NBER WORKING PAPER SERIES

GRANULAR INSTRUMENTAL VARIABLES

Xavier Gabaix
Ralph S. J. Koijen

Working Paper 28204
http://www.nber.org/papers/w28204

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020, Revised June 2022

We thank Manav Chaudhary, Aditya Chaudhry, Antonio Coppola, Zhiyu Fu, Rishab Guha, Dong Ryeol Lee, Constantin Schesch, Lingxuan Wu, and Yingxuan Yang for excellent research assistance. For helpful comments we thank the editor and referees, Isaiah Andrews, Miriam Arden, Stéphane Bonhomme, Gary Chamberlain, Jiafeng Chen, Raj Chetty, Gabriel Chodorow-Reich, Mark Gertler, Robin Greenwood, James Hamilton, Nathan Hendren, Guido Imbens, Oleg Itskhoki, Larry Katz, Lutz Kilian, Sofonias Korsaye, Jens Kvaerner, Karen Lewis, Erik Loualiche, Matteo Maggiori, Isabelle Méjean, Serena Ng, Ashesh Rambachan, Jean-Marc Robin, Kelly Shue, Jim Stock, Elie Tamer, Harald Uhlig, Christian Wolf, Tiemen Woutersen, Dacheng Xiu, and seminar participants at various conferences. We gratefully acknowledge financial support from the Ferrante Economics Research Fund and the Sloan Foundation (Gabaix) and the Center for Research in Security Prices at the University of Chicago Booth School of Business and the Fama Research Fund at the University of Chicago Booth School of Business (Koijen). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Granular Instrumental Variables
Xavier Gabaix and Ralph S. J. Koijen
NBER Working Paper No. 28204
December 2020, Revised June 2022
JEL No. C01,E0,F0,G0

ABSTRACT

We propose a new way to construct instruments in a broad class of economic environments. In the economies we study, a few large firms, industries or countries account for an important share of economic activity. As the idiosyncratic shocks from these large players affect aggregate outcomes, they are valid and often powerful instruments. We provide a methodology to extract idiosyncratic shocks from the data and create “granular instrumental variables” (GIVs), which are size-weighted sums of idiosyncratic shocks. These GIVs allow us to then estimate parameters of interest, including causal elasticities and multipliers. We illustrate the idea in a basic supply and demand framework. GIVs provide a novel approach to identify both supply and demand elasticities based on idiosyncratic shocks to either supply or demand. We then show how to extend the basic procedure to cover a range of empirically relevant situations. As an application, we measure how “sovereign yield shocks” transmit across countries in the Eurozone. We sketch how GIVs could be useful to estimate a host of other causal parameters in economics.

Xavier Gabaix
Department of Economics
Harvard University
Littauer Center
1805 Cambridge St
Cambridge, MA 02138
and NBER
xgabaix@fas.harvard.edu

Ralph S. J. Koijen
University of Chicago
Booth School of Business
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
Ralph.koijen@chicagobooth.edu

1 Introduction

In many settings, there is a dearth of instruments, which hampers economists' understanding of causal relations (Ramey (2016); Stock and Watson (2016); Nakamura and Steinsson (2018); Chodorow-Reich (2019)). We propose a new way to construct instruments. In the economies we study, many decisions are taken by a few large actors, such as firms, industries or countries, whose idiosyncratic shocks (for instance, productivity shocks) affect the aggregate ones.[1] These idiosyncratic shocks at the firm, industry, or country level are valid instruments for aggregate endogenous variables such as prices. We present a method to extract those idiosyncratic shocks, which allows us to construct "granular instrumental variables" (GIVs). The GIVs then allow us to estimate causal relations in a wide variety of economic contexts.

We first illustrate the idea in a basic static setup with supply and demand (Section 2). It is a classic setting, and we show how GIVs allow for a novel estimation procedure: they yield an instrument that allows us to estimate the elasticities of both supply and demand. Indeed, idiosyncratic demand shocks to large firms or countries give a valid instrument for changes in demand – and thus allow one to estimate the elasticity of supply. They also allow us to estimate the elasticity of demand: the idiosyncratic demand shock of a large firm impacts the price, which changes the demand of other firms. We formalize these ideas and present a way to extract and optimally aggregate idiosyncratic shocks, thus constructing optimal GIVs. Once the ideas are in place, we show in Section 4 how the procedure can be extended to handle various extensions. We also discuss conditions under which GIVs provide powerful instruments.

As an illustration, we study sovereign yield spillovers in the Eurozone during the period from 2009 to 2018. If a country has an increase in its sovereign yield spread (i.e., the yield on its government debt minus the comparable yield for German sovereign bonds), how much does that "spill over" to other Eurozone countries? We use GIVs (based on the impact of idiosyncratic country shocks on the aggregate Eurozone yields) to estimate that spillover.

Uses of GIVs GIVs allow to uncover instruments, especially in finance and macroeconomics, where it is generally challenging to discover valid instruments. Typically, finding an instrument is an ingenious affair that depends detailed historical knowledge and applies only to a specific time period. GIVs can provide a more systematic approach to constructing instruments that can provide an alternative when other instruments are unavailable.

Several recent papers have already applied GIV procedures to identify key parameters and elasticities of interest. Chodorow-Reich et al. (2021) study the multiplier of idiosyncratic shocks to an insurer's asset portfolio on the insurer's equity valuation. Camanho et al. (2022) study the

[1] Hence, economies are "granular:" their shocks are made of incompressible "grains" of economic volatility, the idiosyncratic shocks that occur at the level of firms or industries. This theme is laid out in Gabaix (2011), and developed in Acemoglu et al. (2012), di Giovanni and Levchenko (2012); di Giovanni et al. (2014), and Carvalho and Grassi (2019).

impact of currency flows on exchange rates, using idiosyncratic shocks to fund-level rebalancing. Galaasen et al. (2021) use GIVs to study how idiosyncratic shocks to firms impact banks, and how this spills over to other (small, non-granular) firms borrowing from the same bank. In Gabaix and Koijen (2021), we apply the methodology in this paper to measure the elasticity of the aggregate stock market using idiosyncratic demand shocks to large investors or investor sectors. Schubert et al. (2022) study the impact of concentration on wages and use idiosyncratic firm-level shocks to instrument for concentration. Kundu and Vats (2021) estimate how idiosyncratic firm-level shocks in one state affect economic activity in other states via their transmission through the banking system. Ma et al. (2021) estimate how lenders' expectations about a city affect GDP growth in the same geography. Lin (2021) estimates how payouts by non-financial public firms affect deposit flows to the banking sector.

Related literature We relate to a number of literatures. We offer some brief pointers here and we provide a longer discussion in Section D. An active literature discusses identification strategies in macro (Ramey (2016); Nakamura and Steinsson (2018); Chodorow-Reich (2019)). We add to it by proposing to use GIVs, which are quite ubiquitous. There are lots of idiosyncratic shocks, and GIVs allow us to construct them quite systematically.

A growing literature finds that a sizable amount of volatility is "granular" in nature—coming from idiosyncratic shocks to firms or industries (Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); di Giovanni and Levchenko (2012); di Giovanni et al. (2014); Baqaee and Farhi (2019); Carvalho and Grassi (2019); Gaubert and Itskhoki (2021)). We provide tools to isolate idiosyncratic shocks in the presence of common factors. Datasets used in this literature can be revisited GIVs can be constructed to investigate causal relations.

The idea of using idiosyncratic shocks as instruments to estimate spillover effects has been explored in several creative papers, as we discuss in more detail in Section D, such as Leary and Roberts (2014b), Amiti and Weinstein (2018), Amiti et al. (2019) and Sarto (2018). However, the typical approach has been to use idiosyncratic shocks to variables that are excluded from the main estimating equation to construct instruments. We instead use the idiosyncratic shocks in the estimating equation directly. In addition, we allow for more flexible exposures to unobserved common shocks in extracting idiosyncratic shocks.

Outline Section 2 introduces the GIV framework, centered around a simple model of supply and demand. Section 3 gives a practical user's guide. Section 4 presents a number of extensions and robustness checks. Section 5 gives an empirical application. Section 6 concludes. Long proofs are in Section C of the online appendix.

Notations For a vector $X = (X_i)_{i=1\ldots N}$ and a series of weights w_i, we define $X_w = \sum_i w_i X_i$. With relative weights S_i that satisfy $\sum_{i=1}^{N} S_i = 1$, we define $X_E := \frac{1}{N}\sum_{i=1}^{N} X_i$, $X_S := \sum_{i=1}^{N} S_i X_i$ so that X_E is the equal-weighted average of the vector's elements, X_S is their size-weighted average. We also commonly use the notation u_i for shocks that are uncorrelated and with variance $\sigma^2_{u_i}$. Then, we will define the "inverse variance weights" or "quasi-equal weights" (using that term to highlight that they are a small variant of equal weights): $\tilde{E}_i := \frac{1/\sigma^2_{u_i}}{\sum_j 1/\sigma^2_{u_j}}$, which satisfy $\sum_i \tilde{E}_i = 1$, and are equal to $\tilde{E}_i = \frac{1}{N}$ when all the $\sigma^2_{u_i}$ are equal. Then $X_{\tilde{E}} := \sum_{i=1}^N \tilde{E}_i X_i$. We use the notation x^e for an estimator of a variable x, or as a "proxy" for variable x, meaning a variable close to x but not exactly equal to it, even asymptotically – we will make the difference clear. We use $\mathbb{E}_T$ for the sample temporal mean, $\mathbb{E}_T[Y_t] := \frac{1}{T}\sum_{t=1}^T Y_t$; C_t for a vector of controls; ι for a vector of 1's, I for the identity matrix, both of the appropriate dimension given the context; V^Y for the variance-covariance matrix of vector Y_t (so $V^Y = \mathbb{E}[Y_t Y_t']$ if Y_t has mean zero). In Section A, we summarize more specialized notations.

2 The basics of Granular Instrumental Variables

2.1 A benchmark model

For clarity, we lay out a concrete economic model of the equilibrium in, for instance, the oil market. There is a succession of i.i.d. economies indexed by t. Demand by country i at date t is $D_{it} = \bar{Q} S_i (1 + y_{it})$, where $\bar{Q}$ is the average total world production, y_{it} is a demand shift term, and S_i is country i's share of demand, normalized to follow $\sum_{i=1}^N S_i = 1$. The demand shift y_{it} is

$$y_{it} = \phi^d p_t + \lambda_i \eta_t + u_{it}, \qquad \lambda_i \eta_t = \sum_{f=1}^r \lambda_i^f \eta_t^f, \tag{1}$$

where $p_t = \frac{P_t - \bar{P}}{\bar{P}}$ is the proportional deviation from $\bar{P}$, which can be thought as the average price of oil, ϕ^d is the elasticity of demand, η_t is a vector of common shocks, vector λ_i is country i's sensitivity to the common shocks, and u_{it} is the idiosyncratic demand shock by country i.

All shocks are i.i.d. across dates. Then, total world demand is $D_t = \sum_i D_{it} = \bar{Q}(1 + y_{St})$, where $y_{St} := \sum_i S_i y_{it}$ is the size-weighted average demand disturbance. We suppose that supply is $Q_t = \bar{Q}(1 + s_t)$, where the supply shift s_t is

$$s_t = \phi^s p_t + \varepsilon_t, \tag{2}$$

where ϕ^s is the elasticity of demand and ε_t is a supply shock.

