Propagation of Shocks in Networks: Identification and Applications*

Gabriel Chodorow-Reich, Xavier Gabaix, Davide Viviano

March 25, 2026. Preliminary and Incomplete.

Abstract

This paper provides an estimation procedure to causally identify the impact of shocks in fairly general networks, which we term "network GIV" (granular instrumental variables). We show that, using panel data with some large units (such as countries, industries, or firms) and a parametric model of how shocks propagate across these units, one can recover idiosyncratic unit-level shocks and use them to estimate the underlying parameters of the macroeconomic network.

We apply the network GIV to estimate the transmission of TFP shocks across sectors, a key mechanism for economic growth. We study how a TFP shock in one industry propagates to its customers and suppliers. We find that TFP shock amplification is very large: if industries experience, on average, a 1% primitive increase in TFP, then after propagation through the production network, aggregate economy-wide TFP rises by 5%. This fivefold amplification is a long-run effect and reflects two components. First, following a 1% primitive shock, industry-level TFP (measured on gross output) rises on average by 2.7% due to upstream and downstream spillovers—this is the effect that we estimate. Second, this increase translates into a 5% rise in aggregate TFP (measured on net output), reflecting the traditional Hulten effect, which nets out intermediate inputs.

More broadly, we show how the method can be used to causally estimate a range of quantities of interest, including various elasticities, and to study how productivity,

*chodorowreich@fas.harvard.edu, xgabaix@fas.harvard.edu, 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 Aureo de Paula, Basile Grassi, Mikkel Plagborg-Møller, and participants at seminars and conferences.

demand, or financial shocks propagate within and across countries. The network GIV thus provides a general-purpose tool for causal identification in economics, particularly in macro-finance settings where valid instruments are typically scarce.

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 extracts idiosyncratic shocks (via a factor model) and uses them as primitive instruments to estimate for instance demand and supply elasticities in a given market. That procedure worked for one market — here 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. Our goal is to identify and estimate these idiosyncratic shocks and their propagation through the network. 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

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

Here Y_t is the observed vector of outcomes of dimension N. The $N \times N$ matrices $B_0(\gamma^\star)$ and $B_1(\gamma^\star)$ characterize the network (for example, input-output matrices) and are known up to an unknown low-dimensional parameter vector $\gamma \in \mathbb{R}^p$ 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 $\lambda\eta_t \equiv 0$. Note that u_t can have heteroskedastic entries (in principle, one could also model the variance of u_t, omitted for expositional convenience).

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 instrument except for the shock of individual i.

In practice, however, we do not observe the shocks 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.

Applications We apply our method to study TFP spillovers in the US economy: Ii 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. 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 sectorsis 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 of 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 (with 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 the impact of a demand of supply shock to one corner of the the stock market (as in as in Koijen 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 effect 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)) by using idiosyncratic shocks to large firms or sectors for identification. Baumeister and Hamilton (2023) propose a maximum likelihood (MLE) approach 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 2.5.2) 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 2.5.2, 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 process, 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 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.

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.

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 this approach, as it does not require existence of a common exogeneous shock (and in fact it removes the factors from the estimation); instead it leverages independence of unit-specific individual shocks.

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) 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 have 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 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, and in Section 2.5 a variety of extensions, e.g. to autocorrelations, heteroskedasticity, and missing data. In Section G, we present a multi-sector macro model and illustrate how our method applies to it. In Section G.5, we bring the model to the data and estimate it using our proposed methodology. Section 4 concludes.

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,

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.

$Lx_t = x_{t-1}$.

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, and 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 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 of η_t. This fits in the framework, with $Y_t = (q_{1t}, \ldots, q_{nt}, p_t)'$, $\gamma = \left(\phi^d, \psi\right)$, $B = \begin{pmatrix} 0 & \phi^d \\ \psi S' & 0 \end{pmatrix}$, $\chi = \begin{pmatrix} I_n \\ 0 \end{pmatrix}$. This shows how the framework generalize 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.

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 G develops such a model, in see in particular Proposition 7.

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 identically distributed over time, although we could directly relax this assumption by assuming forms of mixing and covariances stationarity (omitted for brevity).

Assumption 2. *The u_t are independent and identically distributed 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 assumption 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 assumption on the unobserved shocks. Our idea proceed 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, with pre-multiplying by e_i the left and right hand side, we obtain

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

The main issue of this expression is that the regressor $e_i^\top B_\gamma\left(\gamma\right) Y$ depends on the endogenous variable Y_t. As our second step, we use the fact that shocks are independent over units and replace $e_i^\top B_\gamma\left(\gamma\right) Y$ in the regression above with an instrument $z_{i,t}(\gamma)$ that approximates $E[e_i^\top B_\gamma\left(\gamma\right) Y | \{u_{j,t}\}_{j\neq i}]$. Such an instrument guarantees under independence of shocks across units that $E[z_{i,t}(\gamma)\tilde{v}_{it}(\gamma)\Big|_{\gamma=\gamma^\star}] = 0$. 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 are left to discuss the construction of the instrument $z_{i,t}(\gamma)$. Recall that $Y_t = A(\gamma^\star) u_t$. Therefore, we have $e_i^\top B_\gamma\left(\gamma\right) Y_t = e_i^\top B_\gamma\left(\gamma\right) A(\gamma) u_t$. Hence, we set

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

so $z_{i,t}$ is the instrument using only shocks to entities different from i.

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 completeness, we present here with all its variants, and discuss each of these in the subsequent subsections.

Algorithm 1 One-step estimation of $\gamma^\star$

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 3 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 A.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 OLS 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 $\Delta\gamma$. Update $\hat{\gamma}$ as $\hat{\gamma}' \coloneqq \hat{\gamma} + \Delta\gamma$ (or e.g. $\hat{\gamma}' \coloneqq \hat{\gamma} + 0.2\Delta\gamma$).

return $\hat{\gamma}'$

We state two propositions giving theoretical guaranties 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 to run a sequence of linear regressions), and the moment conditions to identify the target parameter of interest is 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 have the variance matrix of Y_t, which is symmetric and $N \times N$, so has $d_{VY} = \frac{1}{2} N (N+1)$ 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 $p^{\max} = \frac{1}{2}N(N+1) - N$ parameters, if we have a very long time series.[3] We record that for a large number N of units:

$$p^{\max} \sim \frac{1}{2}N^2 \tag{14}$$

So in principle, we can estimate many parameters with a large time series. 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.

