocks and measurement errors satisfy $\mathbb{E}[u_t] = \mathbb{E}[\varepsilon_t] = 0$ and are serially independent. The structural shocks u_t are independent across units; the measurement errors ε_t have cross-sectional covariance $\mathrm{Cov}(\varepsilon_t) = D_\omega^{1/2} R(\rho) D_\omega^{1/2}$, where $R(\rho)$ is a correlation matrix indexed by a scalar ρ (default: equicorrelation, $R(\rho) = (1-\rho)I + \rho \mathbf{1}\mathbf{1}^\top$, $\rho \in (-1/(N-1), 1)$). We use the notations:

$$M(\gamma) \coloneqq I - B_0(\gamma), \qquad A(\gamma) \coloneqq M(\gamma)^{-1} \tag{A.3}$$

Algorithm 2 First-pass estimator of model (A.1) and (A.2) with missing outcomes.

Require: Observed panel $Y^{\text{obs}} \in (\mathbb{R} \cup \{\text{Not Available}\})^{N \times T}$, initial guess $\hat{\gamma} \in \mathbb{R}^p$, number of factors r.

1: Define the observation weights $s_{it} \coloneqq \mathbf{1}\{Y_{it}^{\text{obs}} \text{ is observed}\}$. At the first iteration, initialize the completed panel by

$$Y_{it} \coloneqq \begin{cases} Y_{it}^{\text{obs}} & \text{if } s_{it} = 1, \\ 0 & \text{if } s_{it} = 0. \end{cases}$$

In the imputed-missing-value variant of the algorithm, after the first iteration set $Y_{it} \coloneqq e_i^\top \tilde{y}_{t|-i}$ whenever $s_{it} = 0$.

2: Define $\tilde{v}_t \coloneqq Y_t - \sum_{\ell=0}^{L} B_\ell(\hat{\gamma}) Y_{t-\ell}$.

3: Apply homoskedastic or heteroskedastic PCA to $\tilde{v}_t$ using the weights s_{it}, so that missing outcomes receive zero weight. Denote the estimated loadings by $\tilde{\lambda}$ and the estimated factors by $\hat{\eta}_t$.

4: For each i, estimate $\hat{\eta}_t^{-i}$ as the slope of the weighted regression $\tilde{v}_{jt} = \tilde{\lambda}_j \eta_t^{-i} + \tilde{u}_{jt}^{-i}$ using only units $j \neq i$, with weights s_{jt}. Collect the idiosyncratic residuals

$$\tilde{u}_{jt}^{-i} \coloneqq \begin{cases} \tilde{v}_{jt} - \tilde{\lambda}_j^\top \hat{\eta}_t^{-i} & \text{if } j \neq i \text{ and } s_{jt} = 1, \\ 0 & \text{if } j = i \text{ or } s_{jt} = 0. \end{cases}$$

5: For each i, set $\tilde{y}_{0|-i} = 0$ and construct recursively

$$\tilde{y}_{t|-i} \coloneqq \big(I - B_0(\hat{\gamma})\big)^{-1} \Big(\tilde{u}_t^{-i} + \sum_{\ell=1}^{L} B_\ell(\hat{\gamma}) \tilde{y}_{t-\ell|-i} \Big).$$

6: For each i, t, construct the instrument $z_{i,t} \in \mathbb{R}^p$ with k-th component

$$z_{it}^{(k)} \coloneqq e_i^\top \sum_{\ell=0}^{L} \frac{\partial B_\ell}{\partial \gamma_k} \tilde{y}_{t-\ell|-i}.$$

7: Estimate the updated $\hat{\gamma}'$ in two steps.

1. For all i, regress $\hat{\eta}_t^{-i}$ on z_{it} with weights s_{it} and collect the residuals $\hat{\eta}_t^{-i,\perp}$: $\hat{\eta}_t^{-i} = \beta_i z_{it} + \hat{\eta}_t^{-i,\perp}$.
2. Run the WLS regression

$$\tilde{v}_{it} = \Delta\gamma \cdot z_{i,t} + b_i \hat{\eta}_t^{-i,\perp} + e_{it} \quad \text{with weights } s_{it},$$

and update $\hat{\gamma}' \coloneqq \hat{\gamma} + \texttt{step}\, \Delta\gamma$ with $\texttt{step} = 1$ (or e.g. 0.3 initially)

return $\hat{\gamma}'$.

A.2 A first-pass algorithm for missing data

It is useful to consider a simple procedure that handles missing outcomes directly, without the latent-panel posterior step. The procedure described here is a first-pass version of the algorithm of Section A: it handles missing outcomes only through observation weights and a simple imputation step, without an explicit treatment of the measurement error in the imputed entries. It is not theoretically exact—in particular, the imputed values are treated as known inputs to the dynamic regressors, which can create a self-referential feedback loop in which larger estimated propagation coefficients lead to larger imputations, which in turn push the next update further upward. Nonetheless, it is much simpler and perhaps more intuitive to implement than the corrected procedure of Algorithm3, it requires no Monte Carlo step, and in our simulations it often delivers reasonable estimates of γ and of the implied multipliers, so we record it here as a practical alternative.

Let $s_{it} \coloneqq \mathbf{1}\{Y_{it}^{\text{obs}} \text{ is observed}\}$ denote the observation indicator, and treat missing outcomes through the weights s_{it} in the PCA step and in the regressions used to update γ; missing outcomes receive zero weight throughout. The completed panel is initialized by zero-filling the missing entries at the first iteration and, in the imputed-missing-value variant, by replacing each missing entry by its leave-one-out predicted value $e_i^\top \tilde{y}_{t|-i}$ in subsequent iterations. The PCA step is a weighted variant of the PCA used in the baseline: the loadings λ and factors η_t solve

$$\min_{\lambda \in \mathbb{R}^{N\times r},\, \eta \in \mathbb{R}^{r\times T}} \sum_{i,t} s_{it}\big(\tilde{v}_{it} - \lambda_i^\top \eta_t\big)^2,$$

which can be implemented as an alternating-least-squares iteration in which unobserved entries are simply excluded. The full procedure is summarized in Algorithm 2.

A.3 A sophisticated algorithm for missing and/or mismeasured data

We next show a more sophisticated, and rigorous, algorithm.

Algorithm 3 Projected-instrument estimator of model (A.1) and (A.2) with missing outcomes and/or measurement error.

Require: Observed panel $\{Y_t, S_t\}_{t=1}^T$, initial value $\hat{\gamma}^{(0)}$, lag length L, factor rank r, tolerances

$\varepsilon_{\mathrm{EM}}, \varepsilon_\gamma > 0$, maximum numbers of iterations M_{EM} and $R_\gamma^{\max}$.

Require: Inverse-gamma (IG) priors with density proportional to $x^{-(a+1)}\exp(-b/x)$.

$$\sigma_{u_i}^2 \equiv \sigma_i^2 \sim IG(a_\sigma, b_\sigma), \qquad \sigma_{\varepsilon_i}^2 \equiv \omega_i^2 \sim IG(a_\omega, b_\omega), \qquad i = 1, \ldots, N,$$

and a Beta prior on the measurement-error correlation ρ: with $\tau = (\rho - \rho_{\min})/(\rho_{\max} - \rho_{\min})$,$\rho_{\min} = -1/(N-1)$,$\rho_{\max} = 1$,$\tau \sim \mathrm{Beta}(\alpha_\rho, \beta_\rho)$ with default $\alpha_\rho = \beta_\rho = 3$.

Require: Measurement-error covariance: $\mathrm{Cov}(\varepsilon_t) = D_\omega^{1/2} R(\rho) D_\omega^{1/2}$, with $D_\omega = \mathrm{diag}(\omega_1^2, \ldots, \omega_N^2)$ the diagonal of marginal measurement-error variances and $R(\rho)$ a correlation matrix (default: equicorrelation, $R(\rho) = (1-\rho)I + \rho \mathbf{1}\mathbf{1}^\top$).

Initialize $\hat\gamma^{(0)}$,$(\Lambda^{(0)}, \{\eta_t^{(0)}\}_{t=1}^T)$, and the variance components using the preliminary values described below:

$$\hat\sigma_i^{2,(0)} = \tilde\sigma_i^2, \qquad \hat\omega_i^{2,(0)} = \tilde\omega_i^2, \qquad D_\sigma^{(0)} = \mathrm{diag}(\hat\sigma_i^{2,(0)}), \qquad D_\omega^{(0)} = \mathrm{diag}(\hat\omega_i^{2,(0)}).$$

for $r_\gamma = 0, 1, \ldots, R_\gamma^{\max} - 1$ **do**

  Set $\hat\gamma = \hat\gamma^{(r_\gamma)}$.

  for $m = 0, 1, \ldots, M_{\mathrm{EM}} - 1$ **do**

    E-step. Given $\hat\gamma, (\Lambda^{(m)}, \{\eta_t^{(m)}\}), D_\sigma^{(m)}$ and $D_\omega^{(m)}$, compute the full posterior moments

$$\hat Y_t^{\star,(m+1)} = \mathbb{E}_m[Y_t^\star \mid Y_{1:T}], \qquad \hat u_t^{(m+1)} = \mathbb{E}_m[u_t \mid Y_{1:T}],$$

and the posterior second moments $\mathbb{E}_m[u_{it}^2 \mid Y_{1:T}]$,$\mathbb{E}_m[\varepsilon_{it}^2 \mid Y_{1:T}]$ for observed entries $s_{it} = 1$, and the cross-pair posterior second moments $\mathbb{E}_m[\varepsilon_{it}\varepsilon_{jt} \mid Y_{1:T}]$ for observed pairs $(i,j) \in \mathcal{O}_t \times \mathcal{O}_t$, $i \neq j$.

    Define $n_i \coloneqq \sum_{t=1}^T s_{it}$ and update the variance components by the MAP steps

$$\hat\sigma_i^{2,(m+1)} \coloneqq \frac{b_\sigma + \frac{1}{2}\sum_{t=1}^T \mathbb{E}_m[u_{it}^2 \mid Y_{1:T}]}{a_\sigma + 1 + \frac{T}{2}}, \qquad \hat\omega_i^{2,(m+1)} \coloneqq \frac{b_\omega + \frac{1}{2}\sum_{t=1}^T s_{it}\mathbb{E}_m[\varepsilon_{it}^2 \mid Y_{1:T}]}{a_\omega + 1 + \frac{n_i}{2}}.$$

Set $D_\sigma^{(m+1)} \coloneqq \mathrm{diag}(\hat\sigma_i^{2,(m+1)})$, $D_\omega^{(m+1)} \coloneqq \mathrm{diag}(\hat\omega_i^{2,(m+1)})$.

    Update the measurement-error correlation ρ by the plug-in moment formula:

$$\bar r^{(m+1)} = \frac{1}{|\mathcal{P}|}\sum_{(i,j,t)\in\mathcal{P}} \frac{\mathbb{E}_m[\varepsilon_{it}\varepsilon_{jt} \mid Y_{1:T}]}{\hat\omega_i^{(m+1)}\hat\omega_j^{(m+1)}}, \qquad \hat\rho^{(m+1)} = \Pi_{[\rho_{\min}+\delta, \rho_{\max}-\delta]}\left(\frac{|\mathcal{P}|\bar r^{(m+1)} + \kappa_\rho\rho_0}{|\mathcal{P}| + \kappa_\rho}\right),$$

with $\mathcal{P} = \{(i,j,t) : i < j,\ s_{it} = s_{jt} = 1\}$,$\rho_0$ the prior mean (default $\hat\rho^{(0)}$), $\kappa_\rho = \alpha_\rho + \beta_\rho - 2$, and Π the projection onto the interval (default $\delta = 10^{-3}$). Set $\Omega^{(m+1)} = (D_\omega^{(m+1)})^{1/2} R(\hat\rho^{(m+1)}) (D_\omega^{(m+1)})^{1/2}$.

    Form the posterior structural residual $\widehat v_t^{(m+1)} = \widehat Y_t^{\star,(m+1)} - \sum_{\ell=0}^L B_\ell(\widehat\gamma)\widehat Y_{t-\ell}^{\star,(m+1)}$.

    Update the factor component by applying the same factor step as in algorithm 1 to $\{\widehat v_t^{(m+1)}\}_{t=1}^T$, with the same normalization as in the baseline factor step:

$$(\Lambda^{(m+1)}, \{\eta_t^{(m+1)}\}_{t=1}^T) \in \arg\min_{\Lambda, \{\eta_t\}} \sum_{t=1}^T \left\| (D_\sigma^{(m+1)})^{-1/2}\left(\widehat v_t^{(m+1)} - \Lambda\eta_t\right)\right\|_2^2.$$

  end for

  Set $\widehat Y_t^\star \coloneqq \widehat Y_t^{\star,(m+1)}$, $\widehat v_t \coloneqq \widehat v_t^{(m+1)}$, $\xi_t \coloneqq \widehat u_t^{(m+1)}$, $\widehat\Lambda \coloneqq \Lambda^{(m+1)}$, $\widehat\eta_t \coloneqq \eta_t^{(m+1)}$.

  for each unit $i = 1, \ldots, N$ **do**

Estimate the leave-one-out factor path $\hat{\eta}_t^{-i}$ from $\{\hat{v}_{jt}\}_{j\neq i}$ and define the leave-one-out shock path $\hat{u}^0_{t|-i} \coloneqq (I - e_i e_i^\top)\big(\hat{v}_t - \hat{\Lambda}\hat{\eta}_t^{-i}\big)$.

Starting from $\tilde{y}^0_{t|-i} = 0$ for $t \le 0$, compute successively:

$$\tilde{y}^0_{t|-i} \coloneqq A(\hat{\gamma})\big(\hat{u}^0_{t|-i} + \sum_{\ell=1}^{L} B_\ell(\hat{\gamma})\tilde{y}^0_{t-\ell|-i}\big), \qquad t = 1, \dots, T. \tag{A.4}$$

Construct the raw derivative instrument $z^0_{it} \in \mathbb{R}^p$ with components

$$z^{0,(k)}_{it} \coloneqq e_i^\top \sum_{\ell=0}^{L} \frac{\partial B_\ell(\hat{\gamma})}{\partial \gamma_k}\tilde{y}^0_{t-\ell|-i}, \qquad k = 1, \dots, p.$$

Compute the pooled projection coefficient from the fitted Gaussian model as described below: $\pi_i \coloneqq \frac{\sum_{t=1}^{T} \mathbb{E}_{\hat{\gamma}}[z^0_{it}\xi_{it}|S_{1:T}]}{\sum_{t=1}^{T} \mathbb{E}_{\hat{\gamma}}[\xi^2_{it}|S_{1:T}]}$.

Form the projected instrument $z^\perp_{it} \coloneqq z^0_{it} - \pi_i \xi_{it}$, $t = 1, \dots, T$.

Regress $\hat{\eta}_t^{-i}$ on $z^\perp_{it}$ to obtain the pooled residual $\hat{\eta}_t^{-i,\perp}$.

end for

Estimate by least squares the coefficients $\Delta\gamma$ and the nuisance loadings b_i in:

$$\hat{v}_{it} = \Delta\gamma \cdot z^\perp_{it} + b_i \hat{\eta}_t^{-i,\perp} + e_{it} \tag{A.5}$$

Update $\hat{\gamma}^{(r_\gamma+1)} \coloneqq \hat{\gamma}^{(r_\gamma)} + \texttt{step}\,\Delta\gamma$ with $\texttt{step} = 1$ (or e.g. 0.3 initially)

end for

return final $\hat{\gamma}$, factor estimates, variance estimates, and posterior shock proxies.

A.3.1 Notation and setup

For the posterior calculations, we use the Gaussian working model $u_t \sim N(0, D_\sigma)$,$\varepsilon_t \sim N\big(0, D_\omega^{1/2} R(\rho) D_\omega^{1/2}\big)$ with diagonal $D_\sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_N^2)$ and $D_\omega = \mathrm{diag}(\omega_1^2, \dots, \omega_N^2)$ the diagonal of marginal measurement-error variances, and $R(\rho)$ a correlation matrix indexed by a scalar ρ (default: equicorrelation $R(\rho) = (1-\rho)I + \rho\mathbf{1}\mathbf{1}^\top$), and for a candidate γ we let

$$\xi_{it}(\gamma) \coloneqq \hat{u}_{it}(\gamma) = \mathbb{E}_\gamma\big[u_{it} \mid Y_{1:T}, S_{1:T}\big] \tag{A.6}$$

denote the full posterior mean of the latent shock.