Then, to equilibrate supply and demand ($D_t = Q_t$), the price must adjust so that $\bar{Q}(1 + y_{St}) = \bar{Q}(1 + \phi^s p_t + \varepsilon_t)$, i.e., $y_{St} = s_t$, which gives

$$p_t = \frac{u_{St} + \lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi^d} = \mu u_{St} + \varepsilon_t^p, \tag{3}$$

where $\mu := \frac{1}{\phi^s - \phi^d}$ is the price impact of a demand shock u_{St}, and $\varepsilon_t^p := \frac{\lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi^d}$ is made of aggregate shocks. The equilibrium quantity produced is given by:

$$s_t = y_{St} = \frac{\phi^s u_{St} + \phi^s \lambda_S \eta_t - \phi^d \varepsilon_t}{\phi^s - \phi^d} = M u_{St} + \varepsilon_t^s, \tag{4}$$

where the multiplier $M := \frac{\phi^s}{\phi^s - \phi^d}$ is quantity impact of a demand shock u_{St}, and $\varepsilon_t^s := \frac{\phi^s \lambda_S \eta_t - \phi^d \varepsilon_t}{\phi^s - \phi^d}$ is made of aggregate shocks. We want to estimate the elasticity of supply and demand, ϕ^s and ϕ^d, and their related quantities, μ and M. A 1% demand shock leads to a μ% price increase and an M% supply increase.

Throughout this paper we will make the mild assumption that all our variables (such as $\eta_t, \varepsilon_t, u_t = (u_{it})_{i=1...N}$) have finite second moments and have been normalized to have zero mean. We allow ε_t and η_t to be correlated.

2.2 Introducing the GIV

We center our discussion on the estimation of ϕ^s, the elasticity of supply, and discuss the other parameters below. The classic problem is that we cannot estimate ϕ^s by OLS. Indeed, a direct regression supply on the price (as suggested naively by (2)) would be biased as ε_t and p_t are typically correlated.

The idea of the GIV is to use idiosyncratic shocks, u_{it}, as instruments. We assume that the u_{it} are idiosyncratic, that is, they are uncorrelated with the common shocks (η_t, ε_t). That is, for all i, $\mathbb{E}\left[u_{it}\left(\eta_t, \varepsilon_t\right)\right] = 0$, or in vector form with $u_t = (u_{it})_{i=1...N}$,

$$\mathbb{E}\left[u_t\left(\eta_t, \varepsilon_t\right)'\right] = 0. \tag{5}$$

Now first suppose that an oracle gives us the u_{it}'s. Then, we could use those to instrument for the price (see (3)), which would allow us to estimate for instance the supply elasticity (see (2)). In practice, we do not have an oracle, so we need to estimate the u_{it}'s, or to construct some proxy u_{it}^e for them that can serve the same purpose as the unobserved idiosyncratic shocks.[2] We now show that this is feasible.

The simplest case is when the factor structure (1) is only a "time fixed effect:" $\lambda_i \eta_t = \eta_t$. This is the case that we recommend keeping in mind to build intuition. To handle more general cases, we develop the theory under Assumption 1 below. We maintain it in Sections 2 and 3. In later sections, we consider further generalizations.

[2] For instance, take the simple model where the econometrician observes $y_{it} = \eta_t + u_{it}$. We cannot exactly recover u_{it}, but by forming $u_{it}^e := y_{it} - y_{Et}$ one can recover $u_{it}^e = u_{it} - u_{Et}$. We shall call u_{it}^e a "proxy" rather than an "estimate" of u_{it}. Likewise, $\eta_t^e := y_{Et} = \eta_t + u_{Et}$ will be a "proxy" for η_t. Section C.5 details this for more general structures.

Assumption 1 (Parametric factors). *We know characteristics X_i (with dimension $1 \times r$) for each entity i, and the loadings are linear functions of those characteristics: $\lambda_i = X_i \dot{\lambda}$ for some matrix $\dot{\lambda}$ of dimension $r \times r$. We suppose that the vector of loadings include, as a first vector, $\iota = (1, \ldots, 1)'$, which captures common shocks, i.e. $X_{i1} = 1$ for all i's.*

We study the parametric case, which covers many situations of interest. First, it covers the basic "time fixed effect" case, where X_i is simply a vector of ones. Second, this modeling approach is consistent with the practice in modern macro and finance, in which exposures to risks align with characteristics (see e.g. Fama and French (1993)). In this context, parametric approaches are preferred as they are more stable than non-parametric approaches. Third, it is much simpler analytically, so that we can concentrate on the key insights of the GIV. We call $\Lambda = (\lambda_i)_{i=1\ldots N}$ the $N \times r$ matrix of loadings. Banafti and Lee (2022) extend our framework to the case of non-parametric factor loadings that are estimated.

We construct the GIV as follows. Suppose that we have a set of weights $\Gamma \in \mathbb{R}^N$ orthogonal to Λ, $\Gamma'\Lambda = 0$, and such that $\Gamma' S \neq 0$. In (12) we provide a concrete construction of Γ. Then the GIV is defined as:

$$z_t := \Gamma' y_t = \sum_{i=1}^{N} \Gamma_i y_{it}. \tag{6}$$

The GIV is constructed from observables, y_{it}. The key observation is that as $\Gamma'\iota = \Gamma'\Lambda = 0$, $z_t := \Gamma' y_t = \Gamma' \left(\iota \phi^d p_t + \Lambda \eta_t + u_t\right) = \Gamma' u_t$, so

$$z_t = \Gamma' u_t. \tag{7}$$

As a result, the GIV z_t is a linear combination of idiosyncratic shocks. By (5), the GIV satisfies the exogeneity condition

$$\text{Exogeneity: } \mathbb{E}\left[(\eta_t, \varepsilon_t) z_t\right] = 0. \tag{8}$$

Also, the relevance condition holds (because $\Gamma' S \neq 0$):

$$\text{Relevance: } \mathbb{E}\left[p_t z_t\right] \neq 0.$$

Given that $s_t - \phi^s p_t = \varepsilon_t$, per (2), and $\mathbb{E}\left[u_t \varepsilon_t\right] = 0$, per (5), we have:

$$\mathbb{E}\left[(s_t - \phi^s p_t) z_t\right] = 0, \tag{9}$$

which implies $\phi^s = \frac{\mathbb{E}[s_t z_t]}{\mathbb{E}[p_t z_t]}$. The GIV can therefore be used to estimate ϕ^s. We now state a formal proposition.[3]

[3] It holds under mild regularity conditions on the joint distribution of $(u_{it}, \eta_t, \varepsilon_t)$ given that the data are i.i.d. across dates.

Proposition 1 (Consistency of the GIV estimator). *Suppose that $\mathbb{E}[u_t \varepsilon_t] = 0$, although the u_{it} can have any arbitrary distribution with mean 0 and finite variance. Form the GIV $z_t := \Gamma' y_t = \sum_i \Gamma_i y_{it}$. Then, z_t identifies the supply elasticity, by $\phi^s = \frac{\mathbb{E}[s_t z_t]}{\mathbb{E}[p_t z_t]}$. For fixed N and as $T \to \infty$, the GIV estimator $\phi_T^{s,e} := \frac{\frac{1}{T}\sum_t s_t z_t}{\frac{1}{T}\sum_t p_t z_t}$ is consistent for the price elasticity ϕ^s.*

Proof. By the law of large numbers, as $T \to \infty$, almost surely $\frac{1}{T}\sum_t s_t z_t \to \mathbb{E}[s_t z_t]$ and $\frac{1}{T}\sum_t p_t z_t \to \mathbb{E}[p_t z_t]$, so that $\phi_T^{s,e} \to \phi^s = \frac{\mathbb{E}[s_t z_t]}{\mathbb{E}[p_t z_t]}$. $\square$

For instance, in the "time fixed effect case", we take $\Gamma = S - E$ (this is optimal, as we shall soon see) that is, $\Gamma_i = S_i - \frac{1}{N}$, so that the GIV is:

$$z_t := y_{\Gamma t} = y_{St} - y_{Et}. \tag{10}$$

In this case, the GIV is the difference between the size-weighted average of consumption, y_{St}, and its equal-weighted average, y_{Et}. The structure implies that

$$z_t = u_{\Gamma t} = u_{St} - u_{Et}$$

so that the GIV is the difference between the size-weighted average of idiosyncratic shocks, u_{St}, and its equal-weighted average, u_{Et}. The GIV puts a high weight on large countries. Intuitively, if China has an idiosyncratic demand shock for oil, then it will change the world demand and total quantity produced, which will allow us to estimate the elasticity of supply for oil. In the next section, we discuss conditions under which this instrument is powerful.

2.3 Optimal GIV weights and the precision of the GIV estimator

We now explore under which conditions the GIV estimator is precise. Under a standard central limit theorem, an appropriately scaled and centered version of the above GIV estimator is asymptotically Gaussian for fixed N and as $T \to \infty$:

$$\sqrt{T}\left(\phi_T^{s,e} - \phi^s\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\phi^s}^2\right),$$

where $\sigma_{\phi^s}^2$ is the asymptotic variance as derived in the next proposition. We can use this result to derive the weights Γ that yield the highest precision, that is, the lowest $\sigma_{\phi^s}^2$ (its proof is in Section C). For this, we use the $N \times N$ projection matrix Q that is orthogonal to the factor loadings, i.e. which satisfies $Q\Lambda = 0$:

$$Q := I - \Lambda \left(\Lambda'\Lambda\right)^{-1} \Lambda' \tag{11}$$

Proposition 2 (Optimal weights Γ for the GIV $y_{\Gamma t}$). *Assume that ε_t has variance σ_ε^2 conditional on the u_{it}'s. The asymptotic variance of the GIV estimator $\phi_T^{s,e}$ in Proposition 1, which is $\sigma_{\phi^s}^2 =$*

$\lim_{T\to\infty} T\mathbb{E}\left[(\phi_T^{s,e} - \phi^s)^2\right]$, *satisfies* $\sigma_{\phi^s}(\Gamma) = \frac{\sigma_\varepsilon \mathbb{E}\left[u_{\Gamma t}^2\right]^{1/2}}{|\mu\mathbb{E}[u_{St}u_{\Gamma t}]|}$. *The GIV weights*

$$\Gamma^{*\prime} = S'Q \tag{12}$$

minimize the asymptotic variance, with $\sigma_{\phi^s}(\Gamma^*)^2 = \frac{\sigma_\varepsilon^2}{\mu^2\mathbb{E}\left[u_{\Gamma^* t}^2\right]}$. *This implies that for any other* Γ *that is not collinear to* Γ^*, *the asymptotic variance* $\sigma_{\phi^s}^2(\Gamma)$ *is larger. Hence, the optimal GIV is*

$$z_t := \Gamma^{*\prime} y_t = S'Qy_t = S'\check{u}_t \tag{13}$$

with $\check{u}_t := Qu_t$. *In other terms,*

$$z_t = \sum_i S_i \check{u}_{it}, \tag{14}$$

with $\check{u}_{it}$ *the residual from the regression of the cross-sectionally demeaned consumption deviations* $\check{y}_{it} := y_{it} - y_{Et}$ *on the demeaned factor loadings* $\lambda_i - \lambda_E = (0, \check{\lambda}_i)$*:*

$$\check{y}_{it} := y_{it} - y_{Et} = \check{\lambda}_i \eta_t^x + \check{u}_{it}, \tag{15}$$

where $\eta_t = (\eta_t^1, \eta_t^x)$ *with* $\eta_t^1 \in \mathbb{R}$ *and* $\check{\lambda}_i, \eta_t^x \in \mathbb{R}^{r-1}$. *We therefore also obtain an estimate of the aggregate shocks* η_t^x.

Each country i affects the price proportionally to its size S_i, see (3). Hence, the economically appropriate weights are S_i. Proposition 2 shows that those are also the statistically appropriate weights. The intuition for (12) is that Γ^* is the vector closest to S, while being orthogonal to the factor loadings.