2.3 Enrichments

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

2.3.1 The case with 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 A.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:

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

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

3. When we add r factors, as we will do later, this removes rN degrees of freedom, so that$p^{\max} = \frac{1}{2}N(N+1) - (r+1)N$

2.3.2 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 $v_t \coloneqq Y_t - B_0(\gamma^\star)Y_t - X_t b_X$, in Step 1 of Algorithm 1 and its variants.[4]

2.4 Formal statements

We present more formal econometric results. We define

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

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

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

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 when $j \neq i$), we get an extra instrument. Indeed regression (9), i.e. (13) in the Algorithm, we could have

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

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

However, it is not clear that X_t is truly exogenous. For instance, suppose that we estimate the elasticity of demand for oil, and X_{jt} is the business cycle in country j. It is a useful "control", but it is very dependent on the price of oil, which itself will depend on the idiosyncratic demand shock by country i. So, in that case, it may be best to just use z_{it} rather than also the z^X_{it}. As always, the choice of instrument depends on careful thinking about exogeneity.

so that Γ gives the 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,

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

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 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 $\{1, \ldots, p\}$:

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

Proof. See Appendix B.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 $-\sum_{i=1}^N \frac{\partial \mathbb{E}[\bar{H}_i(\gamma)]}{\partial \gamma}\Big|_{g=\gamma^\star}$*

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 signs restrictions on the parameters De 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. *Suppose that Assumptions 1, 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,*

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

Proposition 2 follows directly from standard arguments for extremum estimators (Newey and McFadden 1994). The exact expression for the variance is derived in Appendix B.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 $\Gamma\Gamma^\top$. 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. [5]

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.

5. 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.

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 simple OLS estimator proposed here). This efficiency property, together 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. Then Net-GIV's variance achieves the Cramer-Rao lower bound.*

Corollary 1 states that under Gaussian of the shocks, the proposed procedure is also efficient. The proof is derived in Appendix B.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 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 is very small compared to the first term in some leading examples, see the example 4 below.

Example 4 (Sphere of influence)**.** 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 $\sum_{i=1}^N \frac{\partial \mathbb{E}[\bar{H}_i(\gamma)]}{\partial \gamma}\Big|_{g=\gamma^\star} / \sigma_u^2 = \sum_i \sum_{j\neq i} c_{i,j}^2 S_j^2 + c_{i,j}c_{j,i}S_iS_j$. Suppose now that $c_{i,j}$ is iid across i, j with squared mean $\mathbb{E}[c_{i,j}^2]$ independent of (i, j). We can write, by taking expectations over c, the expression reads as $(N-1)\mathbb{E}[c_{i,j}^2]h^2 + \mathbb{E}\left[c_{ij}\right]^2$ where $h^2 = \sum_j S_j^2$ denotes the Herfindhal of country sizes. This implies that we expect the gradient to be of order $N\mathbb{E}[c_{i,j}^2]h^2$ 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).[6] Then

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

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 $1/\sqrt{TN}$ (more precisely, $1/\sqrt{TNh^2}$).[7] 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.[8]

Results with factor models The appendix (Section B.3) provides propositions for the more complex case with factor models and regressors and discuss inference on those scenarios.

2.5 Discussion

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

2.5.1 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 restriction to enrich the model when the model poorly approximate 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.

7. 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 $1/\sqrt{T}$.

8. 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 Herfindhal. 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 keep $Nh^2 \to \infty$, the first term is of larger order larger than the second one.

2.5.2 Connection to MLE

It is important to compare our procedure to MLE for two reason: (i) MLE is efficient under correct model specification; but (ii) MLE can be sensitive to misspecification of the model as discussed below. In Section B.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 of 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:

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 B.2.[9]

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 to the discussion in Andrews, Kitagawa, and McCloskey (2024), who highlight the relevance of using different sources of exogenous variation for identification of models. Our approach allows us to be explicit in this particular source of exogenous variation we leverage for identifications: independence between unit-specific unobservable shocks. But one can easily refine the identification to use only some of 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

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

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.[10] Likewise, we saw in Section 2.5.1 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 our arguably more, transparent, parsimonious and robust (for more complex models) to standard MLE procedures (while retaining efficiency guarantees

9. 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 turn a potentially high-dimensional number of moment conditions to an exact identified system, that can be easily estimated in practice.

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

in simple models).

3 Application: Propagation of TFP in the US

3.1 Setup

In this application, we are interested in the diffusion of TFP growth: 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 R&D 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 depends 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:

$$Y_t = \sum_{h=0}^{2} B_h Y_{t-h} + \lambda \eta_t + u_t. \tag{27}$$

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:

$$A_{ij} = \frac{x_{ij}}{x_i}, \tag{28}$$

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,

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

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

We parametrize the influence at horizon h matrix B_h as:

$$B_h(\gamma) = \gamma_A^{(h)} A + \gamma_{A^{\text{up}}}^{(h)} A^{\text{up}}, \tag{29}$$

with lags $h = 0, 1, 2$. The scalars $\gamma_A^{(h)}, \gamma_{A^{\text{up}}}^{(h)}$ are weights on two different channels: A captures the influence of the supplier's TFP and A^{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).[11]

3.2 Impact multipliers

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

3.2.1 The long run impact matrix

We have (omitting aggregate factors η_t for simplicity)

$$Y_t = \bar{\mathcal{I}}(L)\, u_t, \tag{30}$$

where L is lag operator ($Lx_t = x_{t-1}$) and $\bar{\mathcal{I}}(L)$ is the impact matrix (with dimension $N \times N$):

$$\bar{\mathcal{I}}(L) \coloneqq \left(I - B_0 - B_1 L - B_2 L^2\right)^{-1} \tag{31}$$

We thus define the long run impact matrix:

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

Indeed, the long run impact on cumulative productivity is

$$\frac{d}{du_t} \sum_{h=0}^{\infty} Y_{t+h} = \bar{\mathcal{I}}(1) = \mathcal{I} \tag{33}$$

11. 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.