In the complete-data baseline of Algorithm 1 the local update of γ is identified from the structural residual $\tilde{v}_t(\gamma) = Y_t - \sum_{\ell=0}^{L} B_\ell(\gamma) Y_{t-\ell}$. A missing-outcomes and measurement error extension which would simply replace $\tilde{v}_{it}(\gamma)$ by the posterior shock proxy $\xi_{it}(\gamma)$ would work only in a model without factors. Indeed, with estimated common factors, $\xi_{it}(\gamma) = \hat{v}_{it}(\gamma) - \hat{\lambda}_i(\gamma)^\top \hat{\eta}_t(\gamma)$, so the linearization of $\xi_{it}(\gamma)$ around the current $\hat{\gamma}$ also contains the

derivative of the fitted common component; a projected- ξ update would drop this term and therefore no longer match the baseline Network GIV moment. Therefore, we use the posterior step to recover the latent de-noised panel $\hat{Y}_t^\star = \mathbb{E}\big[Y_t^\star \mid Y_{1:T}, S_{1:T}\big]$, and then form the posterior structural residual

$$\hat{v}_t \coloneqq \hat{Y}_t^\star - \sum_{\ell=0}^{L} B_\ell(\hat{\gamma}) \hat{Y}_{t-\ell}^\star, \tag{A.7}$$

which plays the same role on the left-hand side of the local moment as $\tilde{v}_t(\hat{\gamma})$ in the baseline.

The leave-one-out construction mirrors that of Algorithm 1, but evaluated at the posterior-recovered latent panel. For each target unit i, estimate the leave-one-out factor path $\widehat{\eta}_t^{-i}$ from $\{\widehat{v}_{jt}\}_{j\neq i}$, using the fitted loadings $\widehat{\Lambda}$ and the same leave-one-out factor-score step as in Algorithm 1. Define the leave-one-out idiosyncratic shock path

$$\hat{u}^0_{t|-i} \coloneqq (I - e_i e_i^\top)\big(\hat{v}_t - \hat{\Lambda}\hat{\eta}_t^{-i}\big),$$

where e_i is the i-th canonical basis vector in $\mathbb{R}^N$. Starting from $\tilde{y}^0_{t|-i}(\gamma) = 0$ for $t \leq 0$, define

$$\tilde{y}^0_{t|-i}(\gamma) = A(\gamma)\big(\hat{u}^0_{t|-i} + \sum_{\ell=1}^{L} B_\ell(\gamma)\tilde{y}^0_{t-\ell|-i}(\gamma)\big).$$

For $k = 1, \ldots, p$, define the raw derivative instrument

$$z^{0,(k)}_{it}(\gamma) = e_i^\top \sum_{\ell=0}^{L} \frac{\partial B_\ell(\gamma)}{\partial \gamma_k} \tilde{y}^0_{t-\ell|-i}. \tag{A.8}$$

Stacking over k, write $z^0_{it}(\gamma) = (z^{0,(1)}_{it}(\gamma), \ldots, z^{0,(p)}_{it}(\gamma))^\top$. The raw instrument $z^0_{it}(\gamma)$ is designed to preserve the same identifying variation as in the complete-data Network GIV procedure. However, because it is computed from posterior quantities derived from noisy and incomplete data, $z^0_{it}(\gamma^\star)$ may be correlated with $\xi_{it}(\gamma^\star)$. Therefore the raw moment

$$\mathbb{E}[z^0_{it}(\gamma^\star)\xi_{it}(\gamma^\star)] \neq 0.$$

Fix a current value $\hat{\gamma}$ and suppress dependence on $\hat{\gamma}$ in this paragraph. Let

$$\xi_{it} \coloneqq \xi_{it}(\hat{\gamma}), \qquad z^0_{it} \coloneqq z^0_{it}(\hat{\gamma}),$$

where ξ_{it} is the posterior shock proxy and $z^0_{it} \in \mathbb{R}^p$ is the raw Network GIV derivative instrument. Because missingness and measurement error may induce correlation between the raw instrument z^0_{it} and the shock proxy ξ_{it}, we remove from z^0_{it} the component that is linearly related to ξ_{it}. We do this using a coefficient pooled over time within unit. Define[26]

$$\pi_i \coloneqq \frac{\sum_{t=1}^{T} \mathbb{E}_{\hat{\gamma}}\left[z^0_{it}\xi_{it} \mid S_{1:T}\right]}{\sum_{t=1}^{T} \mathbb{E}_{\hat{\gamma}}\left[\xi^2_{it} \mid S_{1:T}\right]} \in \mathbb{R}^p, \tag{A.9}$$

and set

$$z^{\perp}_{it} \coloneqq z^0_{it} - \pi_i \xi_{it}. \tag{A.10}$$

By construction, the pooled projected instrument satisfies

$$\sum_t \mathbb{E}_{\hat{\gamma}}\left[z^{\perp}_{it}\xi_{it} \mid S_{1:T}\right] = 0.$$

Thus the correction removes only the time-pooled component of the raw instrument that violates the moment condition. It does not project the observed data and therefore does not discard entire directions that may be useful for identifying γ.

For each i, regress $\hat{\eta}_t^{-i}$ on $z^{\perp}_{it}$ and collect the pooled residual $\hat{\eta}_t^{-i,\perp}$. The projected update solves the aggregate moment equation

$$\sum_{i,t} z^{\perp}_{it}\left[\widehat{v}_{it} - (z^{\perp}_{it})^{\top}\Delta\gamma - b_i^{\top}\widehat{\eta}_t^{-i,\perp}\right] = 0, \tag{A.11}$$

solved jointly for $\Delta\gamma \in \mathbb{R}^p$ and the nuisance loadings $b_i \in \mathbb{R}^r$. Operationally, (A.11) is the linear regression

$$\hat{v}_{it} = \Delta\gamma \cdot z^{\perp}_{it} + b_i\hat{\eta}_t^{-i,\perp} + \varepsilon_{it},$$

instrumented by $z^{\perp}_{it}$, from which we estimate the slope vector $\Delta\gamma$ and update $\hat{\gamma}' \coloneqq \hat{\gamma} + \Delta\gamma$.

26. The denominator in (A.9) is assumed to be positive; if it is close to zero, the unit contributes little information about the moment and the projection coefficient can be regularized.

The full estimator is summarized in Algorithm 3.

A.3.2 Practical initialization and empirical-Bayes priors

For implementation, we describe a simple practical initialization. We use common inverse-gamma (IG) hyperparameters across units, separately for the structural-shock variances and the measurement-error variances:

$$\sigma_i^2 \overset{ind}{\sim} IG(a_\sigma, b_\sigma), \qquad \omega_i^2 \overset{ind}{\sim} IG(a_\omega, b_\omega), \qquad i = 1, \dots, N.$$

Preliminary completed panel and initial γ. Let $n_i = \sum_{t=1}^{T} s_{it}$. Construct a preliminary completed panel $\tilde{Y}_t$ by filling missing entries with a simple additive unit-time mean approximation. Let $\bar{Y}_i$ be the mean of observed outcomes for unit i,$\bar{Y}_t$ the mean of observed outcomes at date t, and $\bar{Y}$ the grand mean over observed entries. Set

$$\tilde{Y}_{it} \coloneqq \begin{cases} Y_{it}, & s_{it} = 1, \\ \bar{Y}_i + \bar{Y}_t - \bar{Y}, & s_{it} = 0 \text{ and } n_i > 0, \\ \bar{Y}_t, & s_{it} = 0 \text{ and } n_i = 0. \end{cases}$$

If a reliable baseline estimate is available, initialize $\hat{\gamma}^{(0)}$ at that value. Otherwise, run the complete-data baseline Network GIV procedure on $\{\tilde{Y}_t\}_{t=1}^{T}$.

Initial factor component and preliminary residuals. Given $\hat{\gamma}^{(0)}$, form preliminary structural residuals

$$\tilde{r}_t = \tilde{Y}_t - \sum_{\ell=0}^{L} B_\ell(\hat{\gamma}^{(0)}) \tilde{Y}_{t-\ell}, \qquad t = 1, \dots, T,$$

with $\tilde{Y}_t = 0$ for $t \leq 0$. Initialize $(\Lambda^{(0)}, \{\eta_t^{(0)}\}_{t=1}^{T})$ by the rank- r factor approximation to $\{\tilde{r}_t\}_{t=1}^{T}$, using the same normalization as in the main algorithm. Equivalently, compute

$$(\Lambda^{(0)}, \{\eta_t^{(0)}\}_{t=1}^{T}) \in \arg\min_{\Lambda, \{\eta_t\}} \sum_{t=1}^{T} \|\tilde{r}_t - \Lambda \eta_t\|_2^2$$

and then impose the baseline normalization. Define the preliminary idiosyncratic shock

$$\tilde{u}_t \coloneqq \tilde{r}_t - \Lambda^{(0)}\eta_t^{(0)}.$$

Initial structural-shock variances. Initialize the structural-shock variance using the average residual variance across units and periods: $\tilde{\sigma}^2 \coloneqq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2$. We then set the initial variance equal across units: $\hat{\sigma}_i^{2,(0)} = \tilde{\sigma}^2$.

Initial measurement-error variances. Let $n_{\text{obs}} = \sum_{i,t} s_{it}$ denote the total number of observed unit-period outcomes, and define the pooled mean among observed entries as $\bar{Y}_{\text{obs}} \coloneqq \frac{1}{n_{\text{obs}}}\sum_{i,t} s_{it}Y_{it}$. Initialize the measurement-error variance as a small fraction of the pooled variance of the observed outcomes:

$$\tilde{v}_Y = \frac{1}{\max\{n_{\text{obs}} - 1, 1\}}\sum_{i,t} s_{it}\big(Y_{it} - \bar{Y}_{\text{obs}}\big)^2, \qquad \tilde{\omega}^2 = 0.1\,\tilde{v}_Y.$$

We then set the initial measurement-error variance equal across units: $\hat{\omega}_i^{2,(0)} = \tilde{\omega}^2$ for all i.

Initial measurement-error correlation. From the preliminary observed-cell measurement-error residuals $\hat{\varepsilon}_{it}^{(0)} = Y_{it} - \tilde{Y}_{it}^{\star,(0)}$, compute pairwise sample correlations on $i \neq j$ using only pairs with $s_{it}s_{jt} = 1$, and set $\hat{\rho}^{(0)} = \text{median}_{i\neq j}\hat{r}_{ij}$, projected onto $[\rho_{\min} + \delta, \rho_{\max} - \delta]$ with $\delta = 10^{-3}$. The Beta prior on τ can additionally be empirically-Bayes-centred at the mean by setting $\alpha_\rho = 2$ and $\beta_\rho = 2(1 - \hat{\tau}^{(0)})/\hat{\tau}^{(0)}$, with $\hat{\tau}^{(0)} = (\hat{\rho}^{(0)} - \rho_{\min})/(\rho_{\max} - \rho_{\min})$.

Empirical-Bayes calibration of the inverse-gamma priors. We use a simple empirical-Bayes calibration that centers the priors at the same pooled variance values used for initialization. Set $a_\sigma = a_\omega = 3$. Since an $IG(a, b)$ distribution has mean $b/(a-1)$, choose

$$b_\sigma = 2\tilde{\sigma}^2, \qquad b_\omega = 2\tilde{\omega}^2.$$

Thus the prior mean for each σ_i^2 is $\tilde{\sigma}^2$, and the prior mean for each ω_i^2 is $\tilde{\omega}^2$. The algorithm is initialized at these same pooled values: $D_\sigma^{(0)} = \tilde{\sigma}^2 I$, $D_\omega^{(0)} = \tilde{\omega}^2 I$.

A.3.3 Monte Carlo E-step

Posterior quantities can be obtained either by simulation or, thanks to the Gaussian environment, analytically.

We now describe how to compute the posterior quantities in the E-step by simulation. Fix γ and the current nuisance values

$$\Lambda^{(m)}, \quad \{\eta_t^{(m)}\}_{t=1}^T, \quad D_\sigma^{(m)}, \quad D_\omega^{(m)}.$$

For a candidate latent path $y_{1:T}^\star = (y_1^\star, \ldots, y_T^\star)$, define the structural innovation implied by that path as

$$u_t(y^\star) \coloneqq y_t^\star - \sum_{\ell=0}^{L} B_\ell(\gamma) y_{t-\ell}^\star - \Lambda^{(m)} \eta_t^{(m)}, \qquad t = 1, \ldots, T, \tag{A.12}$$

with $y_t^\star = 0$ for $t \leq 0$. Given the observed panel and masks, the posterior density of the latent path is proportional to

$$\exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} u_t(y^\star)^\top \big(D_\sigma^{(m)}\big)^{-1} u_t(y^\star) - \frac{1}{2} \sum_{t=1}^{T} \big(Y_t^{\mathcal{O}_t} - y_t^{\star,\mathcal{O}_t}\big)^\top \big[\Omega_{\mathcal{O}_t\mathcal{O}_t}^{(m)}\big]^{-1} \big(Y_t^{\mathcal{O}_t} - y_t^{\star,\mathcal{O}_t}\big) \right\}. \tag{A.13}$$

where $\mathcal{O}_t = \{i : s_{it} = 1\}$ is the set of units observed at date t and $\Omega_{\mathcal{O}_t\mathcal{O}_t}^{(m)}$ denotes the principal submatrix of $\Omega^{(m)} = (D_\omega^{(m)})^{1/2} R(\rho^{(m)}) (D_\omega^{(m)})^{1/2}$ on rows/columns $\mathcal{O}_t$. At $\rho = 0$ this collapses to the original diagonal expression.Draw $Y_{1:T}^{\star,1}, \ldots, Y_{1:T}^{\star,B_{\mathrm{MC}}}$ from the posterior density in (A.13). For each draw $b = 1, \ldots, B_{\mathrm{MC}}$, compute

$$u_t^b \coloneqq Y_t^{\star,b} - \sum_{\ell=0}^{L} B_\ell(\gamma) Y_{t-\ell}^{\star,b} - \Lambda^{(m)} \eta_t^{(m)}.$$

For observed entries, define the implied measurement error

$$\varepsilon_{it}^b \coloneqq Y_{it} - Y_{it}^{\star,b}, \qquad \text{for } s_{it} = 1.$$

The posterior means used in the E-step are approximated by

$$\hat{Y}_t^{\star,(m+1)} = \frac{1}{B_{\mathrm{MC}}} \sum_{b=1}^{B_{\mathrm{MC}}} Y_t^{\star,b}, \qquad \hat{u}_t^{(m+1)} = \frac{1}{B_{\mathrm{MC}}} \sum_{b=1}^{B_{\mathrm{MC}}} u_t^b.$$

The posterior second moments for the variance updates are approximated by

$$\widehat{\mathbb{E}}_m[u_{it}^2 \mid Y_{1:T}] = \frac{1}{B_{\mathrm{MC}}} \sum_{b=1}^{B_{\mathrm{MC}}} (u_{it}^b)^2,$$

and, for observed entries,

$$\widehat{\mathbb{E}}_m[\varepsilon_{it}^2 \mid Y_{1:T}] = \frac{1}{B_{\mathrm{MC}}} \sum_{b=1}^{B_{\mathrm{MC}}} (Y_{it} - Y_{it}^{\star,b})^2.$$

These Monte Carlo averages replace the posterior expectations in the M-step. With inverse-gamma priors $\sigma_i^2 \sim IG(a_\sigma, b_\sigma)$, $\omega_i^2 \sim IG(a_\omega, b_\omega)$, the MAP variance updates become

$$\hat{\sigma}_i^{2,(m+1)} = \frac{b_\sigma + \frac{1}{2}\sum_{t=1}^{T} \widehat{\mathbb{E}}_m[u_{it}^2 \mid Y_{1:T}]}{a_\sigma + 1 + \frac{T}{2}},$$

and

$$\hat{\omega}_i^{2,(m+1)} = \frac{b_\omega + \frac{1}{2}\sum_{t=1}^{T} s_{it} \widehat{\mathbb{E}}_m[\varepsilon_{it}^2 \mid Y_{1:T}]}{a_\omega + 1 + \frac{1}{2}\sum_{t=1}^{T} s_{it}}.$$

The Monte Carlo E-step is therefore simple: draw latent complete panels from (A.13), compute the implied shocks and measurement errors from the draws, and replace all posterior expectations by sample averages over the draws.