Simple example with an additive common shock We now go back to the "time fixed effect" case, $\lambda_i \eta_t = \eta_t$. It allows us to develop the main intuition in a transparent way.

Proposition 3 (Precision of the GIV estimator in the time fixed effect case). *In the "time fixed effect" case where* $\lambda_i = 1$, *the optimal GIV estimator takes the form* $\Gamma = S - E$, *that is,* $\Gamma_i = S_i - \frac{1}{N}$*:* $z_t := y_{\Gamma t} = y_{St} - y_{Et}$, *so that* $z_t = u_{\Gamma t} = u_{St} - u_{Et}$. *If* ε_t *is homoskedastic conditional on the* u_{it}*'s, then the asymptotic standard deviation of the scaled and centered GIV estimator is* $\sigma_{\phi^s} = \frac{\sigma_\varepsilon}{\mu \sigma_{u_\Gamma}}$ *with* $\sigma_{u_\Gamma} = h\sigma_u$, *so that*

$$\sigma_{\phi^s} = \frac{\sigma_\varepsilon}{\mu h \sigma_u}, \tag{16}$$

where h *is the excess Herfindahl:*

$$h := \sqrt{-\frac{1}{N} + \sum_{i=1}^N S_i^2}. \tag{17}$$

The proposition highlights what it takes to obtain a precise estimate of ϕ^s and a powerful instrument. We need some large units (in order to have a large excess Herfindahl h) and we need that idiosyncratic shocks are large compared to aggregate shocks (large $\sigma_u/\sigma_\varepsilon$). In our demand-supply model, we need at least one large country, with volatile idiosyncratic shocks relative to the volatility of the supply shocks.

There is another way in which the GIV is optimal, as one can easily verify in the simplest case where $\lambda_i \equiv 1$. It is the optimally-weighted GMM estimator.[4] This implies that other combinations of idiosyncratic shocks (besides weighing by Γ) cannot increase the precision of the estimator.

Using GIVs to estimate the demand elasticity So far, we have focused on estimating ϕ^s, the supply elasticity. We now show how GIVs can also be used to estimate the demand elasticity, ϕ^d. The GIV has the useful technical property that:[5]

$$\mathbb{E}\left[u_{Et}u_{\Gamma t}\right] = 0. \tag{18}$$

This allows us to estimate ϕ^d. Indeed, from (1) we have: $y_{Et} - \phi^d p_t = \lambda_E \eta_t + u_{Et}$, so that $\mathbb{E}\left[\left(y_{Et} - \phi^d p_t\right) z_t\right] = 0$. This gives the demand elasticity ϕ^d,

$$\phi^d = \frac{\mathbb{E}\left[y_{Et} z_t\right]}{\mathbb{E}\left[p_t z_t\right]}, \tag{19}$$

and the estimator is $\phi_T^{d,e} = \frac{\frac{1}{T}\sum_t y_{Et} z_t}{\frac{1}{T}\sum_t p_t z_t}$.

Hence, the same instrument, the GIV z_t, can be used to estimate both supply and demand elasticities. Intuitively, an idiosyncratic shock to a large country affects both world prices and quantities, so it allows us to estimate the elasticity of demand of the other countries.

2.4 Using GIVs to estimate the price impact and multiplier by OLS

So far, we used GIVs as instruments to estimate the supply and demand elasticities, ϕ^s and ϕ^d. In this section, we show that we can use GIVs to directly estimate the price impact, $\mu = \frac{1}{\phi^s - \phi^d}$, and the multiplier, $M = \frac{\phi^s}{\phi^s - \phi^d}$. Importantly, those parameters that are often of interest to economists can be estimated using standard OLS. To this end, we need to make one normalization assumption.

Assumption 2 (Normalization of GIV) *We assume that* $\frac{\mathbb{E}\left[u_{St}u_{\Gamma t}\right]}{\mathbb{E}\left[u_{\Gamma t}^2\right]} = 1$.

[4] Any moment $\mathbb{E}_T\left[\left(s_t - \phi^s p_t\right)\left(u_{it} - u_{Et}\right)\right] = 0$ is a valid GMM moment to identify ϕ^s. It is easy to check that the first-order condition of the efficient GMM objective function involves size-weighting those moments, which is exactly our GIV moment condition $\mathbb{E}_T\left[\left(s_t - \phi^s p_t\right)\left(u_{St} - u_{Et}\right)\right] = 0$.

[5] Indeed, recall that $\Gamma'\Lambda = 0$, and that the first column of Λ is made of ones: it is $\iota = (1, \ldots, 1)'$. So, we have $E = \frac{1}{N}\iota = \Lambda e$ with $e = \frac{1}{N}(1, 0, 0, \ldots)'$. So $\Gamma' E = \Gamma'\Lambda e = 0$. Using $\mathbb{E}\left[u_t u_t'\right] = \sigma_u^2 I$,

$$\mathbb{E}\left[u_{\Gamma t} u_{Et}\right] = \mathbb{E}\left[\left(\Gamma' u_t\right)\left(u_t' E\right)\right] = \Gamma' \mathbb{E}\left[u_t u_t'\right] E = \sigma_u^2 \Gamma' E = 0.$$

Assumption 2 is verified for instance by the optimal GIV (12).[6] Or this can be ensured by a rescaling of z_t. Indeed for a given value of Γ, such that $\mathbb{E}\left[u_{St}u_{\Gamma t}\right] \neq 0$, we can multiply Γ by a scalar to satisfy Assumption 2. This is useful, for instance, if we have data only on some of the i's.[7]

Proposition 4 (OLS using GIV) *Consider the factor model above, when N is fixed and $T \to \infty$, and Assumption 2 holds. Suppose that we regress the price on the GIV,*

$$p_t = \mu z_t + \varepsilon_t^{p,OLS}, \tag{20}$$

or regress the total quantity change on the GIV,

$$s_t = y_{St} = M z_t + \varepsilon_t^{s,OLS}. \tag{21}$$

Then, the estimated μ and M are unbiased and consistent estimators of $\mu = \frac{1}{\phi^s - \phi^d}$ and $M = \frac{\phi^s}{\phi^s - \phi^d}$ respectively. Furthermore, the standard errors returned by OLS for μ and M are correct.

This means that we can use the GIV via standard OLS under Assumption 2. The fact that the standard errors are correct is a bit surprising at first as z_t is a generated regressor. The reason is that the GIV is directly obtained from an exact formula ($z_t := S'Qy_t$ as in (13)) and can thus be constructed without error. For this, Assumption 1 is important. We relax it later in Section 4.4.

From μ and M, we can recover the elasticities ϕ^s and ϕ^d. This is exactly the same estimate as the IV estimators derived earlier.[8] Indeed, one can view regression (20) as the "first stage" regression of the price on the GIV, which yields the instrumented price $p_t^e := \mu^e z_t$. The "second stage" regresses supply on the instrumented price:

$$s_t = \phi^s p_t^e + \varepsilon_t^{s,\text{2SLS}}, \tag{22}$$

which gives an estimate of ϕ^s. To estimate the demand elasticity, we regress the equal-weighted (not size-weighted) demand on the instrumented change in the price, p_t^e, in the second stage:[9] $y_{Et} = \phi^d p_t^e + \varepsilon_t^{y,\text{2SLS}}$.

GIVs and weak instruments For the parameters that can be estimated using OLS, that is, μ and M, the standard errors obtained via standard OLS inference are valid (as per Proposition 4). When a ratio is implicitly performed, for instance to estimate ϕ^d or ϕ^s, the two-stage least square (2SLS) procedure as in (22) will also give correct standard errors when the instrument is strong

[6] Indeed, $\frac{\mathbb{E}[u_{\Gamma^*}u_S]}{\sigma_u^2} = \Gamma^{*\prime}S = S'QS$ and, as $Q = Q' = Q^2$, $\frac{\mathbb{E}\left[u_{\Gamma^*}^2\right]}{\sigma_u^2} = \Gamma^{*\prime}\Gamma^* = S'QQ'S = S'QS = \frac{\mathbb{E}[u_S u_{\Gamma^*}]}{\sigma_u^2}$.

[7] More concretely, suppose we observe demand data for $n < N$ countries. Then we can still form the GIV based on $\sum_{i=1}^n S_i \check{u}_{it}$, where $\sum_{i=1}^n S_i < 1$.

[8] Indeed, the OLS estimators are $M_T^e = \frac{\mathbb{E}_T[s_t z_t]}{\mathbb{E}_T[z_t^2]}$ and $\mu_T^e = \frac{\mathbb{E}_T[p_t z_t]}{\mathbb{E}_T[z_t^2]}$. We have $\phi_T^{s,e} = \frac{M_T^e}{\mu_T^e} = \frac{\mathbb{E}_T[s_t z_t]}{\mathbb{E}_T[p_t z_t]}$, which is the same as in Proposition 1.

[9] Here we used (18), which makes the OLS valid.

enough. A traditional rule of thumb for the strength of the instrument (in the i.i.d., homoskedastic case) is that the F statistic (which is the squared t-statistic on μ) on the first stage (20) should be greater than a threshold around 16 to 19, and this advice is being progressively enhanced in current IV research (see Montiel Olea and Pflueger (2013) and Andrews et al. (2019)).

2.5 Using common factors as controls to increase the precision of the GIV

We now note how adding estimates of η_t in the procedure enhances the precision of the estimates as it "soaks up" some of the noise. We assumed that $\lambda_i = X_i \dot{\lambda}$ and, after a rotation of the η_t, we can assume that $\dot{\lambda} = I$. We estimate η_t using OLS as in (15) to obtain η_t^e.[10] For finite N, which is the case that we consider in this paper, it is not exactly η_t, but a quantity close to it that suffices for our purposes (as in Footnote 2). In addition, the residuals u_{it}^e from the regressions are $\check{u}_t = Q u_t$, with Q in (11).

Proposition 5 (Results with a control for recovered factors η_t^e). *The results of Propositions 1, 2, and 4 also hold when we control for the estimated factors, η_t^e.*[11] *For instance, in the context of Proposition 1, we use the moments $\mathbb{E}\left[\left(s_t - \phi^s p_t - \beta \eta_t^e\right)\left(z_t, \eta_t^e\right)\right] = 0$, which allow for estimating ϕ^s and β. Likewise, the standard error in Proposition 2 is the same, except that σ_ε is replaced by $\sigma_{\varepsilon^\perp}$, where $\varepsilon^\perp$ is the residual of the regression $\varepsilon_t^p = \beta^\varepsilon \eta_t^e + \varepsilon_t^{p,\perp}$.*

3 A user's guide to GIV

We summarize the above arguments in the form of a user's guide. We outline the baseline algorithm in Section 3.1, and we discuss robustness checks and ways to examine potential sources of misspecification in Section 3.2. In Section 4, we discuss extensions of the model presented so far to cover other empirically relevant settings. We provide extensions to the user's guide in Appendix B.

3.1 The baseline GIV algorithm

The model can be summarized by

$$y_{it} = \phi^d p_t + m^y C_{it}^y + \lambda_i \eta_t + u_{it}, \qquad \lambda_i \eta_t = \sum_{f=1}^{r} \lambda_i^f \eta_t^f,$$
$$s_t = \phi^s p_t + m^s C_t^s + \varepsilon_t,$$

[10] Note that η_t^e has dimension $r-1$, while η_t has dimension r. This is because the loading on the first factor ι is $\phi^d p_t + \eta_{1t} + u_{Et}$, so that we cannot recover η_{1t} as ex ante we do not know ϕ^d.