For instance, the impact of a TFP shock y_j in industry j on the TFP of industry i is $\mathcal{I}_{ij}y_j$. 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

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:

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

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 GDP also involves a Hulten (1978) effect:[12] 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

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

12. 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).

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

3.2.4 Propagation Dynamics

We next consider the dynamic 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

$$Y_t = \overline{\mathcal{I}}(L)\, u_t = \sum_{h=0}^{\infty} \overline{\mathcal{I}}_h u_{t-h} \tag{38}$$

Hence, the cumulative impact at horizon H is

$$\overline{\mathcal{I}}_{\leq H} \coloneqq \sum_{h=0}^{H} \overline{\mathcal{I}}_h = \frac{d}{du_t} \mathbb{E}_t \left[Y_t + \cdots + Y_{t+H} \right] \tag{39}$$

Hence, the total multiplier at horizon H is

$$M^{\text{tot}}(H) = \frac{\sum_{i,j} S_i \left(\overline{\mathcal{I}}_{\leq H} \right)_{ij}}{\sum_i S_i} \tag{40}$$

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

$$M^{\text{Full}} = M^{\text{Hulten}} M^{\text{tot}} \tag{41}$$

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

3.2.5 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 $f(0,0)$":[13]

$$S^{x,\text{Shapley}} = \frac{1}{2}\left(f(x,y) - f(0,y) + f(x,0) - f(0,0)\right). \tag{42}$$

We apply the Shapley value to the log of the multiplier, $\ln M(\gamma_A, \gamma_{A^{\text{up}}})$ (we drop the "tot" superscript for simplicity).[14] Hence, we get for the multiplier coming from each linkage direction:[15]

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

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

By construction,

$$M = M_{\text{down}} M_{\text{up}} \tag{43}$$

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 A as follows. The "Use" table U reports the 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

13. For instance, if $f(x,y) = ax + by + c$, the Shapley value "coming from x" returns $S^{x,\text{Shapley}} = ax$, reassuringly.

14. We do that because we deal with multiplicative effects.

15. In practice, $M(0,0) = 1$ simplifies a bit the calculation.

Table 1: TFP propagation: estimated parameters

Model	$\gamma_A^{(0)}$	$\gamma_{A^{\text{up}}}^{(0)}$	$\gamma_A^{(1)}$	$\gamma_{A^{\text{up}}}^{(1)}$	$\gamma_A^{(2)}$	$\gamma_{A^{\text{up}}}^{(2)}$
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 27 with three common factors and with B matrices in Equation 29. The estimation uses heteroskedastic PCA as in Algorithm 4. Standard errors are in parentheses.

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 $A = M\left(U - U^M\right)$. We fix A 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.

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. In all specifications the size channel γ_S is set 0, and we estimate the supplier and customer coefficients γ_A and $\gamma_{A^{\text{up}}}$. 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:

$$M^{\text{tot}} = 2.7 \text{ (SE 0.83)}, \qquad M^{\text{self}} = 1.2 \text{ (SE 0.05)} \tag{44}$$

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 of the TFP influence is upstream, i.e. from customers to suppliers ($M_{\text{up}} > M_{\text{down}}$).[16] 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. It is surprisingly 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 are much higher than the private return (e.g. Griliches 1992; Hall, Mairesse, and

16. We have $\ln \frac{M_{\text{up}}}{M_{\text{down}}} = 0.41$ (SE 0.17), which is statistically significant at the 5% level.

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 about 2.6 times larger than the private returns, using as instruments using 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. Also, one could microfound the slow dynamics in a more refined 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. Finally, It could be that physically closer industries influence each other more.[17] This could be done easily, provided more data. We do wish 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, not 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,[18] 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. 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).

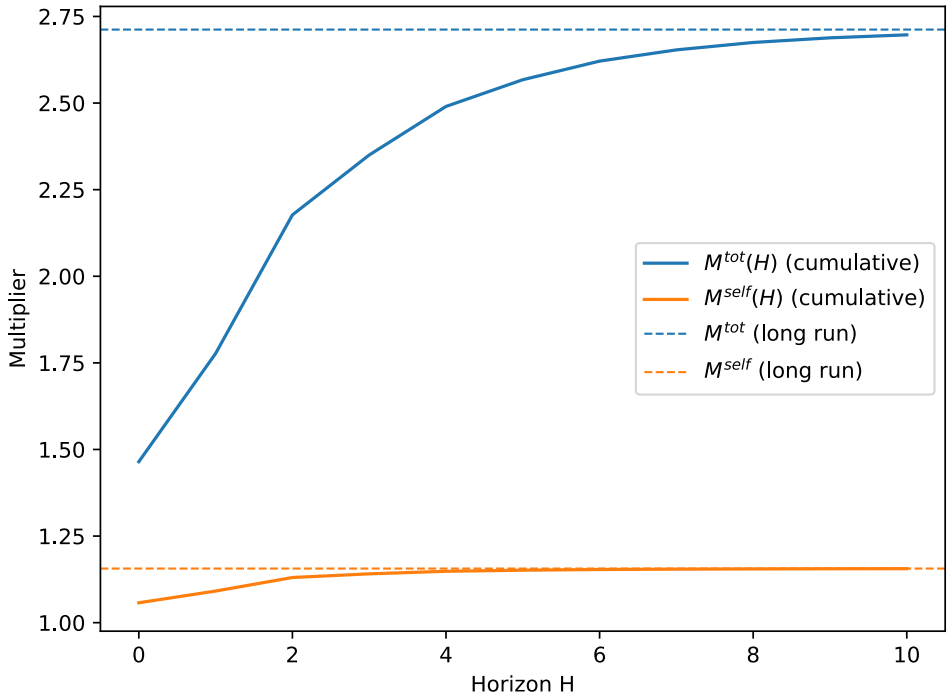
3.4.4 Validation

The online appendix (Section C) contains complements, including simulations, and further diagnostics. We highlight two results here. Figure 2 shows the scatter plot of fitted values

17. That could be modeled respectively as $B_{ij}(\gamma) = (\gamma' X_i) A_{ij}$, with X_i a vector of characteristics such as age; and $B_{ij}(\gamma) = \gamma' A_{ij} P_{ij}$ where P_{ij} is an index of proximity.