A.3.4 Analytic E-step

Under the Gaussian environment, posterior expectations are available in closed form: the posterior density (A.13) is the exponential of a quadratic form in the latent path, so the posterior is exactly Gaussian, and its first two moments deliver every quantity the E-step requires. The analytic E-step computes those moments directly, replacing the simulation draws by a single linear solve. Therefore, the analytic route makes the posterior mean exact, whereas the Monte Carlo mean carries simulation error of order $B_{\mathrm{MC}}^{-1/2}$; and since each posterior draw is itself a solve of the same system, the analytic mean is also the cheaper object, one solve in place of B_{MC} draws.

Stack the latent path and the observed panel over time into vectors in $\mathbb{R}^{NT}$. Stacking the structural innovations $u_t(y^\star)$ of (A.12) the same way makes the map $y^\star \mapsto u$ affine, $u(y^\star) =$

$\mathcal{B}y^\star - \Lambda^{(m)}\eta^{(m)}$, where $\mathcal{B}$ is the $NT \times NT$ operator that sends $y_t^\star$ to $M(\gamma)y_t^\star - \sum_{\ell=1}^{L} B_\ell(\gamma)y_{t-\ell}^\star$. Because each innovation depends only on the current and the L preceding dates, $\mathcal{B}$ is block lower-triangular and block-banded in time with bandwidth L. Two stacked variance objects are: $\mathcal{D}_\sigma = I_T \otimes D_\sigma^{(m)}$, the structural-shock variances of the whole panel, and the measurement-error precision $\mathcal{W}_\omega$, the block-diagonal matrix whose t-th block is$\tilde{S}_t^\top \left[\Omega_{\mathcal{O}_t\mathcal{O}_t}^{(m)}\right]^{-1}\tilde{S}_t$, with $\tilde{S}_t \in \{0,1\}^{|\mathcal{O}_t| \times N}$ the observed-row selection matrix (reducing to $S_t(D_\omega^{(m)})^{-1}S_t$ when $\rho = 0$) and which is therefore zero on the missing cells. Both are simply the panel-stacked forms of the variance matrices $D_\sigma^{(m)}$, $D_\omega^{(m)}$ and the mask S_t.

Substituting $u = \mathcal{B}y^\star - \Lambda^{(m)}\eta^{(m)}$ into (A.13), the negative log-posterior of $y^\star$ is, up to an additive constant,

$$\tfrac{1}{2}(\mathcal{B}y^\star - \Lambda^{(m)}\eta^{(m)})^\top \mathcal{D}_\sigma^{-1}(\mathcal{B}y^\star - \Lambda^{(m)}\eta^{(m)}) + \tfrac{1}{2}(y^\star - Y)^\top \mathcal{W}_\omega (y^\star - Y).$$

This is quadratic in $y^\star$, so the posterior is exactly Gaussian. Write Σ for its covariance; its mean is the panel posterior mean $\hat{Y}^{\star,(m+1)}$ that the E-step reports. The precision and the mean satisfy

$$\Sigma^{-1} = \mathcal{B}^\top \mathcal{D}_\sigma^{-1}\mathcal{B} + \mathcal{W}_\omega, \qquad \Sigma^{-1}\hat{Y}^{\star,(m+1)} = \mathcal{B}^\top \mathcal{D}_\sigma^{-1}\Lambda^{(m)}\eta^{(m)} + \mathcal{W}_\omega Y.$$

The shock posterior mean then follows exactly, $\hat{u}^{(m+1)} = \mathcal{B}\hat{Y}^{\star,(m+1)} - \Lambda^{(m)}\eta^{(m)}$. These are the quantities the Monte Carlo E-step approximates by averaging draws; here they carry no simulation error and cost a single solve. Because Σ^{-1} inherits the block-banded sparsity of $\mathcal{B}$, the system is solved by a sparse factorization.

The inverse-gamma variance updates additionally require the posterior second moments $\mathbb{E}_m[u_{it}^2 \mid Y_{1:T}]$ and, on observed cells, $\mathbb{E}_m[\varepsilon_{it}^2 \mid Y_{1:T}]$. Each splits into a squared-mean term, already exact once $\hat{Y}^{\star,(m+1)}$ is known, and a variance term,

$$\mathbb{E}_m[u_{it}^2 \mid Y_{1:T}] = \left(\hat{u}_{it}^{(m+1)}\right)^2 + \mathrm{Var}(u_{it} \mid Y_{1:T}), \qquad \mathbb{E}_m[\varepsilon_{it}^2 \mid Y_{1:T}] = (Y_{it} - \hat{Y}_{it}^{\star,(m+1)})^2 + \mathrm{Var}(y_{it}^\star \mid Y_{1:T}),$$

the second identity using $\varepsilon_{it} = Y_{it} - y_{it}^\star$ on observed entries. The two variance terms are the

diagonals of the posterior covariance and of its image under $\mathcal{B}$,

$$\mathrm{Var}(y^\star \mid Y_{1:T}) = \mathrm{diag}(\Sigma), \qquad \mathrm{Var}(u \mid Y_{1:T}) = \mathrm{diag}(\mathcal{B}\Sigma\mathcal{B}^\top).$$

Once $\hat{Y}^{\star,(m+1)}$ is known, evaluating these two diagonals is all that remains of the E-step.

There are two ways of evaluating them. A first, exact, way computes $\mathrm{diag}(\Sigma)$ and $\mathrm{diag}(\mathcal{B}\Sigma\mathcal{B}^\top)$ without approximation: the block-banded sparsity of Σ^{-1} makes the required entries of Σ available from the same sparse factorization used for the mean. However, recovering those entries becomes costly as NT grows, so this variant is best suited to panels of moderate size.

A second, computationally faster, option is to use a Hutchinson estimator of the same diagonals which scales to large panels. It rests on the identity $\mathbb{E}[q \odot (Aq)] = \mathrm{diag}(A)$, valid for any probe vector q with independent, mean-zero entries normalized so that $\mathbb{E}[qq^\top] = I$ (for instance Rademacher entries equal to ± 1 with equal probability). Drawing K probes and solving $\Sigma^{-1}x^{(k)} = q^{(k)}$ estimates

$$\mathrm{diag}(\Sigma) \approx \tfrac{1}{K}\sum_{k=1}^{K} q^{(k)} \odot x^{(k)}.$$

Solving $\Sigma^{-1}x^{(k)} = \mathcal{B}^\top q^{(k)}$ and averaging $q^{(k)} \odot (\mathcal{B}x^{(k)})$ estimates $\mathrm{diag}(\mathcal{B}\Sigma\mathcal{B}^\top)$.

A.3.5 Computing the projection coefficient

Finally, we describe how to compute the projection coefficient π_i. The coefficient π_i is a model-implied projection coefficient. It is not computed by conditioning on the realized observed panel alone. Rather, it is computed by averaging over synthetic panels generated from the fitted Gaussian model, holding fixed the realized masks $S_1, \ldots, S_T$ and the current parameter estimates. Fix the current estimates of

$$\widehat{\theta} = \left(\widehat{\gamma}, \widehat{\Lambda}, \{\widehat{\eta}_t\}_{t=1}^T, \widehat{D}_\sigma, \widehat{D}_\omega\right). \tag{A.14}$$

For $b = 1, \ldots, B_{\mathrm{MC}}$, generate a synthetic dataset as follows. First draw

$$u_t^b \sim N(0, \hat{D}_\sigma), \qquad \varepsilon_t^b \sim N\big(0, \hat{D}_\omega^{1/2} R(\hat{\rho}) \hat{D}_\omega^{1/2}\big), \qquad t = 1, \ldots, T,$$

independently over time. Then generate the latent path recursively from

$$M(\hat{\gamma})Y_t^{\star,b} = \sum_{\ell=1}^{L} B_\ell(\hat{\gamma})Y_{t-\ell}^{\star,b} + \hat{\Lambda}\hat{\eta}_t + u_t^b, \qquad Y_t^{\star,b} = 0 \text{ for } t \le 0.$$

Finally, apply the realized mask and measurement error: $\widetilde{Y}_t^b = S_t\big(Y_t^{\star,b} + \varepsilon_t^b\big)$. For each synthetic panel $\widetilde{Y}_{1:T}^b$, compute the posterior means

$$\overline{Y}_t^{\star,b} := E_{\widehat{\theta}}\left[Y_t^\star \mid \widetilde{Y}_{1:T}^b, S_{1:T}\right], \qquad \widetilde{\xi}_{it}^b := E_{\widehat{\theta}}\left[u_{it} \mid \widetilde{Y}_{1:T}^b, S_{1:T}\right],$$

using the same posterior formula as in the E-step, but holding fixed the parameters $\widehat{\theta}$. That is, for each synthetic panel we perform only the posterior-expectation calculation; we do not update γ, the factor path used in the posterior density, or the variance components.

Using the posterior mean of the latent panel, form the synthetic posterior structural residual

$$\widetilde{v}_t^b = \overline{Y}_t^{\star,b} - \sum_{\ell=0}^{L} B_\ell(\widehat{\gamma})\overline{Y}_{t-\ell}^{\star,b}.$$

Then construct $\widetilde{z}_{it}^{0,b}$ by applying the same leave-one-out Network GIV construction used in the realized data to the residual panel $\{\widetilde{v}_t^b\}_{t=1}^T$: for each target unit i, recompute the leave-one-out factor score, the leave-one-out residual path, the propagated state, and the corresponding raw derivative instrument, holding fixed the fitted parameter values. Thus the Monte Carlo projection coefficient is computed using the same raw-instrument statistic that enters the realized-data update.

The pooled projection coefficient is then estimated by

$$\hat{\pi}_i = \frac{\sum_{b=1}^{B_{\mathrm{MC}}}\sum_{t=1}^{T}\widetilde{z}_{it}^{0,b}\widetilde{\xi}_{it}^b}{\sum_{b=1}^{B_{\mathrm{MC}}}\sum_{t=1}^{T}(\widetilde{\xi}_{it}^b)^2}. \tag{A.15}$$

This is the Monte Carlo analogue of (A.9). Given $\hat{\pi}_i$, the projected instrument in the realized data is $z_{it}^\perp = z_{it}^0 - \hat{\pi}_i\xi_{it}$.

B Algorithmic Details for the Econometric Procedure

B.1 Standard errors estimation

We report the standard errors from the last iteration of Algorithm 1. Specifically, for each component of γ, we take the usual OLS standard error from the final regression used to update that component, map it back to the original parameter scale when needed, and then obtain the standard errors of all reported multipliers by the delta method evaluated at the final estimate.

Algorithm 4 Standard error estimation for $\hat{\gamma}$ and multipliers

Require: Vector $\hat{\Delta\gamma}$ from the last iteration of Algorithm 1, vector $\hat{\varepsilon}$ of residuals from regression (13), r number of factors, N number of sectors, T number of periods and $p = \dim(\gamma)$.

1: Estimate residual variance by $\hat{\sigma}^2_\varepsilon = \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{NT - p - r(N+T)}$, as we have NT observations, but $p + r\,(N+T)$ degrees of freedom,

2: The covariance matrix of the last-step estimator is $\hat{\Sigma}_{\Delta\gamma} = \hat{\sigma}^2_\varepsilon (zz^\top)^{-1}$ Hence, the standard error of the k-th component is $\hat{SE}(\hat{\Delta\gamma_k}) = \sqrt{[\hat{\Sigma}_{\Delta\gamma}]_{kk}}$ for $k = 1, \ldots, p$.

3: Next note that $\hat{SE}(\hat{\Delta\gamma_k}) = \hat{SE}(\hat{\gamma_k})$.

4: Let $m(\gamma)$ be any scalar multiplier of interest, for example (34) or (35). Then, using the delta method its variance is

$$\hat{\mathrm{Var}}(m(\hat{\gamma})) = \nabla_\gamma m(\hat{\gamma})^\top \hat{\Sigma}_\gamma \nabla_\gamma m(\hat{\gamma}),$$

so that the standard error is $\hat{SE}(m(\hat{\gamma})) = \sqrt{\hat{\mathrm{Var}}(m(\hat{\gamma}))}$.

5: **return** $\hat{SE}(\hat{\gamma_k})$=, $\hat{SE}(m(\hat{\gamma}))$ for all k, j.

B.2 Heteroskedastic PCA

In this section we present the algorithm to estimate unobserved loadings in the presence of heteroskedastic residuals. We consider a model of the form

$$v_t = \lambda \eta_t + \Sigma_\varepsilon^{1/2} \varepsilon_t, \quad \mathbb{E}[\varepsilon_t^2] = \sigma^2_\varepsilon, \quad \mathbb{E}[\varepsilon_t | \eta_t] = 0,$$

where Σ_ε is a diagonal matrix capturing heteroskedastic residuals. We consider two competing algorithms for heteroskedastic PCA. The first algorithm is an alternating optimization

algorithm, while the second one was first developed by Zhang, Cai, and Wu (2022), and consists of solving the equivalent of a projected gradient descent method that imputes the diagonal entries of the covariance matrix of the residuals, and then performs PCA on the remaining component.

Algorithm 5 Heteroskedastic PCA (Version 1): estimate the loadings from a vector of heteroskedastic residuals v_t via PCA

Require: Vector v_t of residuals, number of iterations I, number of components r

1: Initialize by setting $\Sigma_\varepsilon^0 = \text{diag}((\sigma_{v_i}^2)_{=1\dots n})$, where $\sigma_{v_i}^2$ is the estimated variance of the v_{it}; $\hat{\Sigma}_{v,0} = \frac{1}{T}\sum_t (\Sigma_{\varepsilon,0}^{-1/2} v_t)(\Sigma_{\varepsilon,0}^{-1/2} v_t)'$

2: **for** $k \in \{0, 1, \cdots, I\}$ **do**

3: Do Singular Value Decomposition (SVD) of $\hat{\Sigma}_{v,k}$ and retrieve the r largest eigenvectors denoted as $\tilde{\lambda}^k$

4: Compute $\eta_t^k = \tilde{\lambda}^{k\top} \Sigma_{\varepsilon,k}^{-1/2} v_t$ (where $\tilde{\lambda}^{k\top}\tilde{\lambda}^k = \mathbb{I}$)

5: Compute $\hat{\lambda}^k = \Sigma_{\varepsilon,k}^{1/2} \tilde{\lambda}_k$

6: Compute

$$\Sigma_{\varepsilon,k,i,j} = \begin{cases} \frac{1}{T}\sum_t (v_{i,t} - \hat{\lambda}_i^k \eta_t^k)^2 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$$

7: **end for**

8: **return** $\hat{\lambda}^I, \Sigma_{\varepsilon,I}$

Algorithm 6 Heteroskedastic PCA (Zhang, Cai, and Wu 2022) (Version 2)

Require: Vector v_t, number of iterations I, number of components r

1: Define the covariance matrix $\hat{\Sigma}_v = \frac{1}{T}\sum_t v_t v_t'$

2: Initialize by setting $N^0 = \hat{\Sigma}_v, \text{diag}(N^0) = \mathbf{0}$

3: **for** $k \in \{0, 1, \cdots, I\}$ **do**

4: Perform SVD on N^k and denote $\tilde{N}^k$ its best rank- r approximation

5: Update

$$N_{i,j}^{k+1} = \begin{cases} \tilde{N}_{i,j}^k \quad \text{if } i = j \\ \hat{\Sigma}_{i,j} \text{ otherwise} \end{cases}$$

6: **end for**

7: **return** The first r singular vectors of N^I defined as $\hat{\lambda}$ and the variances of the residuals as

$$\Sigma_\varepsilon = \text{diag}(\hat{\Sigma}_v) - \text{diag}(N^I)$$

B.3 Time fixed effects via demeaning

As mentioned in Step 2 of Algorithm 1, we can control for time fixed effects in the estimation procedure by applying PCA to the demeaned residuals $\check{v}_t$.