[11] This is, run the regressions $p_t = \mu z_t + \beta^p \eta_t^e + \dot{\varepsilon}_t^p$ and $s_t = M z_t + \beta^s \eta_t^e + \dot{\varepsilon}_t^s$.

where $\lambda_i = X_i \dot{\lambda}$ and, after a rotation of the η_t, we can assume that $\dot{\lambda} = I$. We assume that $X_i = (1, x_i)$ and we write $\eta_t = (\eta_t^1, \eta_t^x)$ corresponding to the loadings of 1 and x_i, respectively.[12] C_{it}^y and C_t^s are observable controls, including constants and entity fixed effects. The shocks u_t are idiosyncratic, so uncorrelated with ε_t, η_t, C_t^y and C_t^s. We determine the equilibrium price and quantities as before by imposing market clearing, $y_{St} = s_t$.

We then implement the following algorithm to estimate the relevant parameters. Proposition 6 below justifies this algorithm.

1. *Panel regression:* Estimate a panel regression for y_{it} with a time fixed effect, b_t, and the observable controls in the demand equation:

$$y_{it} = b_t + m^y C_{it}^y + \check{y}_{it}. \tag{23}$$

   The regression yield estimates of b_t,m^y, and residuals $\check{y}_{it}^e$.

2. *Factor estimation:* Estimate the factors using period-by-period cross-sectional regressions of $\check{y}_{it}^e$ on the demeaned characteristics $\check{x}_i := x_i - x_E$ as in (15):

$$\check{y}_{it}^e = \check{x}_i \eta_t^x + \check{u}_{it}. \tag{24}$$

   The regression yields estimates of η_t^x, which we call η_t^e, and residuals $\check{u}_{it}$. We use η_t^e as controls in the subsequent steps to tighten the standard errors (see Proposition 5).

3. *OLS estimation using GIVs:* Form $Z_t := y_{St} - y_{Et}$ and run the regression of p_t (respectively s_t) on Z_t, and controls $C_{St}^y, C_t^s, \eta_t^e$:

$$p_t = \mu Z_t + \beta^{p,y} C_{St}^y + \beta^{p,s} C_t^s + \beta^{p,\eta} \eta_t^e + \varepsilon_t^{p,OLS}, \tag{25}$$

$$s_t = M Z_t + \beta^{s,y} C_{St}^y + \beta^{s,s} C_t^s + \beta^{s,\eta} \eta_t^e + \varepsilon_t^{s,OLS}. \tag{26}$$

   The regression yields estimates of μ, M, and the unimportant ancillary parameters β.

4. *IV estimation using GIVs:* We estimate the elasticity ϕ^d using the instrumental variables regression

$$y_{Et} = \phi^d p_t + m^y C_{Et}^y + \beta^y \eta_t^e + \varepsilon_t^{y,IV}, \tag{27}$$

   where Z_t is an instrument for p_t and with C_{Et}^y and η_t^e as controls. The regression yields estimates of ϕ^d, m^y and β^y. For the supply elasticity, we use the instrumental variables regression

$$s_t = \phi^s p_t + m^s C_t^s + \beta^s \eta_t^e + \varepsilon_t^{s,IV}, \tag{28}$$

[12] To clarify the dimensions, it holds that $X_i \in \mathbb{R}^r$, $x_i \in \mathbb{R}^{r-1}$, $\eta_t^1 \in \mathbb{R}$, and $\eta_t^x \in \mathbb{R}^{r-1}$.

where Z_t is an instrument for p_t and with C^y_{Et} and η^e_t as controls. The regression yields estimates of ϕ^s, m^s and β^s.

We used the following useful equivalence. Provided that we add η^e_t as controls, one can either use the purified GIV $z_t = \sum_i S_i \check{u}_{it}$ or, equivalently, the simple GIV $Z_t := y_{St} - y_{Et} = z_t + (\lambda_S - \lambda_E)\,\eta_t$ as the instrument. Z_t is easier to use than z_t when comparing the stability of estimates as we vary the number of components.

Proposition 6 *The procedure outlined just above gives consistent estimators of* $\left(\mu, M, \phi^d, \phi^s\right)$. *Furthermore, the standard errors returned by OLS for* μ *and* M *are correct.*

After estimating the parameters of interest using GIV, we recommend exploring the diagnostics for misspecification as discussed in the next section.

3.2 Robustness to misspecification and threats to identification

Before discussing the limitations of GIVs and diagnostics for misspecification, we highlights the forms of misspecification to which GIVs are robust.

Forms of misspecification to which GIVs are robust First, it is possible to construct GIVs based on a subset of entities, I_t, that is, $z_t = \sum_{i \in I_t} S_i \check{u}^e_{it}$ (with $\check{u}^e_{it} = u_{it} - u_{Et}$). In practice this can be useful as we can select the top K entities, the entities for which we have data, or to omit entities for which the data may contain measurement error. In this case, all results go through, although a rescaling may be required to ensure that Assumption 2 holds.[13] The estimator is still valid, just not the optimal GIV estimator.

Second, suppose that we misspecify the vector S of size weights, for example, by defining $z_t = \sum_i S^\circ_i \check{u}^e_{it}$ using a wrong vector S°. Then, the parameters estimated using IV are still consistently estimated, but the parameters estimated using OLS can be biased.[14] After all, we still have (9), that is, $\mathbb{E}\left[(s_t - \phi^s p_t)\, z_t\right] = 0$, so that the IV procedure in Proposition 1 still works.

Third, if we assume that the elasticities are homogeneous across actors (of demand in our demand and supply example), while they are actually heterogeneous, then the IV estimates are correct and so are the OLS estimates, assuming that η_t was well-estimated in the cross-section. In this case, the parameters that we estimate are equal-weighted averages of coefficients. For instance, the IV estimates yield ϕ^s and ϕ^d_E, and the OLS coefficients are those corresponding to the interpretation that the elasticity of demand is ϕ^d_E rather that ϕ^d_S. Section H.3 provides the derivations.

[13] For instance, we still have $u_{St} = z_t + \varepsilon^{us}_t$ with $z_t \perp \varepsilon^{us}_t$. Section H.7 gives for a formal analysis.

[14] Calling $\psi = \frac{\mathbb{E}[z_t u_{St}]}{\mathbb{E}[z_t^2]}$ (which is 1 when $S^\circ = S$), then the OLS above gives the estimates (in expectation) $\mu^e = \mu\psi$ and $M^e = M\psi$. For some selection procedures (e.g. selecting the shocks to some pre-specified entities as we discussed), we still have that $\psi = 1$, so that OLS is still valid.

The limits of GIVs and diagnostics for misspecification We now discuss the main limits to the applicability of GIVs and some diagnostics for misspecification.

First, the GIV may not be sufficiently volatile and have low power. In case of the OLS estimators, this simply manifests itself as large standard errors, while in case of the IV estimators, we can encounter weak instruments problems. Propositions 2 and 3 show that we have a precise estimator if there is high concentration and if the idiosyncratic shocks are volatile.

Hence, before conducting a large-scale study and large-scale data collection effort, we recommend doing an ex ante *power analysis*. Recall that in the most basic case, the standard error of the estimator is $s.e.\left(\phi_T^{s,e}\right) = \frac{\sigma_\varepsilon}{\mu h \sigma_u \sqrt{T}}$ (see Proposition 3). To get a quick sense of the power of GIVs, we recommend estimating σ_ε using the volatility of the left-hand side variable as order of magnitude, σ_u using $\sigma_{\check{y}_i}$ so that a simple common factor is removed as well as the component that depends on prices, and the average excess Herfindhal h. These inputs can be used to calibrate an order of magnitude for $\sigma_{\phi_T^{s,e}}$, given the researcher's prior view on μ. This gives a sense of the t-statistics $t = \frac{\phi^s}{\sigma_{\phi_T^{s,e}}}$ that can be obtained with T periods. Hence, before refining the empirical model and perhaps collecting additional data, one has a sense of whether the GIV will be sufficiently powerful. We also recommend reporting the results of such a power analysis alongside the final estimates. We illustrate this procedure in our empirical example, Section 5.4.

Assuming that GIVs are sufficiently powerful, the most important threat to identification is that we do not control properly for common factors. Indeed, $z_t = u_{\Gamma t} + \lambda_\Gamma \eta_t - \lambda_\Gamma^e \eta_t^e$, so there is a danger that, even after controlling for η_t^e in the regression we will not completely eliminate the $\lambda_\Gamma \eta_t - \lambda_\Gamma^e \eta_t^e$ error.[15] This danger is greater when $|\lambda_\Gamma|$ is greater, that is, when loadings are correlated with size. Indeed, omitting factors for which $\lambda_\Gamma = 0$ is inconsequential. This is a small sample problem as we measure η_t and λ accurately when N and T are sufficiently large. We provide three concrete suggestions to mitigate this concern.

First, we can use a "narratively checked" GIV, as we do in our empirical application below. If one checks the top, say, 10 events, and see if they they pass the narrative check, and construct the GIV based only on those narrative events.[16]

Second, we can omit key dates that may be exceptional and that can give rise to "sporadic factors," that is, a factor η_t that affects many entities in unusual ways. For instance, there may be an important policy announcement in the sample. If this happens once in the sample, it is hard to detect using standard factor models.

Third, we can do an overidentification test. For instance, one could construct two GIV based on two types of entities, z_{1t} (respectively z_{2t}) based on the size-weighted sum of idiosyncratic shocks of

[15] As we do control for η_t^e in the regression, the bias is due to the residual of $\lambda_\Gamma \eta_t - \lambda_\Gamma^e \eta_t^e$ after controlling for η_t^e.

[16] This is roughly what the "narrative" approach in the literature (see, for instance, Caldara et al. (2019)) does. But the GIV procedure does help researchers even in the narrative context, since it automates the "pre-selection" of the top K (perhaps $K = 15$) shocks, by selecting the events with the largest K values in $S_i\,|\check{u}_{it}|$. Hence, researchers don't need to know the whole history before selecting their main events – the GIV gives them the most promising candidate events, and the detailed historical search is simply restricted to K events. In addition, the factor analysis in the GIV gives controls η_t^e that are usable when running regressions, which increases the precision of estimators.

e.g. poor (respectively rich) countries, and see if the multipliers that they lead to are statistically indistinguishable. Such a test is obviously only meaningful if it is powerful. If the two estimates are significantly different, it can mean that the factor model is misspecified.

A related concern is that the factor loadings are unstable. In this case, it is possible to use more advanced methods to extract factors, for instance when $\lambda_{it} = X_{it}\dot{\lambda}$, as we discuss in Section 4.4.

What is an idiosyncratic shock? Mathematically, an idiosyncratic shock is plainly a random variable u_{it} such that $\mathbb{E}_{t-1}[\tilde{\eta}_t u_{it}] = 0$, where $\tilde{\eta}_t := (\eta_t, \varepsilon_t)$ includes all the common shocks. But it may be useful to discuss different types of economic settings that map into that definition. In some cases it is quite clear – for example, a random productivity or demand shock. But there are more subtle types of idiosyncratic shocks. One is an "unexpected change in the loading on a common shock." For instance, if China decreases oil consumption by more than anticipated in response to a global economic downturn, then it is an idiosyncratic shock. Formally, if demand is $y_{it} = \phi^d p_t + \left(\lambda_i + \tilde{\lambda}_{it}\right)\eta_t + v_{it}$, with $\mathbb{E}_{t-1}\left[(1, \tilde{\eta}_t)\,\tilde{\lambda}_{it}\right] = 0$, then $u_{it} := \tilde{\lambda}_{it}\eta_t + v_{it}$ is a bona fide idiosyncratic shock.