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

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 $M^{\text{tot}}$and $M^{\text{self}}$are also shown. The time unit is the year.

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.

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 reaction

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 and lead-lag structures. 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.

Andrews, Isaiah, Toru Kitagawa, and Adam McCloskey. 2024. "Inference on winners." *The Quarterly Journal of Economics* 139 (1): 305–358.

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

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

———. 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.

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.

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.*

———. 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.

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 Algorithmic Details for the Econometric Procedure

A.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 2 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.

A.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 perform PCA on the remaining component.

Algorithm 3 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 4 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 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

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

A.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 $i = 1, \dots, N$, 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 3 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, wit 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 $\check{v}_{it}^k(\hat{\gamma}) \coloneqq \tilde{v}_{it}^k(\hat{\gamma}) - \tilde{v}_{Et}^k(\hat{\gamma})$ for $k = 1, 2$, where $\tilde{v}_{Et}^k(\hat{\gamma})$ 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 3 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, $\tilde{\eta}_t(\hat{\gamma}) = (\tilde{v}_{Et}^1(\hat{\gamma}), \tilde{v}_{Et}^2(\hat{\gamma}), \check{\eta}_t(\hat{\gamma}))$. In this way, we effectively consider $r+2$ factors, with the first two capturing a time fixed effect.

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

Gabriel Chodorow-Reich, Xavier Gabaix, Ralph S.J. Koijen and Davide Viviano

March 25, 2026. Preliminary and Incomplete.

B Complements to the Econometric Procedure

B.1 Proof of the main results

B.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 $E[u_{it}|u_{jt}{}_{j\neq i}] = 0$.

B.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

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

B.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 $\bar{V} = \left(N\bar{\Delta}\right)^{-1}$ 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

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

and therefore

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

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

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

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

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

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 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 $\sum_{i \neq j} \Gamma^k_{j,i} \Gamma^{k'}_{j,i} \sigma^4$. 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:

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

B.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 Net-GIV identifies $\gamma^\star$ using alternative moment conditions.

Proposition 3. (Link with MLE) *Suppose the model with no factors and no lags,* $(I - B(\gamma)) Y_t = u_t$. *Define*

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

and

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

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:

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

In addition, the log likelihood satisfies:[19]

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

which gives the efficiency of the MLE estimator :

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

Proof of Proposition 3. 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 $W_u = V_u^{-1}$. 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

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

We observe that

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

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

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

so

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

Also,

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

so

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

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

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

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

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

So, we have

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

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,

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

we have

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

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

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

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

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

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

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

So,

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

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

B.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 examples those formally studied below in Section B.3, our procedure identifies the parameter of interest whereas MLE might fail. The main reason of this failure is because 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 $\eta_t$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 equal 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 here 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 leverage the fact that we do not need to estimate the fixed effects for estimating $\gamma^\star$.

B.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 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 B.3.3 discusses estimation in settings where λ is unobserved. One could directly extends to settings with estimated λ, in which case the standard errors should be adjust 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 below results with 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

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

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

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

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. □

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 remove 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

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

B.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

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

To express the variance of the estimator, we define:

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

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

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

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

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^{λ}_w using the empirical covariance of the moment restrictions, and estimate $N\bar{\Delta}^{\lambda}_w$ 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.

B.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

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

Proposition 6 (Variance)**.** *Let Assumptions 8, 9, 10, 11 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 B.9.1. □

The variance in Proposition 6 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 11.

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)).

B.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

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

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 3. 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 that 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 admit 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.[20]

B.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 the first one. If there are several input observables ($q > 1$), we pick one type of observable (index by $k = 1 \dots q$), e.g. price, or quantities.

First stage. We define ℓ_{it} (as in left-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: 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.

20. 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.

We plot z_{it} on the x axis vs ℓ_{it} on the y axis We expect that the slope is 1, and we report it.

Second stage. We pick a "default" value γ^d. For instance, if we want to visualize component 1, we take $\gamma^d = \left(\gamma_1^D, \gamma_2^e, \ldots, \gamma_K^e\right)$, 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} based on the final (obtained at convergence) value of γ^e (rather than γ^d). We define $L_t \coloneqq y_t - B_0\left(\gamma^d\right) y_t - B_1\left(\gamma^d\right) y_{t-1}$ and plot the regression:

$$L_{it} = \Delta\gamma \cdot z_{i,t} + \text{error}_{it}$$

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

For example, in Section G.5.2, in the first stage, we plot instrumented inflation $z_{it} = \pi_{i,t}^z$ vs actual inflation $R_{it} = \pi_{i,t}$. In the second stage, we plot $L_t \coloneqq \left(I - A^{up}\right)\Delta\hat{y}_t$ vs $z_{it} = \pi_{i,t}^z$.

B.5 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^*.[21]

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.[22] In sum, the predictions about the impulse-responses ($B(\gamma)$) can be robust (and are completely robust in the case we mentioned), even

21. 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 $D\left(V^y(\gamma), V^{y*}\right)$. Indeed, calling V^{Ye} the empirical variance-covariance matrix, and assuming Gaussian shocks, the log-likelihood is (up to a constant), $\mathcal{L} = -TD\left(V^y(\gamma), V^{Ye}\right)$.

22. 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.

when the details of the model of γ is partially wrong.