When there is just one outcome per entity In the simple case where, for each unit $i = 1, \ldots, N$, we observe a single outcome (that is, in our notation, $q = 1$), we do the following:

1. For each i, t, define the demeaned residuals $\check{v_{it}}(\hat{\gamma}) \coloneqq \tilde{v_{it}}(\hat{\gamma}) - \tilde{v}_{Et}(\hat{\gamma})$, where $\tilde{v}_{Et}(\hat{\gamma})$ is the mean outcome in the time- t cross-section;

2. Run PCA with heteroskedasticity using Algorithm 5 passing $\check{v}_t$ as the argument and r as the number of factors. The PCA algorithm iterates to estimate loadings $\check{\lambda}(\hat{\gamma})$ and factors $\check{\eta}_t(\hat{\gamma})$;

3. Add a N-dimensional vector of ones to the estimated loadings, $\tilde{\lambda}(\hat{\gamma}) = \left(\iota, \check{\lambda}(\hat{\gamma})\right)$ where $\iota = (1, \ldots, 1)$, and add the residuals' means to the estimated factors, $\tilde{\eta}_t(\hat{\gamma}) = (\tilde{v}_{Et}(\hat{\gamma}), \check{\eta}_t(\hat{\gamma}))$. In this way, we effectively consider $r + 1$ factors, with the first factor capturing a time fixed effect.

General case, with potentially more than one outcome per entity In general, we may have $q \geq 1$ observable outcomes per unit – for instance, in our macroeconomic application, we have $q = 2$ observables: prices and quantities. When $q = 2$, the procedure becomes as follows (the generalization to all q is natural):

1. For each i, t, demean the two outcomes separately $\check{v}^k_{it}(\hat{\gamma}) \coloneqq \tilde{v}^k_{it}(\hat{\gamma}) - \tilde{v}^k_{Et}(\hat{\gamma})$ for $k = 1, 2$, where $\tilde{v}^k_{Et}(\hat{\gamma})$ is the mean for outcome k in the time- t cross-section;

2. Combine the two vectors of demeaned residuals $\check{v}_t(\hat{\gamma}) = (\check{v}_t{}^1(\hat{\gamma}), \check{v}_t{}^2(\hat{\gamma}))$, and apply heteroskedastic PCA to $\check{v}_t(\hat{\gamma})$. Algorithm 5 returns the estimated loadings $\check{\lambda}(\hat{\gamma})$ and the estimated factors $\check{\eta}_t(\hat{\gamma})$;

3. Add two qN-dimensional vectors of ones and zeros to the estimated loadings, $\tilde{\lambda}(\hat{\gamma}) = \left(\iota^1, \iota^2, \check{\lambda}(\hat{\gamma})\right)$, where $\iota^1 = (1, \ldots, 1, 0, \ldots 0)$ and $\iota^2 = (0, \ldots, 0, 1, \ldots 1)$, and add the two residuals' means to the estimated factors, $\tilde{\eta}_t(\hat{\gamma}) = (\tilde{v}^1_{Et}(\hat{\gamma}), \tilde{v}^2_{Et}(\hat{\gamma}), \check{\eta}_t(\hat{\gamma}))$. In this way, we effectively consider $r + 2$ factors, with the first two capturing a time fixed effect.

C The Network GIV with extra equilibrium relations

Typically, there are extra equilibrium relations, e.g. supply equals demand. We show how the procedure allows for them, with little extra work.

For concreteness, we use the supply and demand model of Example 2, but slightly rewritten as follows:

$$q_{it} = \phi_i p_t + \eta_{1t} + u_{it} \tag{C.1}$$

$$s_t = \zeta p_t + \eta_{2t} \tag{C.2}$$

where ϕ_i is a demand elasticity, and ζ is a supply elasticity. Also, supply equals demand $q_{St} \coloneqq \sum_i S_i q_{it}$, where S_i are size shares:

$$q_{St} = s_t \tag{C.3}$$

Equation (C.1) is the demand by firm $i \in \{1, \ldots, n\}$, which has log demand q_{it} reacting to the log price p_t with demand elasticity ϕ_i. Equation (C.2) is the supply equation. We assume that u_t is uncorrelated with η_t.

The general version of this is: there is a set of variables $X \in \mathbb{R}^{d_X}$, which are decomposed into left-hand side variables $X^L \in \mathbb{R}^{d_L}$, and extra observables $X^R \in \mathbb{R}^{d_R}$: $X = \left[X^L, X^R\right]$, i.e. the concatenation of the two vector. The primitive system is

$$X^L = BX + u + \varepsilon_L \in \mathbb{R}^{d_L} \tag{C.4}$$

and there may be an equilibrium relation:

$$DX = \varepsilon_R \in \mathbb{R}^{d_R} \tag{C.5}$$

where ε_R could be 0, or perhaps a disturbance (e.g. with missing data).

In our oil example, we have $X^L = (q_1, \ldots, q_n, s)'$, $X^R = (p)$. The equilibrium relation (C.5) is (C.3), so $D = (S_1, \ldots, S_n, -1)$ and $\varepsilon_R = 0$.

Stacking together the relations, we have (decomposing e.g. $B = \begin{pmatrix} B_L & B_R \end{pmatrix}$)

$$\Omega X = \begin{pmatrix} u + \varepsilon_L \\ \varepsilon_R \end{pmatrix}, \qquad \Omega \coloneqq \begin{pmatrix} I - B_L & -B_R \\ D_L & D_R \end{pmatrix} \tag{C.6}$$

We assume that Ω is invertible, so we can express $X = \Omega^{-1} \begin{pmatrix} u + \varepsilon_L \\ \varepsilon_R \end{pmatrix}$, which gives

$$X_u \coloneqq \frac{dX}{du} = \left(\Omega^{-1}\right)_{d_X \times d_L}$$

which is the left submatrix of Ω^{-1}, with dimensions $d_X \times d_L$. Idiosyncratic shocks Δu create an impact $\Delta X = X_u \Delta u$, and this shows how to construct matrix X_u.

So, the algorithm is the following, expanding on the intuition in Section 2.2. We start with a guess γ, around the target correct value $\gamma^\star$, and define $\Delta\gamma \coloneqq \gamma^\star - \gamma$. We form

$$v \coloneqq X^L - B(\gamma) X$$

Then, we observe that

$$v = X^L - B(\gamma^\star) X - (B(\gamma) - B(\gamma^\star)) X = \Delta\gamma \cdot B_\gamma X + \varepsilon + O\left(\|\Delta\gamma\|^2\right)$$

where $\Delta\gamma$ is small, so that up to second order terms:

$$v = \Delta\gamma \cdot B_\gamma X + \varepsilon$$

Hence, to estimate $\Delta\gamma$, in the regression with v_{it} on the left-hand side, we use $X^{z,-i} = X_u u^{-i}$ as the instrument the right-hand side X, where u^{-i} is an estimate of u, uncorrelated with u_{it} (for instance, not using observation it, and purified by a factor model, as in the main body of the paper). So, the regression to obtain the regression coefficient $\Delta\gamma$ is:

$$v_{it} = \Delta\gamma \cdot \left(e_i' B_\gamma X_t^{z,-i}\right) + \varepsilon_i, \qquad X^{z,-i} \coloneqq X_u u_t^{-i} \tag{C.7}$$

where e_i' selects the i-th component. Note that we do not have a "first stage" regressing

$e_i' B_\gamma X_t^{z,-i}$ on u^{-i}, as we know (from the model) the relation. Alternatively, though, it could also be estimated.

We next show how this mechanism applies to our simple example.

Application to the estimation of the supply elasticity In our application, take the component of γ equal to ζ. Also, define $E \coloneqq (1, \ldots, 1)/N$, the equal-weighing vector. Then, we have $B_\zeta X = \begin{pmatrix} 0_{n\times 1} \\ p \end{pmatrix}$. Also, solving for the endogenous variables gives $\frac{dp}{du} u = \frac{u_S}{\zeta - \phi}$. Now, we estimate $\eta^e \coloneqq q_E - \phi p$, where $q_E = \frac{1}{N}\sum_i q_i$ is the equal-weighted mean and $u_i^e \coloneqq q_i - \phi p - \eta^e$, so the estimated idiosyncratic demand shock is: $u_i^e = q_i - q_E$. Hence,

$$u_S^e = q_S - q_E =: q_\Gamma, \tag{C.8}$$

where $\Gamma \coloneqq S - E$, the "size-minus equal- weighted shocks" that is central in Gabaix and Koijen (2024). Hence, following the general procedure, the instrumented price is

$$p^z = p_u u^e = \frac{u_\Gamma}{\zeta - \phi} = \frac{q_\Gamma}{\zeta - \phi}$$

Hence, the regression (C.7) is

$$s - \zeta p = (\Delta\zeta)\, p^z \text{ with } p^z = \frac{q_\Gamma}{\zeta - \phi} \tag{C.9}$$

The moment at convergence corresponds to $\Delta\zeta = 0$, i.e.:

$$\mathbb{E}\left[(s - \zeta p)\, p^z\right] = 0$$

exactly like in Gabaix and Koijen (2024). We recover that basic result, in our new general machinery. But the new machinery works in much more complex cases.

Application to the estimation of the demand elasticity. This is similar. We start with the basic case of homogeneous demand elasticities, $\phi_i = \phi$. We have $B_\phi X = \begin{pmatrix} 1_{n\times 1} p \\ 0_{1\times 1} \end{pmatrix}$, and so again we need to instrument the price. For a given target unit i, we have $\eta^{e,-i} =$

$q_{E^{-i}} - \phi p$, where $q_{E^{-i}}$ is the equal mean, among the $n-1$ units different from i. So, $u_j^{-i} = q_j - \phi p - \eta^{e,-i} = q_j - q_{E^{-i}}$. Hence, the instrument for the price is

$$p^{z,-i} = \frac{\sum_{j \neq i} S_j \left(q_j - q_{E^{-i}}\right)}{\zeta - \phi}$$

The regression (C.7) is:

$$v_{it} = \Delta\phi \cdot p_t^{z,-i} + e_{it} \tag{C.10}$$

hence at convergence we impose the moment condition:

$$\mathbb{E}\left[\left(q_{it} - \phi p_t\right) p_t^{z,-i}\right] = 0 \tag{C.11}$$

This is very close to, though not exactly, the value in Gabaix and Koijen (2024). The reason is that they use the homoskedasticity of the u_{it}, which allows them to avoid the"leave one out" detour. Their condition boils down to $\mathbb{E}\left[\sum_i \left(q_{it} - \phi p_t\right) p_t^z\right] = 0$ with $p^z = \frac{q_S - q_E}{\zeta - \phi}$.

If we have heterogeneous elasticities, this is similar. We have $B_{\phi_i} X = \begin{pmatrix} 0_{(i-1)\times 1} \\ p \\ 0_{(n+1-i)\times 1} \end{pmatrix}$. To explain unit i, we form $\eta^{e,-i} \coloneqq \langle q_i - \phi_j p \rangle_{j \neq i}$, and $u_{jt}^{-i} \coloneqq q_j - \phi_j p - \eta^{e,-i}$. So, the instrument for the price is

$$p^{z,-i} = \frac{\sum_{j \neq i} S_j u_{jt}^{-i}}{\zeta - \phi_S}$$

To estimate $\Delta\phi_i = \phi_i^\star - \phi_i$, we regress:

$$v_{it} = \Delta\phi_i \cdot p_t^{z,-i} + e_{it} \tag{C.12}$$

This gives a new estimate $\phi_i' = \phi_i + \Delta\phi_i$, and we loop until convergence.

Online Appendix for "Propagation of shocks in networks: identification and applications"

Gabriel Chodorow-Reich, Xavier Gabaix and Davide Viviano

June 3, 2026. Preliminary and Incomplete.

D Complements to the Econometric Procedure

D.1 Proof of the main results

D.1.1 Proof of Proposition 1

We can write $z_{it}(\gamma^\star) = e_i^\top B_\gamma(\gamma^\star)A(\gamma^\star)u_t^{-i}$, where $u_t^{-i} = u_t - e_i u_{it}$. Therefore the result follows directly under the assumption that $E[u_{it}|u_{jt}{}_{j\neq i}] = 0$.

D.1.2 Proof of Lemma 1

Recall that we can write $z_{it}^{(k)}(\gamma^\star) = e_i^\top \Gamma^{(k)} u_t$. Therefore, we can write

$$\begin{aligned} V^z_{k,k'} &= \sum_i \mathbb{E}[z_{it}^{(k)}(\gamma^\star) z_{it}^{(k')}(\gamma^\star)] = \sum_i \mathbb{E}\left[\left(\sum_j \Gamma_{ij}^{(k)} u_{jt}\right)\left(\sum_{j'} \Gamma_{ij'}^{(k')} u_{j't}\right)\right]. \\ &= \sum_i \sum_j \Gamma_{ij}^{(k)} \Gamma_{ij}^{(k')} \sigma^2_{u_j} \end{aligned}$$

D.1.3 Proof of Proposition 2

The asymptotic normality in Proposition 2 follows directly from standard arguments for extremum estimators (Newey and McFadden 1994). Here, we are left to show the exact expression for the variance.

From standard properties of the methods of moments estimator with exact identification, we can write the variance of the estimator as $\bar{V} = \Big(N\bar{\Delta}\Big)^{-1}\Sigma\Big(N\bar{\Delta}\Big)^{-1\top}$ where $\Sigma = E[\Big(\sum_i H_{it}(\gamma^\star)\Big)\Big(\sum_i H_{it}(\gamma^\star)\Big)^\top]$ and $N\bar{\Delta} = -\sum_i \frac{\partial E[H_{it}(g)]}{\partial g}|_{g=\gamma^\star}$.