The volatility of idiosyncratic shocks can depend on the common shocks. For instance, suppose that $u_{it} = \sigma_t v_{it}$ where σ_t and $\tilde{\eta}_t$ could be correlated (for instance, σ_t could increase when $|\tilde{\eta}_t|$ is high), but $\mathbb{E}_{t-1}[\tilde{\eta}_t \sigma_t v_{it}] = 0$ (a sufficient condition is that v_{it} independent of $\sigma_t \tilde{\eta}_t$); then, u_{it} is an idiosyncratic shock in the sense that $\mathbb{E}_{t-1}[\tilde{\eta}_t u_{it}] = 0$.

4 Extensions

We now present a succession of extensions of the basic GIV procedure that cover a range of cases that one may encounter empirically. For expositional clarity, we continue with our supply and demand example, although this can be vastly generalized, see Section G of the Online Appendix. We provide extensions to the user's guide in Appendix B.

4.1 Time-varying size weights

We first consider the case in which the size weights vary over time, $S_{i,t-1}$, so that demand by country i at date t is $D_{it} = \bar{Q} S_{i,t-1}(1 + y_{it})$ In that case, the aggregate demand disturbance is $\sum_i S_{i,t-1}(1 + y_{it})$. The assumption required for identification then changes to $\mathbb{E}\left[u_{it}(\eta_t, \varepsilon_t)'(1, S_{i,t-1})\right] = 0$. If this condition is satisfied, all our derivations go through by replacing S_i by $S_{i,t-1}$.

4.2 Heterogeneous demand elasticities

We have assumed so far that demand elasticities are constant across entities. We now extend the model so that the elasticities are linear in the characteristics X_i (an r-dimensional vector, with r in practice a small number, and the first entry being 1), so that they can be expressed as:[17]

$$\phi_i^d = X_i \dot{\phi}^d = \sum_{\ell=1}^{k} X_{i\ell} \dot{\phi}_\ell^d, \tag{29}$$

for some $\dot{\phi}^d = \left(\dot{\phi}_\ell^d\right)_{\ell=1\ldots k}$ to be determined. With X the $N \times k$ matrix of characteristics, this is saying that we assume the parametric forms $\phi^d = X\dot{\phi}^d$ and $\Lambda = X\dot{\lambda}$ where $\dot{\phi}^d$ and $\dot{\lambda}$ have dimension $k \times 1$ and $k \times r$, recalling that η_t has dimension $r \times 1$. For expositional simplicity, we take $k = r$ but with more notations one could similarly handle the case $k \neq r$. The following proposition describes how we can consistently estimate ϕ^s and $\dot{\phi}^d$.

Proposition 7 (Estimation of heterogeneous parametric elasticities). *Consider the model with heterogeneous elasticities following* (29). *Define* $R := (X'X)^{-1}X'$, $Q := I - XR$, $\dot{y}_t := Ry_t$ *(which has dimension* k*),* $\check{u}_t := Qy_t$*, and* $z_t := \sum_i S_i \check{u}_{it}$*. We then identify* ϕ^s *and* $\dot{\phi}^d$ *using the IV moments:* $\mathbb{E}\left[(s_t - \phi^s p_t)\, z_t\right] = 0$ *and* $\mathbb{E}\left[\left(\dot{y}_{\ell t} - \dot{\phi}_\ell^d p_t\right) z_t\right] = 0 \qquad$ *for* $\ell = 1 \ldots k$*.*

The concrete procedure is that for each date t we run the cross-sectional regression of y_{it} on X_i:

$$y_{it} = X_i \dot{y}_t + \check{u}_{it} = \sum_{\ell=1}^{k} X_{i\ell} \dot{y}_{\ell t} + \check{u}_{it}, \tag{30}$$

which yields regression slopes $\dot{y}_t = (\dot{y}_{\ell t})_{\ell=1\ldots k}$ and residuals $\check{u}_{it}$. Then, we form the GIV $z_t := \sum_i S_i \check{u}_{it}$, and finally use the moment conditions in Proposition 7 to recover the elasticities.

4.3 Heteroskedasticity

We now discuss the case where the u_{it} are heteroskedastic. We call V^u their variance-covariance matrix. We assume for now that this is known, at least up to a factor of proportionality.

Proposition 8 *(GIV with heteroskedastic idiosyncratic shocks) Consider the case where the* u_{it} *are heteroskedastic with variance-covariance matrix* V^u*. Define* $W = (V^u)^{-1}$ *as the inverse variance-covariance matrix of the idiosyncratic shocks* u_{it}*, and define two matrices, with respective dimensions and* $N \times N$ *and* $r \times N$*:*

$$Q^{\Lambda,W} := I - \Lambda R^{\Lambda,W}, \qquad R^{\Lambda,W} := \left(\Lambda' W \Lambda\right)^{-1} \Lambda' W. \tag{31}$$

so that $Q^{\Lambda,W}$ *is a projection on the space orthogonal to* Λ *and* $R^{\Lambda,W}$ *is a projection on* Λ*.*[18]

[17] Section H.4 considers a non-parametric version, which is more involved.

[18] They have a number of good properties that we record here (dropping the superscripts for simplicity):

$$Q\Lambda = 0, \quad R\Lambda = I, \quad Q'W\Lambda = 0, \quad (I - Q)\, W^{-1} Q' = 0, \quad (I - Q')\, WQ = 0, \quad Q^2 = Q, \quad RW^{-1}Q' = 0. \tag{32}$$

Then Propositions 1-7 hold provided that we replace the orthogonality condition $\Gamma'\Lambda = 0$ by its weighted version $\Gamma' W \Lambda = 0$, replace Q, and R by $Q^{\Lambda,W}$ and $R^{\Lambda,W}$ (e.g., we have $\Gamma^{*\prime} = S' Q^{\Lambda,W}$ in (12), (18), (19)*), replace E by the quasi-equal weights $\tilde{E} := \frac{W\iota}{\iota' W \iota}$ where ι is a vector of ones (e.g. in* (10), (18)*), and that cross-sectional regressions use GLS with weights W instead of OLS (e.g. in* (15), (24), (30)*).*

When idiosyncratic shocks are homoskedastic, $\tilde{E}_i = \frac{1}{N}$, while if they are uncorrelated but heteroskedastic (i.e. $W = Diag\left(1/\sigma^2_{u_i}\right)$), we have $\tilde{E}_i := \frac{1/\sigma^2_{u_i}}{\sum_j 1/\sigma^2_{u_j}}$, so that $\tilde{E}$ may be called the "precision-weighted quasi-equal" weights.[19,20]

4.4 PCA and IPCA

We assumed so far that common factors can be modeled as $\Lambda\eta_t = \sum_{f=1}^r \lambda^f \eta_t^f$ with $\Lambda = X\dot{\lambda} \in \mathbb{R}^{N\times r}$, $X \in \mathbb{R}^{N\times k}$, $\dot{\lambda} \in \mathbb{R}^{k\times r}$, $\eta_t \in \mathbb{R}^{r\times 1}$, and $r = k$, and provide proofs for the case. We now discuss three extensions. While we do not provide proofs for these cases, we explore the most general case (PCA) in simulations in Section F.

First, we can allow for characteristics that change over time, X_t instead of X. In this case, we can form $Q_t = I - X_t(X_t'X_t)^{-1}X_t'$, and compute the residuals as $\check{u}_t = Q_t y_t$. We then proceed as before with one exception. If we add η_t^e as a control, which is not needed but adds precision, we need to add $X_{Et}\eta_t^e$ or $X_{St}\eta_t^e$ as controls instead of only η_t^e.

Second, we can consider the case in which $k > r$, implying that we can model the loadings more flexibly as a function of a larger number of characteristics. Kelly et al. (2020) refer to this model as Instrumented Principal Components Analysis (IPCA) and they develop the asymptotic theory.

Third, we can consider the case where we impose no structure on the loadings and we use PCA to estimate the loadings and the factor realizations. The asymptotic theory for PCA has been developed in Bai (2003) and in the context of GIV by Banafti and Lee (2022). To compute standard errors, we can use the bootstrap and we explore this procedure in simulations in Section F. We provide a detailed discussion in Appendix B as part of the user's guide.

It is also possible to first extract factors using loadings that depend on observed characteristics, $\eta_t^{x,e}$, and then estimate additional factors using PCA on the residuals, $\eta_t^{PCA,e}$. We then use the final residuals in constructing the GIV and use $\eta_t^e = \left(\eta_t^{x,e}, \eta_t^{PCA,e}\right)$ as factors.

[19] If we misspecify the variance of the u_{it} (but keeping them uncorrelated), the impact on the estimates is typically small: as $u_E = O_p\left(\frac{1}{\sqrt{N}}\right)$, we do not need $\mathbb{E}\left[u_{\Gamma t} u_{Et}\right] = 0$ to hold exactly, as the term $\mathbb{E}\left[u_{\Gamma t} u_{Et}\right]$ will still be small, of order $O\left(\frac{1}{\sqrt{N}}\right)$, and will vanish for large N.

[20] Generically, $\Gamma^* \neq 0$. But in cases where the variance is inversely proportional to size, we would have $\Gamma^* = 0$ and the GIV would fail (and this would be detected in practice via extremely large standard errors). Indeed, we would then have $V^u S = a\iota$ for some scalar a, so that $S' Q^{\Lambda,(V^u)^{-1}} = 0$. Fortunately, in most contexts, variance may decay a bit with size S_i, but less violently than in $1/S_i$ (see e.g. Lee et al. (1998) and the discussion in Gabaix (2011)).

When using more flexible factor models, one can estimate the required number of common factors (Bai and Ng (2002); Onatski (2009)). In addition, a missing factor may be detected by testing the stability of estimates across GIVs as we add more factors. As is common practice in the weak factors literature, one can verify the stability of the estimates by adding one or two factors beyond what is estimated by formal procedures for the number of factors. We do this in our empirical example below.

4.5 Fat-tailed shocks and outliers

We have discussed in Section 2.3 how the precision of the GIV estimator increases in the volatility of idiosyncratic shocks. Hence, large idiosyncratic shocks are generally beneficial to GIVs. That said, when the idiosyncratic shocks are fat-tailed or there are large outliers, it can affect the estimation of the common factors, and this can lead to noisier estimates. We can adapt standard methods that have been developed for the robust estimation of covariance matrices by Fan et al. (2019) and Fan et al. (2021), such as estimating means using the Huber loss function and winsorizing data, to our setting.

We provide a detailed discussion in Appendix B as part of the user's guide. We implement those procedures in the simulations in Section F, and we show how fat-tailed idiosyncratic shocks are helpful in obtaining more accurate estimates using GIVs compared to the case where the shocks are thin-tailed. Hence, when handled with care, fat-tailed idiosyncratic shocks are a blessing for GIVs.

4.6 Evidence from simulations

All our results up to and including Section 4.3 are rigorously proven. We now explore some of the extensions in this section, for which we do not provide proofs, using simulations. Specifically, we explore a model with non-parametric factors (Section 4.4), fat-tailed idiosyncratic shocks (Section 4.5), and heteroskedastic shocks with unknown volatilities that need to be estimated (Section 4.3). In the implementation, we follow the user's guide in B. While the simulations indicate that GIVs work in these settings, we leave formal proofs to future research.

The results in Section F show that when shocks are normally distributed yet the loadings are unknown, the baseline GIV algorithm works well. When shocks are fat-tailed, the standard procedure is unbiased but leads to large standard errors. Intuitively, the estimated factor loadings are noisy, which adds to the estimation error. However, when we follow the robust procedure in B, the standard errors are in fact smaller compared to the case with thin-tailed Gaussian shocks. The reason is that the procedure removes the impact of outliers in estimating the factors, and takes advantage of the large idiosyncratic shocks in estimating the parameters of interest.