B.6 Missing data

We next develop how to handle missing data. Here are two examples to keep in mind. First, we could have a network model of a closed economy, but that the data on the growth of some sectors are not available. Second, the model features N prices and N quantities, and we only observe quantities, but not prices.

Here is the proposed procedure. For concreteness, divide the vector of outcomes y_t as $y_t = [y_{ot} : y_{ht}]$ where y_{ot} is $o$bserved (e.g. we have data on it), and y_{ht} is $h$idden (e.g. is from sectors with no data). Similarly, decompose $u_t = [u_{ot} : u_{ht}]$ into the shocks to observed vs hidden units. We partition the $A = (I - B)^{-1}$matrix into blocks, $A = \begin{bmatrix} A_o^o & A_h^o \\ A_0^h & A_h^h \end{bmatrix}$. We assume that, given γ, we still know the full $A(\gamma)$ matrix (e.g. we know the input-output matrix, though not price variation year to year).

B.6.1 Basic procedure for missing data: no factors, no lags

For clarity, we first take the basic model, $y_t = Au_t$, where we keep the dependence on γ implicit. It then holds

$$y_{ot} = A^o u_t = A_o^o u_{ot} + A_h^o u_{ht}. \tag{B.19}$$

We observe that the theoretical variance matrix of y_{ot} (given γ) is $V^{y_o} \coloneqq A^o V^u A^{o\prime}$ so that we can write the model as

$$y_{ot} = \tilde{A} v_t, \qquad \tilde{A} = (V^{y_o})^{1/2} R, \tag{B.20}$$

for R a rotation matrix of dimension $\dim y_{ot}$ (so $RR' = I$), and where $V^v = I$ and where v_t are independent random variables.

Given this, we proceed as in the main procedure, with $\tilde{B} \coloneqq I - \tilde{A}$. We just enrich γ with the parameters that govern V^u. For interpretability, we pick a rotation matrix R via:

$$\min_{R:R'R=I} \left\| R - (V^{y_o})^{-1/2} A_o^o (V^{u_o})^{1/2} \right\|_2, \tag{B.21}$$

so that, if we did not have hidden units, we recover the correct A matrix.[23]

The conclusion is that we can use the general procedure, which assumes complete observability. We record, nonetheless, the influence of unobserved factors. One issue is that the "idiosyncratic shocks" are a bit harder to interpret. For instance, if we have only quantities, but not prices, the shocks are a mix of supply and demand shocks; and in the "unobserved sector" case, it's influence is still felt by influencing the recovered idiosyncratic factors v_t. Still, identification is achieved. We now show how the basic idea above extends to the other enrichments.

B.6.2 Missing data, in a model with factors

Suppose now the enriched model with factors: $y_t = Au_t + A\lambda\eta_t$. It follows that we can write $y_{ot} = A^o u_t + A^o\lambda\eta_t$. As above, we write $A^o u_t = \tilde{A}v_t$, so that,

$$y_{ot} = \tilde{A}v_t + A^o\lambda\eta_t = \left(I - \tilde{B}\right)^{-1} v_t + A^o\lambda\eta_t,$$

it holds, with $\tilde{\lambda} = \left(I - \tilde{B}\right) A^o\lambda$,

$$y_{ot} = \tilde{B}y_{ot} + v_t + \tilde{\lambda}\eta_t. \tag{B.22}$$

Hence, we can use the general procedure, with full observability.

B.6.3 Missing data, in a model with lags

Suppose now a model with factors and lags, $y_t = Au_t + A\lambda\eta_t + Cy_{t-1}$ Allowing C to take arbitrary values would require inputing sequentially each entry of y_{t-1}. This is possible, using, e.g., methods similar in spirit to the Kalman filter or other imputation methods, with y_t the state, and y_{ot} the observed values. The high-level idea is the following: Given a guess for γ (augmented to parametrize variances such as σ_u, σ_η), we have the innovation $\tilde{y}_{ot} \coloneqq y_{ot} - \mathbb{E}_{t-1}\left[y_{ot}|(y_{o,s})_{s<t}\right]$,which has the known distribution (as a function of γ), $V^{\tilde{y}_o}$. Whereas we defer a more comprehensive treatment of this problem to future research, we

23. Indeed, the minimizer rotation is $R = \left(V^{y_o}\right)^{-1/2} A_o^o \left(V^{u_0}\right)^{1/2}$, as the latter is a rotation matrix when $V^{y_o} \coloneqq A_o^o V^{u_0} A_o^{o\prime}$. This leads to $\tilde{A} = A_o^o \left(V^{u_o}\right)^{1/2}$,which is the correct matrix, up to a normalization of the standard deviation of the shocks.

suggest in practice to proceed as follows: obtain a preliminary estimate of A and η e.g., assuming no lagged effects. We start from $t = 0$, and predict $\mathbb{E}[y_0] = 0$. At this point, we can construct $\mathbb{E}_{t=1}[y_1|y_{1o}]$, assuming e.g., joint Gaussianity with the variance-covariance matrix that depends on the corresponding correlations induced by A, λ. We then proceed recursively, each period controlling for each past observed outcome $y_{s\leq t,o}$. We can then iterate multiple times to achieve convergence also over A, λ. We can interpret this approach as similar to a Kalman filter, where here we use preliminary estimates of A, λ to capture the correlation structure.

B.7 Measurement error

A concern in practice is measurement error. Suppose that the true model satisfies our assumptions, e.g. $y_t = B(\gamma) y_t + u_t$, but we observe $y_{ot} = y_t + \varepsilon_t$, where the ε_{it} are mean-zero measurement errors — say uncorrelated with u_s, η_s at all dates, and, for simplicity, i.i.d. with variance σ^2_ε. What to do? We observe $y_{ot} = Au_t + e_t$, with $e_t = A\varepsilon_t$, so

$$V^{y_o} = AV^u A' + V^e \tag{B.23}$$

Hence, we proceed as in Section B.6.1. We have

$$y_t = \tilde{A} v_t, \qquad \tilde{A} = (V^{y_o})^{1/2} R$$

for a rotation matrix R, which we can choose of easy for interpretability as in (B.21). Now, the extended γ vector is made of the previous γ, and σ^2_u, σ^2_ε: $\tilde{\gamma} = (\gamma', \sigma^2_u, \sigma^2_\varepsilon)'$. One can proceed as in the main case. Generically at least, we can identify this extended γ (provided it's not of too high dimension). The interpretation is a bit less easy, but it remains "we look at the impact of idiosyncratic shocks; even though they're a bit mismeasured, the procedure accounts for that".