Denote $\sigma_i^2 = \mathbb{E}[u_i^2]$, and $H_{i,t}(\gamma, g) = \Big(Y_{i,t} - e_i^\top B(\gamma)Y_t\Big)e_i^\top \frac{\partial B(\gamma)}{\partial \gamma} y_{t|-i}(\gamma)$, where $y_{t|-i}(\gamma) = A(\gamma)[u_t - e_i u_{i,t}]$. Also, recall the definition of $G^{(k)} = \frac{\partial B(g)}{\partial g_k}\Big|_{g=\gamma^\star} A(\gamma^\star)$. First observe that because under Assumption 1, $Y_t = A(\gamma^\star)u_t$, we can write at the true $\gamma = \gamma^\star$

$$\frac{\partial\Big(Y_{i,t} - e_i^\top B(\gamma)Y_t\Big)}{\partial \gamma_k} = -e_i^\top \frac{\partial B(\gamma)}{\partial \gamma_k} A(\gamma)u_t,$$

and therefore

$$\begin{aligned}\frac{\partial y_{t|-i}(\gamma)}{\partial \gamma_k} &= \frac{\partial A(\gamma)(\mathbb{I} - e_i e_i^\top)\Big(Y_t - B(\gamma)Y_t\Big)}{\partial \gamma_k} \\ &= -A(\gamma)(\mathbb{I} - e_i e_i^\top)\frac{\partial B(\gamma)}{\partial \gamma_k} A(\gamma)u_t + \frac{\partial A(\gamma)}{\partial \gamma_k}\Big[u_t - e_i u_{i,t}\Big].\end{aligned}$$

Consider the component $\Delta^{(p_1,p_2)}$ first.We can write

$$\begin{aligned}-\frac{\partial \mathbb{E}[H_{i,t}^{(p_1)}(\gamma,\gamma)]}{\partial \gamma_{p_2}}\Big|_{\gamma=\gamma^\star} &= \mathbb{E}\Big[e_i^\top G^{(p_2)} u_t e_i^\top G^{(p_1)}[u_t - e_i u_{i,t}]\Big] + \mathbb{E}\Big[u_{i,t} e_i^\top G^{(p_1)}(\mathbb{I} - e_i e_i^\top)G^{(p_2)}u_t\Big] \\ &= \sum_{j\neq i} G_{i,j}^{(p_1)} G_{i,j}^{(p_2)} \sigma_j^2 + \sum_{j\neq i} G_{i,j}^{(p_1)} G_{j,i}^{(p_2)} \sigma_i^2\end{aligned} \tag{D.1}$$

where we used the fact that the remaining components of the derivative are mean zero (including the component involving the second order derivative of $B(\gamma)$). Using the fact that $\Gamma_{i,i} = 0$, and homoskedasticity, where we write $\sigma_i^2 = \sigma^2$, it follows that

$$-\sum_{i=1}^N \frac{\partial \mathbb{E}[H_{i,t}^{(p_1)}(\gamma,\gamma)]}{\partial \gamma_{p_2}}\Big|_{\gamma=\gamma^\star} = \sigma^2 \operatorname{tr}(\Gamma^\top\Gamma + \Gamma^2).$$

Finally, because

$$\sum_{i=1}^N H_{i,t}^{(k)}(\gamma^\star,\gamma^\star) = u_t^\top \Gamma^{(k)} u_t,$$

we can write the (k, k') entry as

$$E[u_t^\top \Gamma^{(k)} u_t u_t^\top \Gamma^{(k')} u_t] = E[\sum_i u_{i,t} \sum_j \Gamma^k_{i,j} u_{j,t} \sum_v u_{v,t} \sum_l \Gamma^{k'}_{v,l} u_{l,t}].$$

Now using independence across unit i and using the fact that $\Gamma_{i,i} = 0$ for all i we can simplify the expression as follows: first, we note that for any $i \neq j$, $E[u_i u_j] = 0$, $E[u_i^2] = \sigma^2$, and $E[u_i^2 u_j^2] = \sigma^4$. Thus, $E[u_i u_j u_v u_l] \neq 0$ precisely when the indexes are split into two matching pairs. The possible matching are $i = j, v = l, i = v, j = l, i = l, j = v$. However, because $\Gamma_{i,i} = 0$ for all i any term involving $i = j$ is automatically zero in the sum and similarly for $v = l$. Consequently, the only pairings that survive are $i = v \neq j = l, i = l \neq j = v$. Let's consider the first case. Then the sum simplifies to $\sum_{i \neq j} \Gamma^k_{i,j} \Gamma^{k'}_{i,j} \sigma^4$. Let's consider the second case. Then the sum simplifies to $\sum_{i \neq j} \Gamma^k_{i,j} \Gamma^{k'}_{j,i} \sigma^4$ providing us with the desired expression:

$$E[u_t^\top \Gamma^{(k)} u_t u_t^\top \Gamma^{(k')} u_t] = \sigma^4 \sum_{i,j} \left(\Gamma^k_{i,j} \Gamma^{k'}_{j,i} + \Gamma^k_{i,j} \Gamma^{k'}_{i,j} \right)$$

D.1.4 Proof of Proposition 3

For simplicity, we proceed in the case without factors – the case with factors is similar. We have $H(\gamma) Y_t = u_t$ with $H(\gamma) \coloneqq I - B(\gamma)$. The moment (22) is $M \coloneqq \frac{1}{T} \sum_t u_t' G^{ND} u_t = 0$ where G^{ND} is the matrix G with the diagonal terms set to 0, and $G = B_\gamma (I - B)^{-1} = -H_\gamma H^{-1}$. We have $u_t = H Y_t = \tilde{H} \tilde{Y}_t$ with $\tilde{H} = H \Omega^{-1}$. So, the procedure in $\tilde{Y}_t$ space leads, at convergence, to the moment $\tilde{M} \coloneqq \frac{1}{T} \sum_t u_t' \tilde{G}^{ND} u_t = 0$, with

$$\tilde{G} = -\tilde{H}_\gamma \tilde{H}^{-1} = -H_\gamma \Omega^{-1} \Omega H^{-1} = G$$

Hence, $M = \tilde{M}$, and we obtain the same moments.

If the transformation changed the variance of the residuals, and we use the GLS weighted moment, the procedure is still invariant.

D.2 Link with the MLE

We organize this subsection as follows: first we show the equivalence with MLE in the absence of a factor model. This proves our result in Corollary 1. Second, we show lack of

identification of the MLE score equation, proving our claim that MLE fails to identify the parameter $\gamma^\star$ in a simple model where Network GIV identifies $\gamma^\star$ using alternative moment conditions.

Proposition 4. (Link with MLE) *Suppose the model with no factors and no lags,* $(I - B(\gamma)) Y_t = u_t$. *Define*

$$G(\gamma) \coloneqq B_\gamma(\gamma) A(\gamma), \tag{D.2}$$

and

$$\Gamma_{ij} = G_{ij} 1_{i \neq j} \tag{D.3}$$

so that Γ *gives the network impacts "removing self-impacts", as shown by the fact that* $\Gamma_{ii} = 0$*; and the inferred* $u_t(\gamma) \coloneqq B(\gamma) y_t$. *Assume that the* σ_{u_i} *are unknown and need to be estimated. With* L *the log likelihood, the MLE estimator corresponds to the following moment condition:*

$$L_\gamma = 0, \qquad L_\gamma = \sum_t u_t'(\gamma) \Gamma(\gamma) u_t(\gamma) \tag{D.4}$$

In addition, the log likelihood satisfies:[27]

$$\mathbb{E} L_{\gamma\gamma} = -2T \operatorname{tr}\left(\Gamma\Gamma' + \Gamma^2\right) \tag{D.6}$$

which gives the efficiency of the MLE estimator :

$$\sqrt{T}\left(\gamma_T^{e,MLE} - \gamma\right) \sim N(0, V^\gamma), \qquad V^\gamma = \left(\operatorname{tr}\left(\Gamma\Gamma' + \Gamma^2\right)\right)^{-1} \tag{D.7}$$

Proof of Proposition 4. We take the case without factors and controls first, so $y_t = A(\gamma) u_t$. We write $u_{it} = \sigma_{u_i} v_{it}$, where v_{it} has variance 1. Call $V_u = Diag\left(\sigma_{u_i}^2\right)$ and $W_u = V_u^{-1}$.

27. Very explicitly, with $G^k = B_{\gamma_k} A$ a $N \times N$ matrix,

$$-\mathbb{E} L_{\gamma_k \gamma_\ell} = -2T \operatorname{tr}\left(\Gamma^k \Gamma^{\ell\prime} + \Gamma^k \Gamma^\ell\right) \tag{D.5}$$

So, $u_t = (I - B)\, y_t$. The log likelihood L is given by:

$$\begin{aligned}\frac{1}{T}L &= -\frac{1}{2}\frac{1}{T}\sum_t y_t'\,(I-B)'\,W^u\,(I-B)\,y_t + \frac{1}{2}\ln\det\left((I-B)'\,W^u\,(I-B)\right) - \frac{1}{2}N\ln 2\pi\sigma_u^2 \\ &= -\frac{1}{2}\frac{1}{T}\sum_t y_t'\,(I-B)'\,W^u\,(I-B)\,y_t + \ln\det\left((I-B)\right) - \frac{1}{2}\ln\det W^u - \frac{1}{2}N\ln 2\pi\sigma_u^2\end{aligned}$$

As $W_u = \sum_i e_i e_i' \sigma_{u_i}^{-2}$, optimizing $\frac{1}{T}L$ over the variance $\sigma_{u_i}^{-2}$ gives

$$0 = \frac{2}{T}\frac{d}{d\sigma_{u_i}^{-2}}L = \frac{1}{T}\sum_t u_t' e_i e_i' u_t - \sigma_{u_i}^2$$

i.e. the MLE of variance is simply

$$\sigma_{u_i}^2 = \frac{1}{T}\sum_t u_{it}^2 \tag{D.8}$$

We observe that

$$\frac{d}{d\gamma}\,(I-B)\,y_t = -B_\gamma y_t = -B_\gamma A u_t = -G u_t$$

so

$$\frac{d}{d\gamma}\left(-\frac{1}{2}y_t'\,(I-B)'\,W^u\,(I-B)\,y_t\right) = \frac{1}{2}\left(u_t' W^u G u_t + u_t' W^u G u_t\right) = u_t' W^u G u_t$$

Also,

$$\frac{d}{d\gamma}\ln\det\,(I-B) = -\operatorname{tr}\left(B_\gamma\,(I-B)^{-1}\right) = -\operatorname{tr} G$$

so

$$\frac{1}{T}L_\gamma = \frac{1}{T}\sum_t u_t' W^u G u_t - \operatorname{tr} G \tag{D.9}$$

Now, write $u_t = u_t^{-i} + u_{it} e_i$, where $u_t^{-i} = \left(u_{jt} 1_{j\neq i}\right)_{j=1\ldots N}$. Then,

$$\begin{aligned}u_t' W^u G u_t &= u_t'\left(\sum_i e_i e_i' \sigma_{u_i}^{-2}\right) G u_t = \sum_i u_t' e_i e_i' \sigma_{u_i}^{-2} G\left(u_t^{-i} + u_{it} e_i\right) \\ &= \sum_i u_{it}\sigma_{u_i}^{-2} e_i' G u_t^{-i} + \sum_i u_{it}\sigma_{u_i}^{-2} e_i' G u_{it} e_i \\ &= \sum_i u_{it}\sigma_{u_i}^{-2} e_i' G u_t^{-i} + \sum_i G_{ii} u_{it}^2 \sigma_{u_i}^{-2}\end{aligned}$$

Hence, calling $z_{it} = \sigma_{u_i}^{-1} e_i' G u_t^{-i}$ and $\tilde{v}_{it} = \sigma_{u_i}^{-1} u_{it}$

$$\begin{aligned}
-\frac{1}{T}\sum_t u_t' W^u G u_t + \operatorname{tr} G &= \frac{1}{T}\sum_t \sum_i u_{it}\sigma_{u_i}^{-2} e_i' G u_t^{-i} - \sum_i G_{ii}\left(\frac{1}{T}\sum_t u_{it}^2 \sigma_{u_i}^{-2}\right) + \operatorname{tr} G \\
&= \frac{1}{T}\sum_{t,i} \tilde{v}_{it} z_{it} - \sum_i G_{ii} + \operatorname{tr} G \\
&= \frac{1}{T}\sum_{t,i} \tilde{v}_{it} z_{it}
\end{aligned}$$

So, we have

$$\frac{1}{T} L_\gamma = \frac{1}{T}\sum_{t,i} \tilde{v}_{it} z_{it} \tag{D.10}$$

which is the moment in Algorithm 1, provided we normalize by $\frac{1}{\sigma_{u_i}}$.

Now, let us study the higher order terms. Calling Γ the G matrix without diagonal terms,

$$\Gamma_{ij} = G_{ij} 1_{i \neq j} \tag{D.11}$$

we have

$$\frac{1}{T} L_\gamma = \frac{1}{T}\sum_i u_t' W_u \Gamma u_t \tag{D.12}$$

So we have

$$\mathbb{E}\frac{1}{T} L_{\gamma, W_{ii}^u} = \mathbb{E} u_{it} e_i' \Gamma u_t = 0$$

Hence, to calculate the variance of the estimator γ, we just need to consider $L_{\gamma\gamma}$. As $u_t = (I - B)\, y_t$, we have

$$\frac{d}{d\gamma} u_t = -B_\gamma y_t = -B_\gamma A y_t = -G y_t$$

hence

$$\begin{aligned}
\mathbb{E}\frac{1}{T} L_{\gamma\gamma} &= \mathbb{E}\left[\frac{d}{d\gamma} u_t' W^u \Gamma u_t + u_t' W^u \Gamma \frac{d}{d\gamma} u_t + u_t' W^u \Gamma_\gamma u_t\right] \\
&= -\mathbb{E}\left[u_t' G' W^u \Gamma u_t + u_t' W^u \Gamma G u_t\right] + \operatorname{tr}\left(V^u W^u \Gamma_\gamma\right)
\end{aligned}$$

As $\mathrm{tr}\left(\Gamma\left(\gamma\right)\right)=0$ for all γ, $\mathrm{tr}\,\Gamma_{\gamma}=0$. So, we have

$$-\mathbb{E}\frac{1}{T}L_{\gamma\gamma}=\mathrm{tr}\left(V^{u}G'W^{u}\Gamma+V^{u}W^{u}\Gamma G\right)=\mathrm{tr}\left(V^{u}G'W^{u}\Gamma+\Gamma G\right)$$

But, as $\Gamma_{ii}=0$, we have

$$\mathrm{tr}\left(V^{u}G'W^{u}\Gamma\right)=\sum_{i,j}V_{ii}^{u}G'_{ij}W_{jj}^{u}\Gamma_{ji}=\sum_{i,j}V_{ii}^{u}\Gamma'_{ij}W_{jj}^{u}\Gamma_{ji}=\mathrm{tr}\left(\Gamma'\Gamma\right)$$

as the term $G'_{ii}\Gamma_{ii}=0$. Likewise

$$\mathrm{tr}\left(\Gamma G\right)=\sum_{ij}\Gamma_{ij}G_{ji}=\sum_{ij}\Gamma_{ij}\Gamma_{ji}=\mathrm{tr}\left(\Gamma^{2}\right)$$

So,

$$-\mathbb{E}\frac{1}{T}L_{\gamma\gamma}=\mathrm{tr}\left(\Gamma\Gamma'+\Gamma^{2}\right) \tag{D.13}$$

which gives the efficiency of the MLE estimator, $Tvar\left(\gamma^{e}-\gamma\right)\sim\left(-\frac{1}{T}\mathbb{E}L_{\gamma\gamma}\right)^{-1}$. □

D.2.1 Example where MLE lacks identification

In the next lines we provide a simple example illustrating that, however, in simple yet more complex models with simple factor structures as for example those formally studied below in SectionD.3, our procedure identifies the parameter of interest whereas MLE might fail. The main reason for this failure is that MLE will try to recover both the factors and the parameters of interest. However, the factor may not be identified, whereas the parameter of interest $\gamma^{\star}$ might.

One simple example proceeds as follows: consider a model of the form

$$Y_{t}=\gamma^{\star}\sum_{j}S_{j}Y_{j,t}+\eta_{t}+u_{it},$$

with $\sum_{j}S_{j}=1$, and where for the purpose of this example we consider η_{t} as a fixed effect and u_{it} having unit variance (note that if η_{t} is a random effect, MLE must model its variance covariance matrix which can be difficult in practice; we therefore consider the comparison with a fixed effect interpretation of η_{t} as a natural benchmark). It is easy to show that in

this case the MLE estimates

$$\eta_t^e = \frac{1}{N}\sum_i Y_{it} - \gamma^\star \sum_j Y_{jt} S_j.$$

which implies that the first order conditions for $\gamma^\star$ are not data dependent, as these reads as $\sum_i (Y_{i,t} - \bar{Y}_t) \sum_j S_j Y_{jt} - \frac{\frac{1}{2}\log\det(I-B(\gamma))^\top (I-B(\gamma))}{\partial\gamma}$ where the first component equals zero by construction. That is, the MLE fails in this context to estimate $\gamma^\star$ due to the lack of identification of η_t. The key difference with our proposed approach instead is to consider a moment condition of the form $E[(Y_{it} - B(\gamma)Y_t) z_{i,t}(\gamma)] = 0$ which preserves validity irrespective of the value of η_t, and leverages the fact that we do not need to estimate the fixed effects for estimating $\gamma^\star$.

D.3 The model with factors

In this section we extend our discussion in the presence of time-varying factors.