We also show that when we estimate the volatilities of the idiosyncratic shocks, even though the shocks are homoskedastic, the estimates are hardly affected. Hence, in our simulations, it does not hurt to use a procedure that takes into account potential heteroskedasticity. If the shocks are heteroskedastic, and we do not account for this, the estimates are biased. However, in the presence of fat-tailed shocks and in particular when we estimate the volatilities, the estimator is again unbiased. These examples illustrate the performance of the GIV estimator in realistic settings and how the generalized procedure in Section B can be used to estimate the model's parameters.

4.7 Generalization of the GIV to other setups

While we focus on the demand and supply setup for our main analysis, one can extend the GIV idea to many settings and generalizations.

Multidimensional GIV One can handle multidimensional "outcomes" y_{it} and shocks u_{it}, for instance, a firm i could have two-dimensional shocks u_{it} to both productivity and labor demand. We show how GIVs can be used in this setting in Section H.1.

Beyond supply and demand for one good One can study equilibrium models that are significantly richer than with just supply and demand for one good, see Section G, which features multiple general equilibrium channels.

GIV with a more complex matrix of influences The GIV can also be extended to non-homogeneous influences in the context of networks. Consider a model $y_{it} = \gamma \sum_j G_{ij} y_{jt} + \lambda_i \eta_t + u_{it}$ with a known network structure G_{ij}, and we would like to estimate the influence factor γ. Section H.16 shows formally how this is possible— we again use the idiosyncratic shocks as instruments. Hence, GIV generalizes to "spatial" models with common shocks, which are often omitted in spatial models.[21] Similarly, GIVs allow a way to identify the influence parameter γ in the "reflection problem" (Manski (1993)).[22] The traditional literature on the reflection problem does not use idiosyncratic shocks in its identification strategies. Hence a GIV approach can be useful to complement existing approaches. Section H.17 develops this.[23]

Estimating structural vector autoregressions with GIVs One can do vector autoregressions and impulse responses with GIVs: if $Y_t = AY_{t-1} + X_t$, one can use the GIV z_t to instrument for some of the shocks to the innovations X_t, and achieve partial or full identification. The GIV is then an "external instrument", and one can follow the methods spelled out in Stock and Watson (2018).[24] One can also do Jordà (2005) style local projections, regressing $\mathbb{E}_t [Y_{t+h}] = \beta_h X_t$, and

[21] See Brownlees and Mesters (2021) for a potential way to extend this approach when G is unknown.

[22] In the simplest case where $G_{ij} = S_j$, this is identifying γ in a model $y_{it} = \gamma y_{St} + \lambda_i \eta_t + u_{it}$. This model is a special case of our supply and demand model, replacing $\phi^d p_t$ by γy_{St}, i.e. setting $s_t = p_t := y_{St}$, $\phi^s = 1$, $\varepsilon_t = 0$, $\phi^d = \gamma$. Hence our results show how the GIV identifies γ.

[23] Somewhat related, Graham (2008) explores the identification of peer effects using conditional variance restrictions on the outcomes by exploiting differences in the sizes of the peer group. Intuitively, smaller peer group sizes lead to a larger contribution of each individual peer on the peer component.

[24] See Plagborg-Møller and Wolf (2021) for a recent development in this area.

instrumenting some of the regressors X_t by a GIV. This shows how GIVs can be used to identify parameters in structural VARs, complementing an active literature that uses sign restrictions, as in Uhlig (2005), or narrative restrictions combined with sign restrictions, as in Antolín-Díaz and Rubio-Ramírez (2018) and Ludvigson et al. (2020).

Comparison with Bartik instruments Bartik (1991) instruments are widely used in economics and we discuss the link between Bartik and GIV in Online Appendix H.5, and clarify the difference in identifying assumptions using the econometric framework of Borusyak et al. (2022).[25] Here we briefly highlight some points in that discussion. First, in a number of cases where a cross-section is used (e.g. Autor et al. (2013)), Bartik applies, but GIV does not apply, for instance because there is no large idiosyncratic shock that one can use. On the other hand, in a number of cases (such that the ones we have discussed in this paper), GIV applies naturally, particularly when there are large idiosyncratic shocks that affect the aggregates. Hence GIVs complement the toolkit of economists, and depending on the setting, Bartik instruments or GIVs are more appropriate.

When aggregate shocks are at least partially made of idiosyncratic shocks GIVs extend to economies where aggregate shocks η_t are themselves at least partially made of idiosyncratic shocks u_{it}, as in Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019)) – provided some suitable modifications of our basic assumption (5). We develop this in Sections H.13–H.16. These sections show that we can identify important parameters even if we have only crude proxies for the primitive shocks such as TFP.

5 Empirical application: Estimating sovereign yield spillovers

We study spillovers in sovereign yield markets in the Euro area as an application of GIV. We are interested in the transmission and amplification of idiosyncratic shocks during the European sovereign debt crisis.

5.1 An empirical model of sovereign yield spillovers

Section E provides a microfounded economic model of sovereign yield spillovers, and we summarize its empirical implications here. In that model, governments may default on their debt, and losses in one country will be partially shared with other countries, implying that shocks to sovereign yields in one country spill over to other countries.

We index countries by i. We define the yield spread, y_{it}, as the yield in country i relative to Germany's yield. Note that y_{it} is not the demand shifter in our initial example. The economic model

[25] Also known as shift-share estimators, they have been the study of much recent econometric work, see also Goldsmith-Pinkham et al. (2020); Adao et al. (2019); Borusyak et al. (2022).

implies that relative changes in yield spreads, $r_{it} := \frac{\Delta y_{it}}{y_{i,t-1}}$, should satisfy the following empirical model

$$r_{it} = a_i + \gamma r_{St} + \lambda_i' \eta_t + u_{it}, \tag{33}$$

where γ is a contagion parameter that we wish to estimate, η_t are aggregate shocks, u_{it} idiosyncratic shocks, and the size weights are, with B_{it} the outstanding government debt of country i:

$$S_{i,t-1} = \frac{B_{i,t-1} y_{i,t-1}}{\sum_j B_{j,t-1} y_{j,t-1}}. \tag{34}$$

This structural model brings two lessons to the empirics. First, the proper "size" of country i here is its "debt at risk," $B_{i,t-1} y_{i,t-1}$, the expected euro loss on its debt.[26] Second, the spillover impact is such that $\frac{\Delta y_{it}}{y_{i,t-1}}$, rather than Δy_{it}, depends linearly on γr_{St}. This means that a country with almost no default risk ($y_{i,t-1} \simeq 0$) should have almost no sensitivity of its yield Δy_{it}, as there is no risk in the first place. This intuition is likely to hold in alternative models, and those models will then imply a similar functional form.[27] It is essential to control for common factors η_t, as we do below. It is well understood, see for instance Forbes and Rigobon (2002), that omitted factors and endogeneity impact measures of spillovers and contagion.[28]

5.2 Data

We use daily data on 10-year zero coupon yields from Bloomberg. We discuss the data in detail in Appendix E.4. We use data on general government gross debt for each country from Eurostat. The full sample is from July 1, 2009 to May 31, 2018, and the multiplier is estimated from September 1, 2009 to May 31, 2018. The earlier period is used to estimate the volatility of yield changes as we discuss below. The countries included are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Slovenia, and Spain. As spreads are computed relative to Germany, we have 11 countries in the main analysis.

5.3 Estimation procedure

We estimate the model using the standard GIV procedure, accounting for heteroskedasticity. Empirically, we use $r_{it} = \frac{\Delta y_{it}}{0.01 + y_{i,t-1}}$ to avoid problems when spreads, $y_{i,t-1}$, get close to zero.

1. We compute the rolling standard deviation of relative changes in yield spreads using the 60 most recent observations, including the current date. We refer to this estimate as $\sigma_t(r_{it})$. We

[26] This is under the risk-neutral measure, i.e. adjusting for the price of risk.

[27] Spillovers in sovereign bond markets may also operate via intermediaries. For instance, if losses in one country impact the intermediaries' constraints, then this can impact the pricing of bonds in other countries in which the intermediaries are active. We strongly suspect that a functional form like (33) would still hold in such a model.

[28] Caporin et al. (2018) study spillovers in European sovereign debt markets and show that quantile regressions can be used to test for contagion if contagion is defined as a change in interlinkages. Our definition of contagion (captured by a nonzero γ in equation (33)) is different from theirs.

then define $\sigma_{it} = \max\left(\sigma_t\left(r_{it}\right), m_t\right)$, where $m_t = \text{median}_i\left(\sigma_t\left(r_{jt}\right)\right)$, that is, the cross-sectional median at time t. We define the quasi-equal weights as usual as $\tilde{E}_{i,t-1} = \frac{1/\sigma_{it}^2}{\sum_i 1/\sigma_{it}^2}$. We apply the max operator in the construction of σ_{it} to avoid that the quasi-equal weights $\tilde{E}$ put too much weight on a single country if yield spreads for that country happen to be stable and close to zero. The main objective of adjusting for heteroskedasticity is to put less weight on extremely volatile countries.[29]

2. We compute $\check{r}_{it} \coloneqq r_{it} - r_{\tilde{E}t} = \check{a}_i + \check{\lambda}_i'\eta_t + \check{u}_{it}$, where $r_{\tilde{E}t} = \sum_i \tilde{E}_{i,t-1} r_{it}$. We select our main sample (September 1, 2009 to May 31, 2018) and estimate the factors, η_t, by solving

$$\min_{(\check{a}_i, \check{\lambda}_i, \eta_t)} \sum_{i,t} \left(\frac{\check{r}_{it} - \check{a}_i - \check{\lambda}_i'\eta_t}{\sigma_{it}}\right)^2,$$

using alternating least squares.[30] Specifically, solving the first-order conditions implies:

$$\begin{aligned}
\check{a}_i &= \frac{\sum_t \left(\check{r}_{it} - \check{\lambda}_i'\eta_t\right)/\sigma_{it}^2}{\sum_t 1/\sigma_{it}^2}, \\
\check{\lambda}_i &= \left(\sum_t \eta_t\eta_t'/\sigma_{it}^2\right)^{-1} \sum_t \eta_t \left(\check{r}_{it} - \check{a}_i\right)/\sigma_{it}^2, \\
\eta_t &= \left(\sum_i \check{\lambda}_i\check{\lambda}_i'/\sigma_{it}^2\right)^{-1} \sum_i \check{\lambda}_i \left(\check{r}_{it} - \check{a}_i\right)/\sigma_{it}^2,
\end{aligned}$$

which we solve recursively until convergence.[31] We refer to the estimated factors as η_t^e.

3. We estimate the multiplier $M = \frac{1}{1-\gamma}$ via the regression (with $\tilde{\Gamma}_t = S_{t-1} - \tilde{E}_{t-1}$,so $r_{\tilde{\Gamma}t} = r_{St} - r_{\tilde{E}_{t-1}}$)

$$r_{St} = k + M r_{\tilde{\Gamma}t} + \lambda_S'\eta_t^e + e_t. \tag{35}$$

To identify the largest shocks and to verify narratively that the shocks are truly idiosyncratic, we regress $r_{\tilde{\Gamma}t}$ on the factors, $r_{\tilde{\Gamma}t} = c + \beta'\eta_t^e + u_{St}^e$, and we analyze the days with the largest $|u_{St}^e|$ in detail in Table 2.