B.8 Summary statistics for the impulse-responses

B.8.1 The notion of a "Typical impulse response of a multidimensional shock"

Suppose that we have one outcome measure,

$$o_t = b' x_t \tag{B.24}$$

with x_t exogenous, with dimension N, and b with dimension N as well.

For a "typical" or "examplar" observation o_t, we can take the following:

$$o_t^e = b' x_t^e$$

(where the e reminds us that this is an example), where the "examplar" x_t^e might be with one of the following:

1. The simplest is: $x_t^e = (\sigma_{x_i})_{i=1...N}$. This corresponds to "a one standard deviation shock for all units"

2. We could have a "typical" shock, that respects the standard deviation of shocks: we first define $x_t^{e,0} = (\sigma_{x_i})_{i=1...N}$ as the shock before rescaling and then rescale it as $x_t^e = \frac{\sigma_o}{b' x^{e,0}} x^{e,0}$

3. A variant is: $x_t^e = \frac{\sigma_o}{b'\iota}\iota$. This is a typical shock where all components are proportional to 1. For instance, if x_{it} is a TFP shock for good i, x_t^e is a uniform TFP shock in all industries. The prefactor $\frac{\sigma_o}{b'\iota}$ ensures that $o_t^e = \sigma_{b'x}$, so reflects the typical size of the response o_t.

Now, consider an IRF, where we're interested in a series of impacts at different horizons $h \geq 0$,

$$o_{t+h} = \beta_h' x_t \tag{B.25}$$

where the β_h come from the model. We apply select the above x_t^e for $b \coloneqq \frac{1}{H}\sum_{h=0}^{H-1} \beta_h$ for some horizon H (e.g. up to 3 years).[24] This ensures the the "average" is correct up to

24. There could be lots of variants, e.g. , or perhaps $b = \beta_H$ for a given horizon H.

horizon H. Then, we plot

$$o^e_{t+h} = \beta_h' x_{t^e}$$

B.8.2 Concrete application

We trace the IRF of y_{t+h} and π_{t+h} given the various shocks.

Some plots of summary quantities First, we plot some summary stats, $\delta^s_t \coloneqq \sum_i S^y_i \delta^s_{it}$, $\delta^{c,d}_t \coloneqq \sum_i S^c_i \delta^{c,d}_{it}$, aggregate firms demand shocks $\delta^{y,d}_t \coloneqq \sum_i S^y_i \delta^{y,d}_{it}$. But the estimation of IRFs is not done with those aggregate measures, it's done with the disaggregated numbers, e.g. δ^s_{it} rather than δ^s_t.

Typical IRFs We use the method in B.8.1 to plot the "typical impulse response" for the TFP shocks δ^s_{it}, firm demand shocks $\delta^{y,d}_{it}$ and consumer demand shocks $\delta^{c,d}_{it}$. They have hats in (H.25), but in this section we remove the hats.

We have all those shocks handy, except the sector-by-sector demand shocks $\delta^{y,d}_{ijt}$. Those can be extracted as a residual from the demand equation (XX Xavier or Constantin to write down the precise way to do that; but I (Xavier) is arriving in Paris with the family, and won't have the time to do that for a few days probably). But in the meantime, we can look a consumer demand and firm supply shocks.

What fraction of fluctuations comes from which shocks? We can see what's the largest fraction explaining GDP (and later, of inflation). Given

$$y_t = \sum_{h \geq 0} b_h \delta_{t-h} \tag{B.26}$$

decompose the shocks into types k, (e.g. 3 types of shocks: to consumer demand, firms demand, firm supply)[25]:

$$y_t = \sum_k y^{(k)}_t, \qquad y^{(k)}_t = \sum_{h \geq 0} b^{(k)}_h \delta^{(k)}_{t-h} \tag{B.27}$$

25. With perhaps with a special role, later, for the government; or the foreign sector. We can also decompose into idiosyncratic vs aggregate shocks.

The fraction coming from shocks of type k is

$$s_k = \frac{cov\left(y_t, y_t^{(k)}\right)}{var\left(y_t\right)} \tag{B.28}$$

so that $\sum_k s_k = 1$.

If there's TFP shock, that increases the rational demand, so affects the measure of δ_{it}^d. Hence, we might also have a "purged" demand shocks. One way to do that it to regress:

$$\delta_{Et}^d = \beta \delta_{St}^s + \varepsilon_t \tag{B.29}$$

where $S_{it} = \frac{\text{Sales}_{it}}{\text{GDP}_{it}}$ is the Domar weight of sector i, and by Hulten,

$$\delta_{St}^s = \sum_i S_{i,t-1} \delta_{it}^s \tag{B.30}$$

is the total TFP shock. We estimate β, and define

$$\delta_{it}^{d,\text{purged}} := \delta_{it}^d - \beta \delta_{St}^s \tag{B.31}$$

which is the demand shock "purged from the effect of supply shocks". Here, there is a shortcut: ideally, we'd regress $\delta_{it}^d = \sum_j \beta_{ij} S_{j,t-1} \delta_{jt}^s$, and estimate all the β_{ij}. But there isn't enough statistical power to that, so instead as use the simpler parametrization of a constant β_{ij}. Alternatively, we could have a structural model, e.g. if TFP increase, consumers should increase their spending, which might look like a demand shock here. But the correct amount depends on the persistence of supply shocks, and the MPC of consumers. So instead, we opt for the simpler approach to purge demand shocks.