Assumption 4. *[Model with factors] We make Assumption 1, except that* $Y_t = B(\gamma^\star)Y_t + \lambda\eta_t + u_t$*, with* $\mathbb{E}\left[u_{it} | \{u_{jt}\}_{j\neq i}, \eta_t\right] = 0$*. Here,* $\lambda \in \mathbb{R}^N$ *is a fixed vector of loadings and* η_t *is a random vector of factors. The* $\eta_t, \{u_{i,t}\}_{i=1}^N$ *are independent random variables over time* t*.*

Assumption 5. *[Known* λ*] Assume that* λ *is observed.*

Assumption 5 holds in settings where the vector of loadings is *proportional* to some individual-level covariates X_i as in Gabaix and Koijen (2024). Section D.3.3 discusses estimation in settings where λ is unobserved. One could directly extend to settings with estimated λ, in which case the standard errors should be adjusted accordingly using, e.g., inferential theory from Bai (2003) (in practice, when λ must be estimated, estimators admit a smooth representation and therefore the variance could be estimated via e.g., the bootstrap). Here, η_t can be unobserved and assumed to be independent over time for simplicity, although one could relax the inferential results below to allow for weakly dependent factors.

The core idea is that we can orthogonalize the regression equation with respect to the presence of factors. We discuss this in the following lines.

We construct a set of instruments that take the following form

$$z_{i,t}^{\lambda}(\gamma) = B_{\gamma}(\gamma) y_{t|-i}^{\lambda}(\gamma), \quad y_{t|-i}^{\lambda}(\gamma) = A(\gamma) u_t^{-i}(\gamma), \qquad u_t^{-i}(\gamma) = \tilde{Q}_i^{\lambda}(\mathbb{I} - e_i e_i^{\top})(\mathbb{I} - B(\gamma)) Y_t. \tag{D.14}$$

The interpretation is the following. u_t^{-i} is the vector of residuals obtained by running the factor model $Y_t - B(\gamma^{\star})Y_t = \lambda \eta_t + u_t$ over the units $j \neq i$, which (at the correct γ) avoids a contamination by u_{it}, after we remove the contribution from all u_{it} from the interactive components $\eta^{-i}\lambda_t$, where η^{-i} denote the vector of loadings of all units except unit i. Then, we construct the instrument $z_{i,t}^{\lambda}(\gamma)$, similarly as in the case with no factors. The intuition behind this procedure is simple. As before, we approximate

$$Y_t - B(\gamma^{\star})Y_t = Y_t - B(\gamma)Y_t - B_{\gamma}(\gamma)Y_t + u_t + \lambda \eta_t + O(||\Delta\gamma||^2)$$

We then replace $B_{\gamma}(\gamma)Y_t$ with a prediction $E[B_{\gamma}(\gamma)Y_t | u_{jt\, j\neq i}]$. In particular, we replace $E[B_{\gamma}(\gamma)Y_t | u_{jt\, j\neq i}]$ with the simple expression for $z_{i,t}^{\lambda}(\gamma)$, which assumes that $E[\eta_t] = 0$. Note that this restriction on the factors is not necessary for validity, but it simplifies the estimation procedure. We show below validity of the procedure.

Proposition 5. *Suppose that Assumption 4, 5 hold. Then for any $i \in \{1, \cdots, N\}, t \in \{1, \cdots, T\}$,*

$$\mathbb{E}\left[\left(Y_{i,t} - e_i^{\top} B(\gamma^{\star}) Y_t\right) z_{i,t}^{\lambda}(\gamma^{\star})\right] = \mathbf{0}.$$

Proof. We can write $\tilde{Q}_i^{\lambda}(\mathbb{I} - e_i e_i^{\top})(\mathbb{I} - B(\gamma^{\star}))Y_t = \tilde{Q}_i^{\lambda}(\mathbb{I} - e_i e_i^{\top})(\lambda_{-i}\eta_t + \lambda_i \eta_t e_i + u_t) = \tilde{Q}_i^{\lambda}(u_t - u_{i,t} e_i)$. The proof is immediate since $u_{i,t}$ is mean independent of the idiosyncratic shocks of the other units. □

The proposition shows that $z_{i,t}^{\lambda}(\gamma^{\star})$ provides us with a valid instrument. The main advantage of this expression is that for estimation we do not need to know or estimate the factors η_t, since the matrix Q_i in practice removes the contribution of the factor by orthogonalizing the expression with respect to the (known) loadings. We proceed with the iterative estimation procedure outlined in Algorithm 1. The corresponding moment conditions equals

$$\bar{H}_i^{\lambda}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} \left(Y_{i,t} - e_i^{\top} B(\gamma) Y_t\right) e_i^{\top} \frac{\partial B(\gamma)}{\partial \gamma} y_{t|-i}^{\lambda}(\gamma), \qquad \bar{H}_i^{\lambda}(\gamma) \in \mathbb{R}^p, \qquad p = \dim \gamma. \tag{D.15}$$

D.3.1 Estimation and inference

In this section, we present our results for inference on the moment estimator that solves for $\hat{\gamma}^\lambda \in \mathcal{S} \subset \mathbb{R}^p$ the equation in $\mathbb{R}^p$: for given weights w_i that are not data dependent (e.g., $w_i = 1$), we set

$$\sum_{i=1}^{N} w_i \bar{H}_i^\lambda(\hat{\gamma}^\lambda) = \mathbf{0}_p \tag{D.16}$$

To express the variance of the estimator, we define:

$$\Delta_i^\lambda = -\mathbb{E}\Big[\frac{\partial \bar{H}_i^\lambda(\gamma)}{\partial \gamma}\Big|_{\gamma=\gamma^\star}\Big], \quad N\bar{\Delta}_w^\lambda = \sum_{i=1}^{N} w_i \Delta_i^\lambda \tag{D.17}$$

Assumption 6 (Local identification)**.** *Suppose that there exist a unique $\gamma^\star \in \mathcal{S}$ in the interior of $\mathcal{S}$, with $\mathcal{S}$ denoting a compact space, such that $\sum_{i=1}^N w_i \mathbb{E}[\bar{H}_i(\gamma^\star)] = \mathbf{0}$ and $N\bar{\Delta}_w^\lambda$ is an invertible matrix.*

Assumption 6 imposes local identification of the parameter of interest (local with respect to the pre-specified set $\mathcal{S}$).

Assumption 7 (Moment restrictions)**.** *Assume that for each i, $H_{i,t}^\lambda(\gamma)$ is continuously differentiable almost surely, $\mathbb{E}\left[\sup_\gamma ||H_{i,t}^\lambda(\gamma)||_\infty\right], \mathbb{E}\left[\sup_\gamma \left|\left|\frac{\partial H_{i,t}^\lambda(\gamma)}{\partial \gamma}\right|\right|_\infty\right] < \infty$, $\mathbb{E}\left[\sup_\gamma \left|\left|\frac{\partial^2 H_{i,t}(\gamma)}{\partial \gamma^{(p_1)} \partial \gamma^{(p_2)}}\right|\right|_\infty\right] < \infty$, $\mathbb{E}\Big[\Big(\frac{1}{N}\sum_{i=1}^N w_i H_{i,t}^\lambda(\gamma^\star)\Big)\Big(\frac{1}{N}\sum_{i=1}^N w_i H_{i,t}^\lambda(\gamma^\star)\Big)^\top\Big]$ is positive definite, and $\frac{1}{N}\sum_{i=1}^N w_i H_{i,t}^\lambda(\gamma^\star)$ has finite third moment.*

Proposition 6. *Suppose that Assumptions 4, 5, 6 and 7 hold. Define $\Sigma_w^\lambda = \mathbb{E}\Big[\Big(\sum_{i=1}^N w_i H_{i,t}^\lambda(\gamma^\star)\Big)\Big(\sum_{i=1}^N w_i H_{i,t}^\lambda(\gamma^\star)\Big)^\top\Big]$. Then as $T \to \infty$,*

$$\sqrt{T}\Big(V_w^\lambda(\gamma^\star)\Big)^{-1/2}\Big(\hat{\gamma}^\lambda - \gamma^\star\Big) \to_d \mathcal{N}(0, \mathbb{I}), \quad \text{where} \quad V_w^\lambda(\gamma^\star) = \Big(N\bar{\Delta}_w^\lambda\Big)^{-1}\Sigma_w^\lambda\Big(N\bar{\Delta}_w^\lambda\Big)^{-1\top}.$$

The proof follows directly by arguments in Newey and McFadden (1994), using the fact that $u_{i,t}$ are independent over time. We can estimate Σ_w^λ using the empirical covariance of the moment restrictions, and estimate $N\bar{\Delta}_w^\lambda$ using any numerical procedure to estimate the derivative of the moment function. Consistency of the estimated variance follows directly from the law of large numbers as $T \to \infty$ under Assumption 7.

D.3.2 Variance characterization in the presence of loadings

In the following proposition, we provide an expression for the variance and of the gradient, choosing $w_i = 1$, for simplicity. Define

$$\tilde{u}^Q_{i,t} = \tilde{Q}_i[u_t - u_{i,t}e_i], \quad \tilde{\Gamma}^1_{i,j} = \left[\Gamma\tilde{Q}_i\right]_{i,j,1}, \quad \tilde{\Gamma}_{i,j,2} = \left[\tilde{Q}_i\Gamma\right]_{j,i} \quad v^\lambda_t = \lambda\eta_t + u_t.$$

Proposition 7 (Variance)**.** *Let Assumptions 4, 5, 6, 7 hold and suppose that* $\mathbb{E}[u^2_{i,t}] = \sigma^2$*. Then*

$$N\bar{\Delta}^{\lambda,(p_1,p_2)}_{w=1} = \sigma^2\left[\mathrm{tr}\Big(\Gamma^{(p_1)\top}\tilde{\Gamma}^{1(p_2)}\Big) + \mathrm{tr}\Big(\Gamma^{(p_1)\top}\tilde{\Gamma}^{2(p_2)}\Big)\right],$$
$$\Sigma^{\lambda,(p_1,p_2)}_{w=1} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\Big(v^{\lambda\top}_t G^{(p_1)}\tilde{u}^Q_t\Big)\Big(v^{\lambda\top}_t G^{(p_2)}\tilde{u}^Q_t\Big)^\top\right].$$

Proof. See Appendix D.7.1. □

The variance in Proposition 7 follows similarly to the case with no factors with some minor modifications to account for the dependence induced by the factors. Variance estimation follows similarly to typical GMM procedures using a sandwich estimator. We can estimate $\Sigma^{\lambda,(p_1,p_2)}$ using the empirical covariance of the moment restrictions, and estimate $N\bar{\Delta}^\lambda$ using any numerical procedure to estimate the derivative of the moment function. Consistency of the estimated variance follows directly from the law of large numbers as $T \to \infty$ under Assumption 7.

It is also possible that factors exhibit (weak) serial correlation, in which case practitioners can replace $\Sigma^{\lambda,(p_1,p_2)}$ with the long-run variance-covariance matrix which can be estimated using standard methods in the time-series literature (see e.g., White (2014)).

D.3.3 Complete algorithm with unobserved loadings

In this section, we turn to the description of the complete algorithm, by integrating all variants. Specifically, we consider generalized heteroskedasticity, lagged outcomes, unobserved loadings, and auto-correlated shocks. We consider a model of the form

$$Y_t = B_0(\gamma)Y_t + B_1(\gamma)Y_{t-1} + \chi(\gamma)u_t + \lambda\eta_t, \tag{D.18}$$

where $\chi(\gamma)$ is a diagonal (invertible) matrix, $B_0(\cdot), B_1(\cdot), \chi(\cdot)$ are known functions, and u_t have uncorrelated entries with arbitrary heteroskedasticity. We further let $u_t = \rho u_{t-1} + \varepsilon_t$, where ε_t are uncorrelated over time. The first step of the algorithm is described in Algorithm 1 and proceeds as follows. It takes one guess of the parameters (ρ, γ) and takes as an input the number of factors r. For simplicity, suppose that the initial guess sets $\rho = 0$. The algorithm then constructs residuals $\tilde{v}_t$ in Algorithm 1 after removing the dependence with the component $\chi(\gamma)$ (the heteroskedasticity component that depends on γ). It then estimates the loadings using an iterative principal component algorithm (PCA). The PCA algorithm allows for heteroskedasticity in the variance of the residuals and is formally described in Algorithm 5. We construct an *estimator* of the loadings and an estimated variance of the idiosyncratic shocks. We then construct an instrument $\tilde{y}_{t|-i}$ using the estimated loadings and the guessed value of γ and impose the orthogonal moment condition.

When constructing the orthogonal moment conditions, we operate with a model that divides each outcome by the standard deviation of its corresponding idiosyncratic shock to improve efficiency, and we reweight each observation by given weights w_i. We use $w_i = 1$ as the default, but other weights are possible (e.g., based on efficient GMM formula). Also, note that the algorithm can directly work if we also incorporate covariates of the form $Y_t = B_0(\gamma)Y_t + B_1(\gamma)Y_{t-1} + X_t\beta + \chi(\gamma)u_t + \lambda\eta_t$, by controlling for both the covariates in the regression steps of the algorithm. Finally, in the last computation of Algorithm 1, we regress the residuals obtained from the previous regression (and evaluated at the $\hat{\gamma}$ from the previous iteration), the component orthogonal to the residuals from our moment condition. If the current $\hat{\gamma}$ was close to the true parameter of interest, this regression should have coefficient zero. When instead the parameter estimated in the previous iteration is different but in a neighborhood near $\gamma^\star$, we would expect this to be captured by the estimated coefficient from a simple Taylor approximation.

Remark 1. (Variance with estimated loadings). In the presence of estimated factors, the variance of the estimator may be affected by the estimation error of the loadings. Whenever estimation of the factors admits a standard central limit theorem, see e.g., Bai (2003), one could use the bootstrap method to approximate the distribution of the estimated parameter. We also note that there are edge cases where estimation of the factors may not affect the

asymptotic distribution.[28]

D.4 Plots to visualize the identification

We propose a way to (partially) visualize the outcomes in Algorithm 1. The following is for one given component of the γ vector, say component k.

1. *First stage.* We define $z_{i,t}^{\text{No Instr.}}$ (as in right-hand side) to be the value of $z_{i,t}$ in (12), but replacing $\tilde{y}_{t|-i}(\hat{\gamma})$ by the actual value Y_t:[29] this is the "non-instrumented" value of the right-and side. We do that only for the k-th component, in this first stage, and in the second stage. We plot $z_{i,t}$ on the x axis vs $z_{i,t}^{\text{No Instr.}}$ on the y axis. We expect that the slope is 1, and we report it.

2. *Second stage.* We pick a "default" value γ^d. For instance, if we want to visualize component 1, we take γ^d defined by $\gamma_\ell^d \coloneqq \gamma_\ell^e 1_{k\neq\ell} + \gamma_k^D 1_{k=\ell}$, where γ^e is our final estimate of γ, and γ_1^D is a default value (e.g. corresponding to 0 elasticity). We form the instrument $z_{it}^{(k)}$ based on the final (obtained at convergence) value of γ^e (rather than γ^d). We define $v_t \coloneqq y_t - B_0\left(\gamma^d\right) y_t - B_1\left(\gamma^d\right) y_{t-1}$ and plot the regression:

$$v_{it} = \Delta\gamma_k \cdot z_{i,t}^{(k)} + \text{error}_{it}$$

   On the left-hand side we have the values at the "default" value of γ, and on the right-hand side, we see how the instrument dictates a deviation $\gamma - \gamma^d$ from the default. The

28. Using a first-order Taylor approximation, it is easy to show that under mild conditions, we can approximate our moment condition (choosing $w_i = 1$ for expositional convenience only)

$$0 = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} H_{i,t}^{\hat{\lambda}}(\hat{\gamma}) = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\Big(H_{i,t}^{\lambda}(\gamma^\star) + \frac{\partial H_{i,t}^{\lambda}(\gamma)}{\partial\gamma}\Big|_{\gamma=\gamma^\star}(\hat{\gamma}-\gamma^\star) + \frac{\partial H_{i,t}^{\lambda'}(\gamma^\star)}{\partial\lambda'}\Big|_{\lambda'\in[\hat{\lambda},\lambda]}(\hat{\lambda}-\lambda) + O_p\Big(||\hat{\gamma}-\gamma^\star||_2^2\Big)\Big).$$

Observe that we can write $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial H_{i,t}^{\lambda}(\gamma^\star)}{\partial\lambda} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^\top\eta_t f(\lambda_{-i}) + o_p(1)$ for some function $f(\cdot)$ which may depend on $\gamma^\star, \eta_t$. Therefore, whenever loadings are asymptotically independent and centered around zero after appropriately recentering by controlling for the loading of ones (e.g., different industries react to aggregate shocks with approximately independent industry specific effects after controlling for covariates), we may expect that $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial H_{i,t}^{\lambda}(\gamma^\star)}{\partial\lambda} = O_p(N^{-v})$, for some $v > 0$ as both N, T grow. This implies that the error for estimating the loadings is asymptotically negligible for the estimated $\hat{\gamma}$, whenever the rate of convergence of $\hat{\gamma}$ is slower than $N^{-v}T^{-1/2}$. However, this does not necessarily occur if the Jacobian with respect to the loadings is non-vanishing asymptotically.