5.4 Empirical results

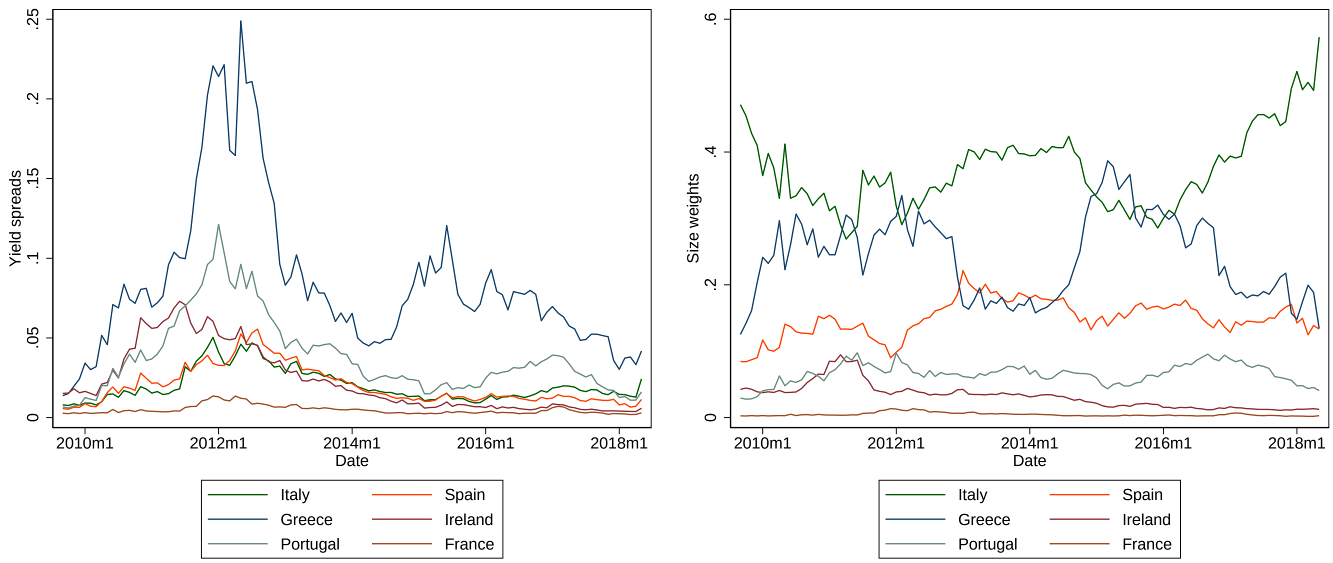
We plot the dynamics of spreads, y_{it}, in the left panel, and size weights, using the definition in (34), in the right panel of Figure 1 for France, Greece, Ireland, Italy, Portugal, and Spain. The sample

[29]The results are quantitatively very similar when using $\sigma_{it} = \max\left(\sigma_t\left(r_{it}\right), \kappa m_t\right)$, with $\kappa = 0.75$.

[30]We initialize the algorithm by running PCA on $\check{r}_{it}/\sigma_{it}$ to get starting values for $\check{\lambda}_i$ and η_t.

[31]In each step, we normalize $\check{\Lambda}$ as it is identified only up to a rotation. Concretely, we compute $\check{\Lambda}'\check{\Lambda} = LL'$, where L is a lower triangular matrix, after updating $\check{\lambda}_i$ for all i. We then replace $\check{\Lambda}$ by $\check{\Lambda}\left(L'\right)^{-1}$ before estimating η_t.

Figure 1: Dynamics of sovereign yield spreads and size weights. The figure reports the yield spreads, relative to Germany, for Italy, Spain, Greece, Ireland, Portugal, and France in the left panel from September 1, 2009 to May 31, 2018. The spreads are based on 10-year government yields and are constructed using data from Bloomberg. The right panel displays the size weights based on the definition in (34) for the same countries and the same sample period.



is from September 1, 2009 to May 31, 2018. We distinguish three broad periods. First, from 2010 to 2012, the yield spread dynamics are driven by the European sovereign debt crisis. During 2015, yield spreads in Greece widen once again, but the low-frequency dynamics in other countries are more muted and spreads tighten in most countries. This period is characterized by political turmoil in Greece related in part to negotiations of a bailout deal. During the last months of our sample, there is a jump in Italian yields due to political uncertainty regarding budget plans following the general election. We will revisit these episodes in more detail when analyzing the largest and most influential idiosyncratic shocks in Table 2.

Before conducting a large scale analysis (or collecting much data), it is useful to do a rough ex ante power analysis to see if the standard errors can be hoped to be tight enough (see p.14). We illustrate that here. If there are no common factors, $\eta_t = 0$, and in the absence of heteroskedasticity, we can approximate the standard errors by

$$s.e.(M) \simeq \frac{\sigma(r_{St})}{h\sigma(\check{r}_{it})\sqrt{T}} = \frac{0.020}{0.38 \times 0.015 \times \sqrt{2283}} = 0.072,$$

where $h = \sqrt{\sum_i \bar{S}_i^2 - 1/N}$, $\bar{S}_i = \frac{1}{T}\sum_t S_{it}$. This number is likely a conservative estimate as we assume that $\check{r}_{it}$ is homoskedastic, which is not the case empirically. This power calculation suggests that there will be enough power to obtain sufficiently accurate estimates of M to be economically informative.

Table 1: Multiplier estimates of sovereign yield spillovers. The table reports the estimates of the multiplier in (35). The first to the third column include zero to two principal components (see the η_{kt} rows) to isolate idiosyncratic shocks. In the fourth column, we set u_{St} to zero for all periods, except for the 10 days on which $|u_{St}|$ is largest. In Table 2, we provide a narrative analysis of those 10 shocks. In the fifth column, we also set the shock on November 11, 2011 to zero, as the shock on this date may be a sporadic common shock. In the final column, we re-run the analysis, but excluding Greece. The model is estimated using daily data from September 2009 until May 2018. Standard errors are in parentheses.

				Narratively-checked shocks		
	Baseline			Top 10 shocks	Excluding sporadic factors	Excluding Greece
M	1.641	1.492	1.417	1.457	1.479	1.595
	(0.0225)	(0.0367)	(0.0362)	(0.0896)	(0.0917)	(0.0402)
η_{1t}		0.0828	0.1255	0.5944	0.5940	0.0590
		(0.0162)	(0.0157)	(0.0111)	(0.0112)	(0.0178)
η_{2t}			-0.202	-0.252	-0.256	-0.029
			(0.0103)	(0.0128)	(0.0128)	(0.0090)
T	2283	2283	2283	2283	2283	2283
R^2	70%	70%	75%	62%	62%	74%

Table 1 reports the estimates of the multiplier, M. The first column regresses r_{St} on $Z_t = r_{\tilde{\Gamma}t}$. The second and third column add principal components. The multiplier estimate drops after adding the first principal component from 1.64 to 1.49, but adding the second principal component does not change the multiplier much. Given the small cross-section (there are only 11 countries), we cannot include additional factors beyond a time fixed effect and two principal components.

The high R-squared in the first column does not estimate the fraction of the variation in aggregate yield spread changes that is due to idiosyncratic shocks, as $r_{\tilde{\Gamma}t}$ is correlated with η_t^e. To estimate the importance of idiosyncratic shocks, we regress r_{St} on u_{St}^e, which provides exactly the same point estimate of the multiplier as in the final column of Table 1. The R-squared of this regression is 17%, implying that 17% of the variation in aggregate yield spread changes is due to idiosyncratic shocks.

To further inspect the variation that the GIVs are exploiting to estimate the multiplier, we narratively check the largest shocks in Table 2. In particular, we order the dates based on the size of $|u_{St}^e|$. In Panel A of Table 2, we report r_{it} for each of the countries. In Panel B, we provide the narratives. If we inspect some of the largest shocks in Table 2, then is is clear that most of them are plausibly idiosyncratic shocks. Examples include the decision by Greece to close all banks or the outcome of the referendum. That said, it is obviously challenging to remove aggregate shocks during times of financial turmoil. For instance, the shock on November 11, 2011 is harder to assign to a single country and may be best interpreted as a sporadic common shock as several spreads

Table 2: Summary of the largest idiosyncratic shocks and narratives. The table reports relative yield changes, $r_{it} = \frac{\Delta y_{it}}{0.01+y_{i,t-1}}$ in Panel A. In Panel B, we provide narratives associated with these events.

Panel A: Main shocks

Date	Austria	Belgium	Finland	France	Greece	Ireland	Italy	Netherlands	Portugal	Slovenia	Spain
12-Mar-2012	1.9%	1.6%	0.8%	2.5%	-31.7%	0.0%	2.8%	0.9%	0.5%	-0.9%	2.2%
29-Jun-2015	3.8%	5.0%	2.8%	5.1%	33.4%	6.9%	16.6%	3.1%	17.4%	4.7%	16.6%
10-Jul-2015	-3.0%	-3.5%	-2.2%	-3.7%	-22.1%	-5.0%	-8.8%	-3.0%	-7.6%	-9.1%	-9.0%
16-Jan-2012	-1.6%	-1.3%	-3.5%	-1.9%	-0.8%	1.0%	-0.2%	-4.2%	14.3%	-1.1%	-0.5%
29-May-2018	5.2%	2.7%	2.3%	2.9%	7.2%	5.1%	14.6%	2.3%	7.3%	4.7%	7.7%
06-Jul-2015	1.8%	2.5%	0.7%	2.0%	18.9%	3.3%	6.8%	1.2%	8.6%	1.1%	7.7%
08-Aug-2011	-1.9%	-3.1%	0.2%	4.4%	0.1%	0.6%	-14.4%	-0.3%	-1.2%	-4.0%	-15.9%
03-Feb-2015	-1.2%	-0.7%	-0.7%	-0.1%	-11.5%	-1.2%	-3.0%	-0.7%	-3.7%	-2.9%	-2.0%
13-Jul-2015	-0.5%	1.0%	-0.2%	1.1%	-9.6%	0.1%	1.0%	0.3%	-0.9%	-0.5%	1.1%
11-Nov-2011	-3.5%	-3.9%	-0.7%	-7.4%	0.2%	-2.3%	-8.6%	0.4%	-0.4%	3.5%	-2.5%

Panel B: Narrative analysis

Date	Event
12-Mar-2012	Greece Bailout Package Signed Off by EU Leaders
29-Jun-2015	Greece closes banks
10-Jul-2015	The Greek government submitted its highly anticipated plan for the country's economic overhaul to bailout authorities
16-Jan-2012	Portugal is downgraded to junk status by Standard and Poor's before the weekend
29-May-2018	Italian political turmoil (snap election plus new budget plan) cause largest 1-day decline in Italian bonds in 25 years
06-Jul-2015	Greece bailout referendum on July 5th where voters reject austerity package
08-Aug-2011	ECB decides to start buying Italian and Spanish bonds as part of the Securities Markets Program
03-Feb-2015	Greek government said to retreat from a demand for a debt writedown
13-Jul-2015	Greek PM Alexis Tsipras conceded to a further swathe of austerity measures and economic reforms
11-Nov-2011	Positive news about Italy's government and/or ECB purchase program (see main text)

move at the same time.[32] In the absence of a clear narrative, the prudent approach may be to explore the sensitivity of the results to that observation as we do below.

In the fourth column of Table 1, we regress r_{St} on u_{St} and two principal components, but we only use the 10 largest values of u_{St} and set all others to zero. This leads to a multiplier of 1.46 with a standard error of 0.09. While this estimate is less precise, the estimate is still informative. As the shock on November 11, 2011 may be a sporadic common shock, we re-run the analysis once more and now setting u_{St} also to zero on November 11, 2011. In this case u_{St} is non-zero on nine days only, and all of those shocks are narratively verified in Table 2. The estimated multiplier equals 1.48 with a standard error of 0.09. As a final robustness check, we repeat the analysis omitting Greece. The results are presented in the final column. Using the shocks from other countries leads to a similar multiplier at 1.6.