So we report the impact of TFP shocks δ_t^s and purged demand shocks, $\delta_t^{d,\text{purged}}$

B.8.3 Supply vs demand shocks in a network economy: conceptual discussion

There is a conceptual difficulty: suppose that firm i decides to use less steel, and fewer plastics (controlling for prices): this is a technological shocks — a supply shock. But, it will appear to steel producers to be a demand shock.

So, to classify as demand shocks: the demands from the consumer side (δ_i^c); and as supply

shocks, the demands from the firm side, δ_i^y, and the own-TFP shock δ_i^s. We decompose the aggregate demand shock of i as (see (H.25))

$$\hat{\delta}_i^d = \Theta_i^c \hat{\delta}_i^{d,c} + \Theta_i^y \hat{\delta}_i^y \tag{B.32}$$

We directly measure $\hat{\delta}_i^{d,c}$, as a residual, and likewise measure $\hat{\delta}_i^y$.

Other version: from the flows from j to i,we get $\hat{\delta}_{ij}^y$, the change in the demand by i for j, that is not explained by price changes. We define the total demand shock from i as:

$$\hat{\delta}_{i\to}^d = \frac{\sum_j A_{ij} \hat{\delta}_{ij}}{\sum_j A_{ij}}$$

so that the "relative" demand shocks for j is $\hat{\delta}_{ij}^d - \hat{\delta}_{i\to}^d$: this one is a TFP shock; while $\hat{\delta}_{i\to}^d$ is a demand shock.

B.9 Omitted proofs for results in the online Appendix

B.9.1 Proof of Proposition 6

Denote $\sigma_{i,t}^2 = \mathbb{E}[u_{i,t}^2]$, and $H_{i,t}(\gamma) = \Big(Y_{i,t} - e_i^\top B(\gamma) Y_t\Big) e_i^\top \frac{\partial B(g)}{\partial g} A(\gamma) y_{t|-i}^\lambda(\gamma)$, 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 8, $Y_t = A(\gamma^\star) v_t^\lambda, \quad v_t^\lambda = u_t + \lambda \eta_t$, we can write

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

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

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

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.

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

Here we give complements to Section 3.

C.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 lags 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$(\text{SE} - \text{SD of } \hat{\gamma})^2$
$\gamma_{\text{A}}^{(0)}$	0.30	0.44	0.10	0.20	0.25	0.01
$\gamma_{\text{A}^{\text{up}}}^{(0)}$	0.30	0.34	0.08	0.13	0.13	0.00
$\gamma_{\text{A}}^{(1)}$	0.21	0.18	0.10	0.13	0.13	0.00
$\gamma_{\text{A}^{\text{up}}}^{(1)}$	0.21	0.20	0.08	0.10	0.10	0.00
$\gamma_{\text{A}}^{(2)}$	0.00	-0.01	0.11	0.11	0.11	0.00
$\gamma_{\text{A}^{\text{up}}}^{(2)}$	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 4. The data generating process simulates heteroskedastic shocks.

C.2 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

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

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

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

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

$$\Delta \ln TFP_t = \sum_{h=0}^{2} b_h Z_{t-h} + \sum_{h=0}^{2} c_h \lambda_S \eta_{t-h} + \varepsilon_t \tag{C.3}$$

where $\lambda_S \coloneqq S'\lambda$.[26] In this case, the estimated b^{tot} is 2.0. The results are in Table 3. We also report the analogue, with growth rate of GDP per capita on the left-hand side. Reassuringly, the $b^{\text{tot}}$is a bit higher, consistent with the factor 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.

Explaining sector-level shock We plot y_{it} vs $\tilde{y}_{i,t|-i}$, which is value of y_{it} that would come from idiosyncratic shocks, defined in Algorithm 1 (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 instrument TFP shocks to sector i induce a fall of prices.

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

Table 3: TFP growth and GDP p.c. growth on idiosyncratic shocks

	TFP growth		GDP p.c. growth	
	(1)	(2)	(3)	(4)
b_0	1.17***	1.39***	0.78**	1.32***
	(0.16)	(0.04)	(0.37)	(0.24)
b_1	0.14	0.29***	0.46	0.93***
	(0.16)	(0.04)	(0.38)	(0.24)
b_2	0.35**	0.34***	0.18	0.28
	(0.17)	(0.04)	(0.39)	(0.24)
b^{tot}	1.66***	2.01***	1.42*	2.54***
	(0.31)	(0.08)	(0.72)	(0.46)
Controls	No	Yes	No	Yes
R^2	0.51	0.97	0.09	0.68
N obs.	57	57	57	57

This table shows the regression results of (C.2). Controls are Domar-weighted common factors as in (C.3). $b^{\text{tot}} = \sum_{h=0}^{2} b_h$. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

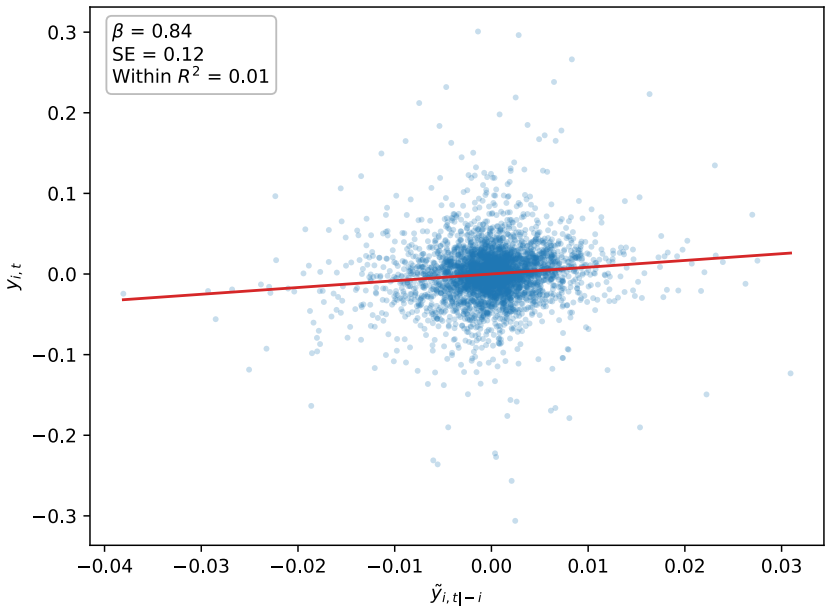
Table 4: Variance decomposition of PCA-estimated common factors

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 4.