29. This is, $z_{i,t}^{(k)} \coloneqq e_i^\top\left(\frac{\partial B_0}{\partial\gamma_k}Y_t + \frac{\partial B_1}{\partial\gamma_k}Y_{t-1}\right)$.

"second stage" graph plots v_{it} (on the y axis) vs z_{it} (on the x axis), and the slope is $\left(\gamma_1 - \gamma_1^D\right)$. We can also plot the following.

3. How good is $\tilde{y}_{t|-i}$ as an instrument for Y_{it}: plot Y_{it} (y axis) vs $\tilde{y}_{t|-i}$ (x axis). At convergence, the slope should be 1.

4. Outlier detection. At the convergence values, we plot v_{it} (y axis) vs $z_{it}^{(k)}$ (x axis). The slope should be 0, but we may detect some outliers.

D.5 Non-linear version of the model

The basic model is linear in u_t, Y_t, though nonlinear in γ. We show that the model generalizes well to shocks entering non-linearly in u_t, y_t, as in:

$$Y_t = A\left(u_t + \lambda\eta_t, \gamma\right) \tag{D.19}$$

where $A\left(\cdot, \gamma\right) : \mathbb{R}^N \to \mathbb{R}^N$ is a general smooth function, *which could be nonlinear*. It's parametrized by γ.

Call $b\left(y, \gamma\right) = A^{-1}\left(y, \gamma\right)$, where the inversion in y, while γ is just a parameter. In our linear case $A\left(u, \gamma\right) = \left(I - B\left(\gamma\right)\right)^{-1} u$ so $b\left(y, \gamma\right) = y - B\left(\gamma\right) y_t$, and $b_\gamma = -B_\gamma$.

We show how our linear logic generalizes naturally. Indeed, let us following the same derivation. Suppose that we have $\gamma = \gamma^* - \gamma^\delta$, for a small γ^δ. We have

$$b\left(y_t, \gamma^*\right) = u_t$$

so

$$\begin{aligned} b\left(y_t, \gamma\right) &= b\left(y_t, \gamma^* - \gamma^\delta\right) = b\left(y_t, \gamma^*\right) - \gamma^\delta \cdot b_\gamma\left(y_t, \gamma^*\right) + O\left(\left\|\gamma^\delta\right\|^2\right) \\ &= -\gamma^\delta \cdot b_\gamma\left(y_t, \gamma^*\right) + u_t + O\left(\left\|\gamma^\delta\right\|^2\right) \end{aligned}$$

Hence, the instrument we take is as follows. We form $u_t^{-i} \coloneqq \left(I - e_i e_i^\top\right) b\left(y_t, \gamma\right)$ (where Q^{e_i} simply sets to 0 the i-th component), and

$$z_{it} = -e_i' b_\gamma\left(A\left(u_t^{-i}, \gamma\right), \gamma\right) \tag{D.20}$$

At convergence z_{it} depends only on u_{jt} for $j \neq i$, so is uncorrelated with $b_i(y_t, \gamma^*) = u_{it}$.

So, we regress

$$b_{it}(y_t, \gamma) = \gamma^\delta \cdot z_{it} + u^e_{it}$$

and we obtain γ^δ. We iterate $\gamma' \coloneqq \gamma + \gamma^\delta$, and do that again, until convergence.

At the true value γ^* we do have $\gamma^\delta = 0$, i.e.

$$\mathbb{E}[b_{it}(y_t, \gamma^*) z_{it}] = 0. \tag{D.21}$$

A different way to say the same thing is the following. We construct the counterfactual economy with just idiosyncratic shock:

$$Y_t^{-i}(\gamma) = b^{-1}\left(u_t^{-i}, \gamma\right) \tag{D.22}$$

and the residual for component i, given a potentially different $\gamma + \Delta\gamma$

$$u^e_{it}(\Delta\gamma) \coloneqq u^e_{it}\left(Y_t^{-i}(\gamma), \gamma + \Delta\gamma\right) \coloneqq e_i' b\left(Y_t^{-i}(\gamma), \gamma + \Delta\gamma\right) \tag{D.23}$$

Then, the moment is

$$\mathbb{E}[u_{it} \partial_{\Delta\gamma} u^e_{it}(\Delta\gamma)] = 0. \tag{D.24}$$

An estimate of $\Delta\gamma$ comes from the regression

$$u_{it} = -\Delta\gamma \cdot \partial_{\Delta\gamma} u^e_{it}(\Delta\gamma) + \varepsilon_{it} \tag{D.25}$$

potentially in the GLS form with weight $\sigma^{-2}_{u_i}$. The empirical moment is then

$$\sum_i \sigma^{-2}_{u_i} u_{it} \partial_{\Delta\gamma} u^e_{it} = 0. \tag{D.26}$$

D.6 Model misspecification

Like most procedures, there are problems if the model is misspecified. When our $B(\gamma)$ does not span the correct matrix B^* (e.g., because it would require more parameters than allowed by our low-dimensional γ), then the chosen γ is, morally, so that $B(\gamma)$ is a "best

approximation" of B^*.[30]

More optimistically, if $B(\gamma)$ does span the correct matrix B^*, then γ may be biased, but the impulse response $B(\gamma)$ will still be correct.[31] In sum, the predictions about the impulse-responses ($B(\gamma)$) can be robust (and are completely robust in the case we mentioned), even when the details of the model of γ are partially wrong.

D.7 Omitted proofs for results in the online Appendix

D.7.1 Proof of Proposition 7

Denote $\sigma_{i,t}^2 = \mathbb{E}[u_{i,t}^2]$, and $H_{i,t}(\gamma) = \left(Y_{i,t} - e_i^\top B(\gamma) Y_t\right) e_i^\top \frac{\partial B(g)}{\partial g} A(\gamma) y_{t|-i}^\lambda(\gamma)$, where $y_{t|-i}^\lambda(\gamma) = \tilde{Q}_i(I - e_i e_i^\top)(I - B(\gamma))Y_t$. Also, recall the definition of $G^{(k)} = \frac{\partial B(g)}{\partial g_k}\Big|_{g=\gamma^\star} A(\gamma^\star)$.

First observe that because under Assumption 4, $Y_t = A(\gamma^\star) v_t^\lambda, \quad v_t^\lambda = u_t + \lambda \eta_t$, we can write

$$\frac{\partial\left(Y_{i,t} - e_i^\top B(\gamma) Y_t\right)}{\partial \gamma_k} = -e_i^\top \frac{\partial B(\gamma)}{\partial \gamma_k} A(\gamma) v_t^\lambda,$$
$$\frac{\partial A(\gamma) y_{t|-i}^\lambda(\gamma)}{\partial \gamma_k} = \frac{\partial A(\gamma)}{\partial \gamma_k} y_{t|-i}^\lambda(\gamma) - A(\gamma)\tilde{Q}_i(\mathbb{I} - e_i e_i^\top)\frac{\partial B(\gamma)}{\partial \gamma_k} A(\gamma) v_t^\lambda.$$

Consider the component Δ_{p_1,p_2}^λ first. Because $E[\eta_t y_{t|-i}^\lambda(\gamma^\star)] = 0$ by construction, by simplifying the remaining terms (including those that depend on the second derivative of $B(\gamma)$), we can write

$$-\frac{\partial \mathbb{E}[H_{i,t}^{(p_1)}(\gamma)]}{\partial \gamma_{p_2}}\Big|_{\gamma=\gamma^\star} = \sum_{j\neq i} G_{j,i}^{(p_2)} [G^{(p_1)} \tilde{Q}_i]_{i,j} \sigma_{j,t}^2 + \sum_{j\neq i} G_{i,j}^{(p_1)} [\tilde{Q}_i G^{(p_2)}]_{j,i} \sigma_{i,t}^2$$

30. For instance, in the basic case without factor, where $Y_t = (I - B)^{-1} u_t$, the predicted variance of is $V^y(\gamma) = (I - B(\gamma))^{-1} V^u (I - B'(\gamma))^{-1}$. Call $V^{y,*}$ the true correct variance of Y_t, and $D(G, H) = \text{tr}\left(G^{-1}H\right) - \ln\left|\det\left(G^{-1}H\right)\right| - N$ the Stein (1981) pseudo-distance between matrices G and H (which minimized when $G = H$). So our procedure, roughly speaking, tries to find γ and V^u (the diagonal matrix of variances of u), to minimize $D(V^y(\gamma), V^{y*})$. Indeed, calling V^{Ye} the empirical variance-covariance matrix, and assuming Gaussian shocks, the log-likelihood is (up to a constant), $\mathcal{L} = -TD\left(V^y(\gamma), V^{Ye}\right)$.

31. Indeed, suppose that the true model has $B^*(\gamma^*)$, but we misspecified the model as $B(\gamma)$. But suppose that the two models have the same "expressive power", i.e. one can write $B(\gamma) = B^*(\phi(\gamma))$ for some bijective function ϕ from $\mathbb{R}^p$ to $\mathbb{R}^p$, for instance $\phi(\gamma) = \gamma + b$. Then, the model will identify a coefficient γ such that $\phi(\gamma) = \gamma^*$, so in our example $\gamma = \gamma^* - b$. So, the coefficient γ will be biased. But the impulse-response $B(\gamma)$ will be exactly equal to $B^*(\gamma^*)$: so the matrix of impulse-response will be exactly identified.

The final expression follows from rearrangement and homoskedasticity. Finally, because

$$\sum_{i=1}^{N} \bar{H}_i(\gamma^\star) = v_t^\top G \tilde{u}_t^Q,$$

the proof completes.

E Complements to the Empirical Study of the Propagation of TFP in the US

Here we give complements to Section 3.

E.1 TFP Transmission simulated data results

In this section we replicate the results in Section 3 using simulated data. We perform 100 Monte Carlo replications of the TFP series generated under a fixed set of true parameters for $T = 50$ periods, and estimate the parameters in each replication. Table 2 shows the results for a model with two lags (the second lag is set to 0).

Table 2: TFP propagation: Monte Carlo simulation estimates, 2 lags

Parameter	True value	Mean $\hat{\gamma}$	Mean OLS SE	SD of $\hat{\gamma}$	RMSE	Mean$(\text{SE} - \text{SD of } \hat{\gamma})^2$
$\gamma_{\text{down}}^{(0)}$	0.30	0.44	0.10	0.20	0.25	0.01
$\gamma_{\text{up}}^{(0)}$	0.30	0.34	0.08	0.13	0.13	0.00
$\gamma_{\text{down}}^{(1)}$	0.21	0.18	0.10	0.13	0.13	0.00
$\gamma_{\text{up}}^{(1)}$	0.21	0.20	0.08	0.10	0.10	0.00
$\gamma_{\text{down}}^{(2)}$	0.00	-0.01	0.11	0.11	0.11	0.00
$\gamma_{\text{up}}^{(2)}$	0.00	0.00	0.08	0.09	0.09	0.00

This table reports Monte Carlo results for the TFP spillover coefficients based on 100 replications. The TFP series are generated from the two-lag model under the reported true parameter values for $T = 50$ periods. The OLS SE are the SE reported at the last iteration of the algorithm (Algorithm 1). The table also reports the mean estimated standard errors, the Monte Carlo standard deviation of $\hat{\gamma}$, $\mathbb{E}[(\text{SE} - \text{SD of } \hat{\gamma})^2]$ and RMSE $\mathbb{E}[(\hat{\gamma} - \gamma)^2]^{1/2}$. The estimation uses heteroskedastic PCA as in Algorithm 6. The data generating process simulates heteroskedastic shocks.

E.2 Propagation Dynamics

We consider the dynamics of propagation to the"long run". Decompose the impact operator as $\overline{\mathcal{I}}(L) = \sum_{h=0}^{\infty} \overline{\mathcal{I}}_h L^h$, so that

$$Y_t = \overline{\mathcal{I}}(L)\, u_t = \sum_{h=0}^{\infty} \overline{\mathcal{I}}_h u_{t-h} \tag{E.1}$$

Hence, the cumulative impact at horizon H is

$$\overline{\mathcal{I}}_{\leq H} \coloneqq \sum_{h=0}^{H} \overline{\mathcal{I}}_h = \frac{d}{du_t} \mathbb{E}_t \left[Y_t + \cdots + Y_{t+H} \right] \tag{E.2}$$

Hence, the total multiplier at horizon H is

$$M^{\text{tot}}(H) = \frac{\sum_{i,j} S_i \left(\overline{\mathcal{I}}_{\leq H} \right)_{ij}}{\sum_i S_i} \tag{E.3}$$

The self impact at horizon H is similarly:

$$M^{\text{self}}(H) = \frac{\sum_i S_i \left(\overline{\mathcal{I}}_{\leq H} \right)_{ii}}{\sum_i S_i}$$

We note that the true impact is

$$M^{\text{Full}} = M^{\text{Hulten}} M^{\text{tot}} \tag{E.4}$$

i.e. our newly estimated M^{tot} is multiplied by the "mechanical" Hulten effect, $M^{\text{Hulten}} \simeq 1.8$.

E.3 Further empirical results

Diagnostic plots We next present a few diagnostic plots.

Explaining aggregate TFP with idiosyncratic shocks Also, take u_{it}, our estimates of the primitive idiosyncratic shocks. Form

$$Z_t = \sum_i \frac{S_{i,t-1}}{Y_{t-1}} u_{it} \tag{E.5}$$

Table 3: TFP growth on idiosyncratic shocks

	(1)	(2)
b_0	1.17***	1.39***
	(0.16)	(0.04)
b_1	0.14	0.29***
	(0.16)	(0.04)
b_2	0.35**	0.34***
	(0.17)	(0.04)
b^{tot}	1.66***	2.01***
	(0.31)	(0.08)
Controls	No	Yes
R^2	0.51	0.97
N obs.	57	57

This table shows the regression results of (E.6). Controls are Domar-weighted common factors as in (E.7). $b^{\text{tot}} = \sum_{h=0}^{2} b_h$. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

where $\frac{S_{i,t-1}}{Y_{t-1}}$ is the Domar weight of sector i (Sales / GDP, so $\sum_i \frac{S_{i,t-1}}{Y_{t-1}} = M^{\text{Hulten}}$). We regress

$$\Delta \ln TFP_t = \sum_{h=0}^{2} b_h Z_{t-h} + \varepsilon_t \tag{E.6}$$

and report $b^{\text{tot}} = \sum_h b_h$. Then, we should have close to $b^{\text{tot}} \simeq M^{\text{tot}}$. The estimated b^{tot} is 1.7 while, as noted above, $M^{\text{tot}} = 2.7$ and $M^{\text{tot}}(2) = 2.2$. We also investigate the following regression

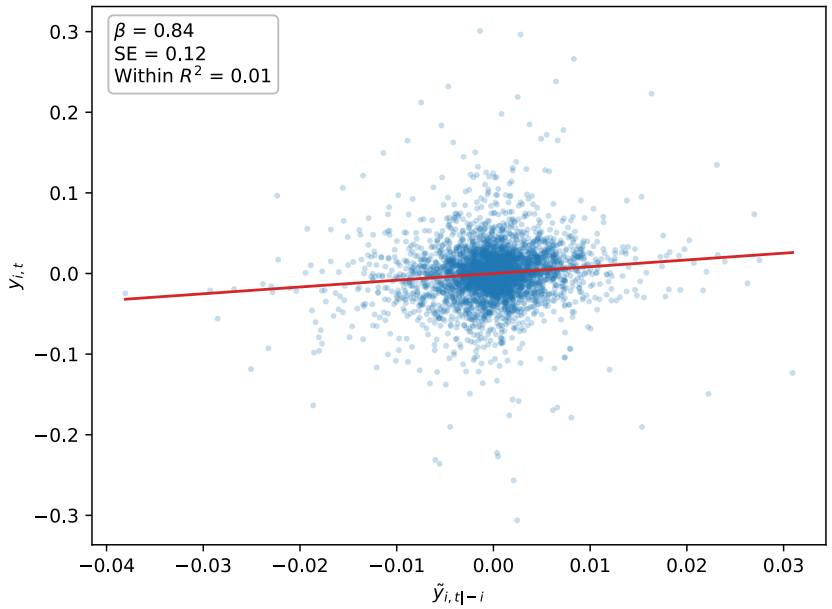
$$\Delta \ln TFP_t = \sum_{h=0}^{2} b_h Z_{t-h} + \sum_{h=0}^{2} c_h \lambda_S \eta_{t-h} + \varepsilon_t \tag{E.7}$$

where $\lambda_S := S'\lambda$.[32] In this case, the estimated b^{tot} is 2.0. The results are in Table 3. Reassuringly, the b^{tot} is a bit higher, consistent with the fact that a TFP shock induces extra effects, such as higher use or accumulation of factors.

This means that indeed we "pass the test": we find $b^{\text{tot}} \simeq M^{\text{tot}}$ to a good degree of approximation. The regression results are shown in Table 3.

32. This way, we avoid estimating a model with too many free parameters.

Figure 2: TFP growth and sector-level TFP shocks



This figure shows the scatterplot of TFP growth y_{it} of sector i vs $\tilde{y}_{i,t|-i}$, which is the TFP growth of sector i induced by the idiosyncratic shocks to sectors outside of i; together with the fitted line. $\tilde{y}_{i,t|-i}$ is the value of y_{it} that would come from idiosyncratic shocks, defined in Algorithm 1. The regression is $y_{it} = a_i + \beta\tilde{y}_{i,t|-i} + e_{it}$.

Explaining sector-level shock We plot y_{it} vs $\tilde{y}_{i,t|-i}$, which is the value of y_{it} that would come from idiosyncratic shocks, defined in Algorithm1 (it's the i'th component of the $\tilde{y}_{t|-i}$ vector).

The regression $y_{it} = a_i + \beta\tilde{y}_{i,t|-i} + e_{it}$ yields a β of 0.84 (S.E. 0.12). In the baseline model (with no measurement errors), we should have $\beta = 1$, and indeed we cannot reject that benchmark.

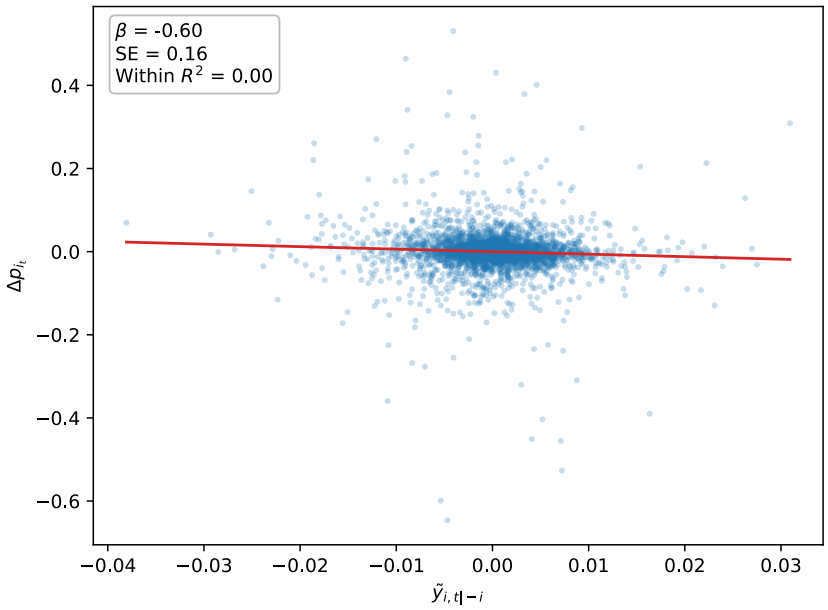
Figure 2 shows the scatterplot of y_{it} vs $\tilde{y}_{i,t|-i}$. Figure 3 shows the same kind of scatterplot but with price growth on the y-axis; this regression yields a β of -0.60. This is consistent with the view that instrumented TFP shocks to sector i induce a fall in prices.

Table 4: Variance decomposition of PCA-estimated common factors

Component	Eigenvalue	Explained variance ratio	Total variance
1	0.012	0.147	0.084
2	0.011	0.126	0.084
3	0.008	0.100	0.084

This table shows the variance decomposition of the three common factors estimated with heteroskedastic PCA as in Algorithm 6.

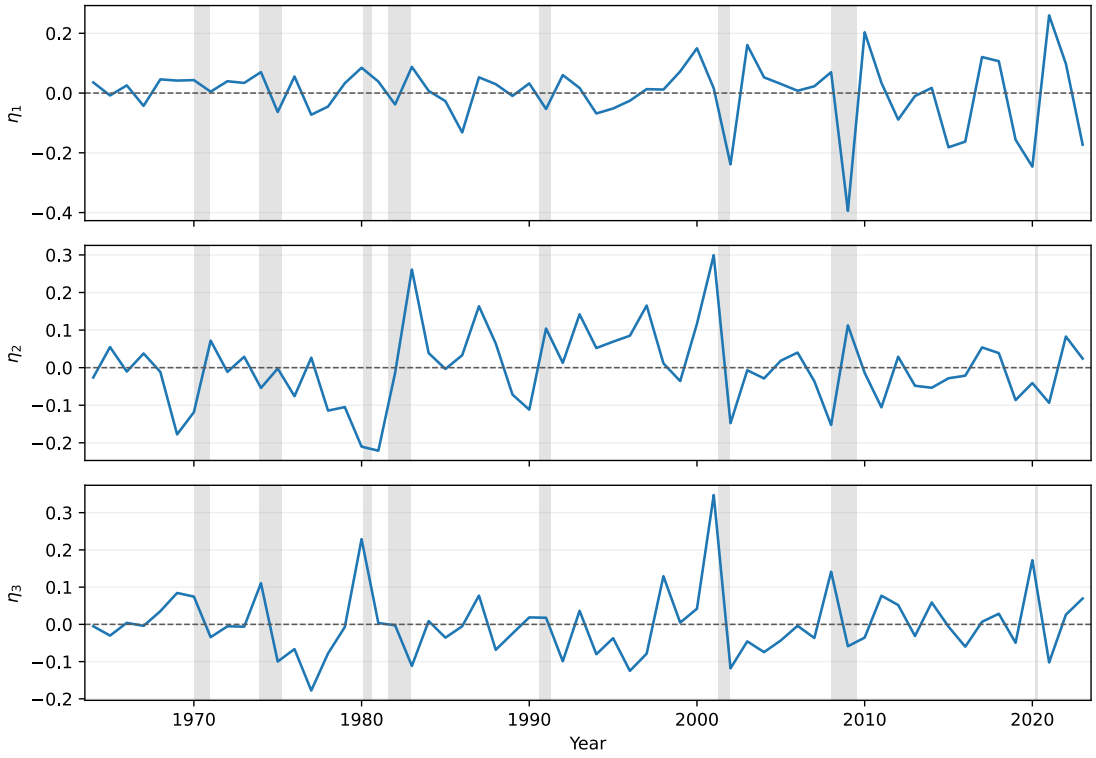
Figure 3: Price growth and sector-level shocks



This figure shows the scatterplot of sector-level price growth vs $\tilde{y}_{i,t|-i}$ together with the fitted line. $\tilde{y}_{i,t|-i}$ is the value of y_{it} that would come from idiosyncratic shocks, defined in Algorithm 1. The regression is $\Delta \ln P_{it} = a_i + \beta \tilde{y}_{i,t|-i} + e_{it}$.

Time-series for the aggregate shocks Figure 4 shows the time-series for the three factors η_t estimated with heteroskedastic PCA as in Algorithm 6. As always, time-series factors are hard to interpret. One can surmise that the first factor η_{1t} is a "2008 recession" factor, while the other two help fit the other major recessions, from the 1970s and early 1980s. They explain from 10 to 15 % of the variance (Table 4).

Figure 4: TFP propagation: PCA-estimated common factors η_t



This figure shows the time-series of the three PCA-estimated common factors η_t. The estimation used heteroskedastic PCA as in Algorithm 6. Shaded regions indicate NBER recession periods. Recession periods are measured monthly, while the estimated factors are annual.

F Monte-Carlo simulations

F.1 Simulated data results for the algorithms with missing data and measurement error

In this section we bring the missing outcomes and measurement error algorithm to a test using simulated data. We compare the results of Algorithm 3 and Algorithm 2.

In particular, we generate a latent panel of 60 periods of TFP growth starting from a set of true parameters, and then we independently generate random measurement error and random missing outcomes. For measurement error we set $\text{Var}(\varepsilon_{it}) = 0.10\text{Var}(Y^\star)$, but in general we can have any share of latent-outcome variance and even heteroskedastic measurement error. For missing outcomes we randomly mask 20% of observations as missing. We then estimate the two algorithms with a 100-replication Monte Carlo simulation.

It should be noted that Algorithm 2 is not designed to handle measurement error.

When using Algorithm 3, the E-step uses the analytic version.

Decomposing total error. Since we are simulating different sources of error, it is useful to adopt a procedure to identify the effects of each source on the final estimate. To do that, we decompose the Monte Carlo error of each estimator using paired counterfactual panels generated from the same simulation draw. For each replication b and parameter γ, we estimate γ on four versions of the same data: the latent complete panel, the noisy complete panel (i.e. the panel with measurement error), the latent panel after applying the missingness mask, and the final noisy masked panel. Denote the corresponding estimates by

$$\hat{\gamma}_b^{LC}, \qquad \hat{\gamma}_b^{NC}, \qquad \hat{\gamma}_b^{LM}, \qquad \hat{\gamma}_b^{NM},$$

where LC is latent-complete, NC is noisy-complete, LM is latent-missing, and NM is noisy-missing. This yields the decomposition

$$\underbrace{\mathbb{E}_b\left[\hat{\theta}_b^{NM} - \theta_0\right]}_{\text{total observed-panel error}} = \underbrace{\mathbb{E}_b\left[\hat{\theta}_b^{LC} - \theta_0\right]}_{\text{finite-sample error}} + \underbrace{\mathbb{E}_b\left[\hat{\theta}_b^{LM} - \hat{\theta}_b^{LC}\right]}_{\text{missing-outcome effect}} + \underbrace{\mathbb{E}_b\left[\hat{\theta}_b^{NM} - \hat{\theta}_b^{LM}\right]}_{\text{measurement-error effect given the mask}}$$

Because all terms are computed within the same replication, the comparisons remove simulation-draw variation and isolate the incremental effect of each data imperfection. The latent-complete estimate measures the error that remains even in the absence of missing outcomes and measurement error; the latent-masked minus latent-complete difference isolates the effect of the missing-outcome pattern holding the latent data fixed; and the noisy-masked minus latent-masked difference isolates the additional effect of measurement error on the already-masked panel.

Tables 6 and 5 show the results of the Monte Carlo simulations under Algorithms 3 and 2 respectively. In that sample at least, the sophisticated Algorithm3 and the much simpler Algorithm2 have quite similar performance.

Both algorithms display finite sample bias, especially for the estimates of $\gamma_{\text{down}}^{(0)}$ and $\gamma_{\text{up}}^{(0)}$; this can be observed already in Table 2. The relevant finding however is that Algorithm 3 produces smaller missing outcomes and measurement error bias, especially for the multiplier M^{tot}. For the first-pass algorithm instead it can be noted that the finite sample bias and

Table 5: Algorithm 2 (simple first-pass algorithm) results with simulated measurement error and missing outcomes

Parameter	True Value	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Mean OLS SE	RMSE	SD of $\hat{\gamma}$	Mean(SE − SD of $\hat{\gamma}$)^2
$\gamma^{(0)}_{\text{down}}$	0.30	0.40	0.44	0.14	0.24	0.22	0.01
$\gamma^{(0)}_{\text{up}}$	0.30	0.33	0.33	0.11	0.14	0.14	0.00
$\gamma^{(1)}_{\text{down}}$	0.21	0.20	0.21	0.14	0.16	0.16	0.00
$\gamma^{(1)}_{\text{up}}$	0.21	0.21	0.22	0.11	0.11	0.11	0.00
$\gamma^{(2)}_{\text{down}}$	0.00	-0.02	-0.03	0.14	0.15	0.15	0.00
$\gamma^{(2)}_{\text{up}}$	0.00	0.00	0.00	0.11	0.12	0.13	0.00
M^{tot}	1.77	2.09	1.97	0.45	0.70	0.63	0.09
M^{self}	1.09	1.11	1.11	0.03	0.05	0.05	0.00

Error decomposition

Parameter	True value	Finite sample	Missing outcomes	Measurement error	Total
$\gamma^{(0)}_{\text{down}}$	0.30	0.08	0.04	-0.02	0.10
$\gamma^{(0)}_{\text{up}}$	0.30	0.02	0.04	-0.03	0.03
$\gamma^{(1)}_{\text{down}}$	0.21	-0.01	0.02	-0.01	-0.01
$\gamma^{(1)}_{\text{up}}$	0.21	0.01	0.01	-0.01	0.00
$\gamma^{(2)}_{\text{down}}$	0.00	-0.02	0.00	0.00	-0.02
$\gamma^{(2)}_{\text{up}}$	0.00	-0.01	0.00	0.01	0.00
M^{tot}	1.77	0.16	0.31	-0.14	0.32
M^{self}	1.09	0.01	0.02	-0.01	0.02

Table 6: Algorithm 3 results with simulated measurement error and missing outcomes

Parameter	True Value	Mean $\hat{\gamma}$	Median $\hat{\gamma}$	Mean OLS SE	RMSE	SD of $\hat{\gamma}$	Mean(SE − SD of $\hat{\gamma}$)^2
$\gamma^{(0)}_{\text{down}}$	0.30	0.41	0.44	0.11	0.27	0.25	0.02
$\gamma^{(0)}_{\text{up}}$	0.30	0.35	0.36	0.08	0.16	0.15	0.01
$\gamma^{(1)}_{\text{down}}$	0.21	0.20	0.16	0.11	0.16	0.16	0.00
$\gamma^{(1)}_{\text{up}}$	0.21	0.21	0.20	0.09	0.11	0.11	0.00
$\gamma^{(2)}_{\text{down}}$	0.00	-0.03	-0.02	0.11	0.18	0.18	0.00
$\gamma^{(2)}_{\text{up}}$	0.00	0.01	0.00	0.09	0.13	0.13	0.00
M^{tot}	1.77	2.08	2.02	0.34	0.59	0.50	0.05
M^{self}	1.09	1.11	1.11	0.03	0.04	0.04	0.00

Error decomposition

Parameter	True value	Finite sample	Missing outcomes	Measurement error	Total
$\gamma^{(0)}_{\text{down}}$	0.30	0.08	0.03	0.00	0.11
$\gamma^{(0)}_{\text{up}}$	0.30	0.03	0.04	-0.02	0.05
$\gamma^{(1)}_{\text{down}}$	0.21	0.00	-0.01	0.00	-0.01
$\gamma^{(1)}_{\text{up}}$	0.21	0.02	-0.01	-0.01	0.00
$\gamma^{(2)}_{\text{down}}$	0.00	-0.02	0.00	0.00	-0.03
$\gamma^{(2)}_{\text{up}}$	0.00	0.00	0.00	0.01	0.01
M^{tot}	1.77	0.22	0.16	-0.07	0.31
M^{self}	1.09	0.02	0.01	0.00	0.02

the measurement error bias are of opposite sign; therefore the similarity in results between the two algorithms is, to some degree, only apparent.