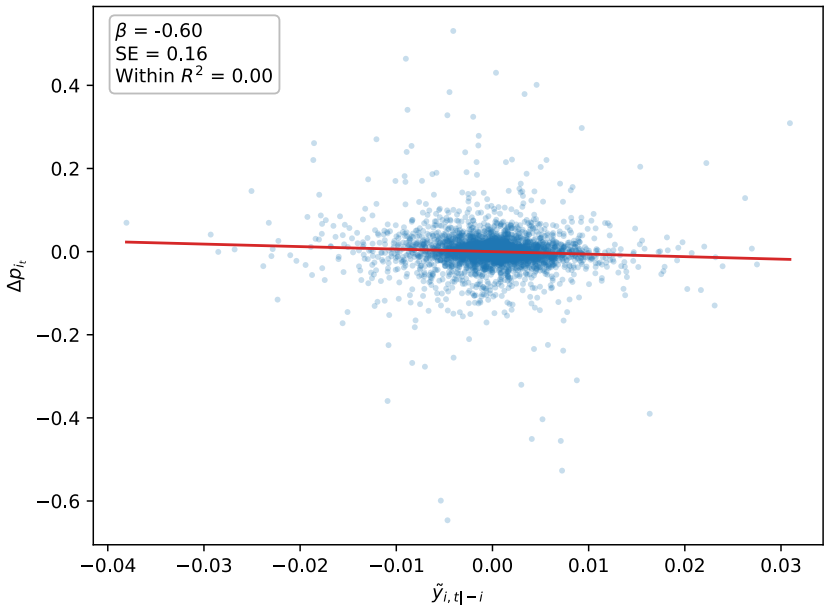
Time-series for the aggregate shocks Figure 4 shows the time-series for the three factors η_t estimated with heteroskedastic PCA as in Algorithm 4. As always, time-series factors are hard to interpret. One can surmise that the first factor η_{1t} is a "2008 recession" factor, while the other two help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

Figure 2: TFP growth and sector-level TFP shocks



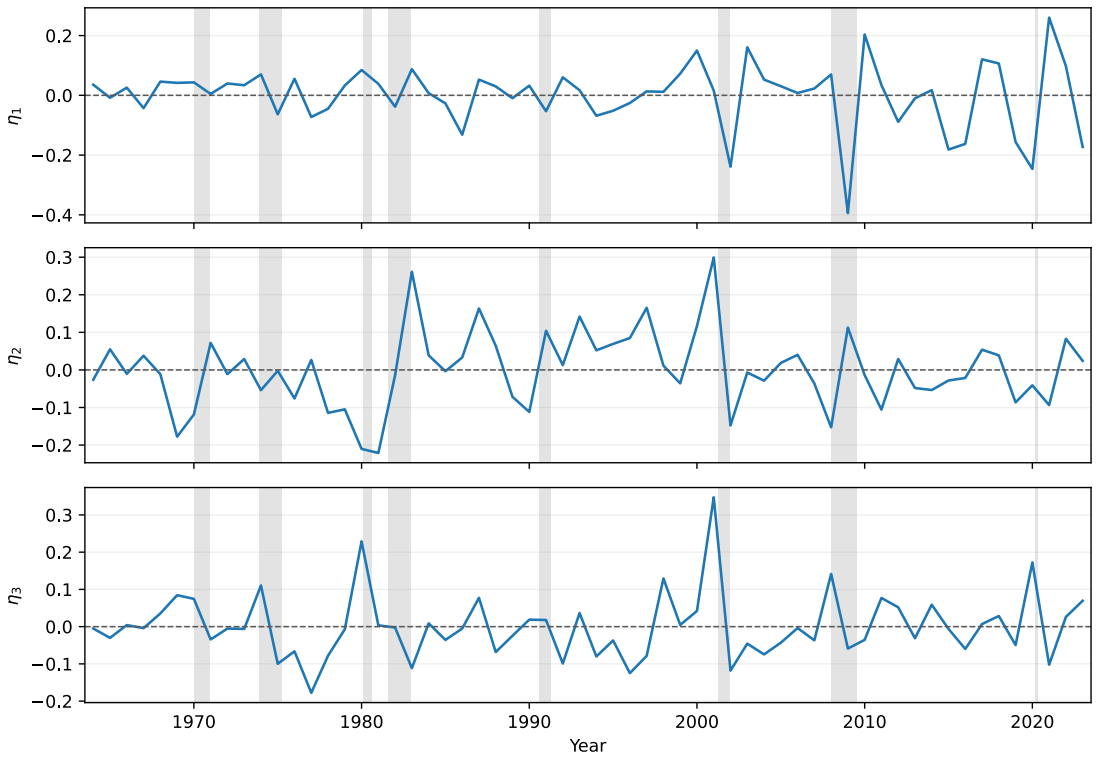
This figure shows the scatterplot of TFP growth y_{it} of sector i vs $\tilde{y}_{i,t|-i}$, which is the TFP growth of sector i induced by the idiosyncratic shocks to sectors outside of i; together with the fitted line. The regression is $y_{it} = a_i + \beta\tilde{y}_{i,t|-i} + e_{it}$. $\tilde{y}_{i,t|-i}$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

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. The regression is $\Delta \ln P_{it} = a_i + \beta\tilde{y}_{i,t|-i} + e_{it}$. $\tilde{y}_{i,t|-i}$ is the value of y_{it} that would come from idiosyncratic shocks, defined in Algorithm 1.

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 4. Shaded regions indicate NBER recession periods. Recession periods are measured monthly, while the estimated factors are annual.