[32] Media narratives discuss increased confidence in Italy's government by voting in favor of austerity measures and due to the growing support for Monti to take over from Berlusconi as prime minister. This would be a valid idiosyncratic shock. However, there are also news reports discussing the potentially important role of the ECB's asset purchase program (the SMP) in supporting bond prices. This would be would a sporadic common shock.

5.5 Interpretation of the coefficients

We find a multiplier $M = \frac{1}{1-\gamma} \simeq 1.5$ and hence a spillover parameter $\gamma \simeq \frac{1}{3}$. The interpretation is as follows:[33] suppose that Italy suffers a bad shock that makes its debt likelier to default, so that the market value of Italy's debt falls by 1 billion euros. The multiplier $M \simeq 1.5$ means that the aggregate debt of all European governments falls by 1.5 billion euros – the spillover consists of an extra 0.5 billion euros in expected losses in European sovereign debt markets.[34] Note that the expectation is under the risk-neutral measure, so could correspond to a higher likelihood of default, or a higher price of risk for that default.[35]

6 Conclusion

We developed granular instrumental variables (GIVs). The generative insight is that idiosyncratic shocks offer a rich source of instruments. We lay out econometric procedures to extract them from panel data and optimally aggregate them to obtain the most powerful instruments. We provided an empirical application to illustrate the implementation of GIVs.

We discuss various econometric extensions that might be useful to explore in future research. We hope that GIVs will aid identification in new settings and help researchers investigate and understand causal relationships in the economy.

References

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

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Amiti, Mary and David E Weinstein, "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data," *Journal of Political Economy*, 2018, *126* (2), 525–587.

[33] Here we use, omitting aggregate shocks, $r_{it} = \gamma M u_{St} + u_{it}$, with $r_{it} = \frac{\Delta y_{it}}{y_{i,t-1}}$, and $r_{St} = M u_{St}$.

[34] Indeed, suppose that country i (Italy) has an idiosyncratic shock u_{it}, so that that value of its debt falls by $V_{it} = B_{i,t-1} y_{i,t-1} u_{it}$ euros. As $r_{St} = M S_{i,t-1} u_{it}$, the value of all Eurozone sovereign debt falls by

$$V_t = (\sum B_{j,t-1} y_{j,t-1}) r_{St} = (\sum B_{j,t-1} y_{j,t-1}) S_{i,t-1} M u_{it} = M B_{i,t-1} y_{i,t-1} = M V_{it}.$$

[35] To get some more intuition, consider that Italy, near the peak of the crisis, has a relative size of 0.4. Suppose an idiosyncratic shock to Italy makes the Italian yield spread double ($u_i = 1$ for $i =$ Italy); that is, the Italian yield spread goes from 2% to 4%. That makes the other yield spreads go up by a relative value of $\gamma M \times S_i \times u_i = 0.5 \times 0.4 \times 1 = 0.2$, so that the average yield increases from 1% to 1.20%. In other terms, as the Italian yield spread goes up by 200bp, the other countries' yield spreads go up by 20bp, implying a "pass-through" of 0.1.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International shocks, variable markups, and domestic prices," *The Review of Economic Studies*, 2019, *86* (6), 2356–2402.

Andrews, Isaiah, James H Stock, and Liyang Sun, "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 2019, *11*, 727–753.

Antolín-Díaz, Juan and Juan F Rubio-Ramírez, "Narrative sign restrictions for SVARs," *American Economic Review*, 2018, *108* (10), 2802–29.

Autor, David H, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

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

Bai, Jushan and Serena Ng, "Determining the number of factors in approximate factor models," *Econometrica*, 2002, *70* (1), 191–221.

Banafti, Saman and Tae-Hwy Lee, "Inferential Theory for Granular Instrumental Variables in High Dimensions," *Available at arXiv:2201.06605*, 2022.

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

Bartik, Timothy J, *Who benefits from state and local economic development policies?*, W.E. Upjohn Institute for Employment Research, 1991.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Brownlees, Christian and Geert Mesters, "Detecting granular time series in large panels," *Journal of Econometrics*, 2021, *220* (2), 544–561.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello, "Oil price elasticities and oil price fluctuations," *Journal of Monetary Economics*, 2019, *103*, 1–20.

Camanho, Nelson, Harald Hau, and Hélene Rey, "Global portfolio rebalancing and exchange rates," *Review of Financial Studies*, 2022.

Caporin, Massimiliano, Loriana Pelizzon, Francesco Ravazzolo, and Roberto Rigobon, "Measuring Sovereign Contagion in Europe," *Journal of Financial Stability*, 2018, *34*, 150–181.

Carvalho, Vasco and Xavier Gabaix, "The great diversification and its undoing," *American Economic Review*, 2013, *103* (5), 1697–1727.

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

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

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

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

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

Fama, Eugene F and Kenneth R French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 1993, *33* (1), 3–56.

Fan, Jianqing, Weichen Wang, and Yiqiao Zhong, "Robust Covariance Estimation for Approximate Factor Models," *Journal of Econometrics*, 2019, *208* (1), 5–22.

Fan, Jianqing, Weichen Wang, and Yiqiao Zhong, "A Shrinkage Principle for Heavy-tailed Data: High-dimensional Robust Low-rank Matrix Recovery," *Annals of Statistics*, 2021, *49* (3), 1239–1266.

Forbes, Kristin J. and Roberto Rigobon, "No Contagion, Only Interdependence: Measuring Stock Market Comovements," *The Journal of Finance*, 2002, *57* (5), 2223–2261.

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

Gabaix, Xavier and Ralph SJ Koijen, "In search of the origins of financial fluctuations: The inelastic markets hypothesis," Technical Report, National Bureau of Economic Research 2021.

Galaasen, Sigurd, Rustam Jamilov, Ragnar Juelsrud, Hélène Rey et al., "Granular credit risk," Technical Report, NBER Working Paper No. 27994 2021.

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

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

Graham, Bryan S, "Identifying social interactions through conditional variance restrictions," *Econometrica*, 2008, *76* (3), 643–660.

Huber, Peter J, "Robust estimation of a location parameter," *Annals of Mathematics Statistics*, 1964, *35*, 73–101.

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American Economic Review*, 2005, *95* (1), 161–182.

Kelly, Bryan T, Seth Pruitt, and Yinan Su, "Instrumented principal component analysis," *Available at SSRN 2983919*, 2020.

Kundu, Shohini and Nishant Vats, "Banking networks and economic growth: from idiosyncratic shocks to aggregate fluctuations," Technical Report 2021.

Leary, Mark T and Michael R Roberts, "Do peer firms affect corporate financial policy?," *The Journal of Finance*, 2014, *69* (1), 139–178.

Lee, Youngki, Luís A Nunes Amaral, David Canning, Martin Meyer, and H Eugene Stanley, "Universal features in the growth dynamics of complex organizations," *Physical Review Letters*, 1998, *81* (15), 3275.

Lin, Leming, "Depositing Corporate Payout," *Available at SSRN 3959166*, 2021.

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

Ludvigson, Sydney C, Sai Ma, and Serena Ng, "Shock restricted structural vector-autoregressions," Technical Report, National Bureau of Economic Research 2020.

Ma, Yueran, Teodora Paligorova, and Jose-Luis Peydro, "Expectations and Bank Lending," *Working Paper*, 2021.

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

Montiel Olea, José Luis and Carolin Pflueger, "A robust test for weak instruments," *Journal of Business & Economic Statistics*, 2013, *31* (3), 358–369.

Nakamura, Emi and Jon Steinsson, "Identification in Macroeconomics," *Journal of Economic Perspectives*, August 2018, *32* (3), 59–86.

Onatski, Alexei, "Testing hypotheses about the number of factors in large factor models," *Econometrica*, 2009, *77* (5), 1447–1479.

Plagborg-Møller, Mikkel and Christian K Wolf, "Instrumental variable identification of dynamic variance decompositions," Technical Report, National Bureau of Economic Research 2021.

Ramey, Valerie A., "Macroeconomic Shocks and Their Propagation," in "Handbook of Macroeconomics, John B. Taylor and Harald Uhlig eds," Vol. 2, Elsevier, 2016, pp. 71–162.

Sarto, Andres, "Recovering macro elasticities from regional data," *Unpublished manuscript, Princeton University*, 2018.

Schubert, Gregor, Anna Stansbury, and Bledi Taska, "Monopsony and outside options," *Available at SSRN*, 2022.

Stock, James H and Mark W Watson, "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in "Handbook of Macroeconomics, John B. Taylor and Harald Uhlig eds," Vol. 2, Elsevier, 2016, pp. 415–525.

Stock, James H and Mark W Watson, "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 2018, *128* (610), 917–948.

Uhlig, Harald, "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 2005, *52* (2), 381–419.

A Appendix: Notations

d, s : Indicates demand and supply. E.g., ϕ^d, ϕ^s are the elasticities of demand and supply.
ε_t, η_t: Aggregate shocks.

ε_t^x, $\dot{\varepsilon}_t^x$, $\varepsilon_t^{x,\mathrm{OLS}}$,etc.: Aggregate shocks affecting variable x (e.g., supply if $x = s$). Depending on the specification, this variable can have 0, 1 or more dots. The superscript OLS means that it is the error arising in OLS.
λ: Factor loadings.
u_t: Idiosyncratic shocks.
$\check{u}_t := Qu_t$: Idiosyncratic shocks residualized by a projection matrix Q.
$z_t = \Gamma y_t = \Gamma u_t$: GIV.
$Z_t = y_{St} - y_{Et}$: Difference between size- and equal-weighted average of the outcome variable y_{it}.

B Appendix: Extensions to the user's guide

In this section, we provide extensions to the user's guide presented in Section 3 to cover extensions to the basic model as discussed in Section 4. In Section F, we use simulations to show the performance of these extensions in the presence of non-parametric factors, heteroskedasticity, and fat-tailed idiosyncratic shocks. We also discuss how bootstrap methods can be used to construct confidence intervals around the estimates.

Heteroskedasticity If the idiosyncratic shocks are heteroskedastic, $\sigma_i = \sigma(u_i)$, then we stop the baseline algorithm after Step 2. We then estimate $(\sigma_i^e)^2 = \frac{1}{T}\sum_t \check{u}_{it}^2$.[36] We return to Step 1, and estimate Step 1 and Step 2 with $\frac{1}{(\sigma_i^e)^2}$ as regression weights. Analogously, in Step 3, we replace y_{Et} by $y_{\tilde{E}t} = \sum_i \tilde{E}_i y_{it}$ with $\tilde{E}_i := \frac{(\sigma_i^e)^{-2}}{\sum_j (\sigma_j^e)^{-2}}$.

As the regression weights are estimated, it is useful in practice to explore the sensitivity of the estimates to winsorizing large values of the regression weights, $\frac{1}{(\sigma_i^e)^2}$, to avoid that one entity gets too much weight in estimating the factors if the estimated σ_i^e happens to be relatively low.

Robustness to large idiosyncratic shocks Fat-tailed idiosyncratic shocks are helpful in constructing GIVs as they increase the volatility of the instrument. At the same time, fat-tailed shocks or outliers can distort the estimation of the factors and fixed effects. We now discuss how to adjust the algorithm, so we can take advantage of the fat-tailed shocks without distorting the factor estimates. Section C.11 provides an analysis justifying this procedure (in particular Proposition 11).

We start from the model in Section 3.1 and write it as

$$y_{it} = a_i + b_t + \check{\lambda}_i \eta_t + \check{u}_{it},$$

where we simplify the controls $m^y C_{it}^y$ to an entity fixed effect a_i. As before, we estimate $(a_i, b_t, \check{\lambda}_i, \eta_t)$,

[36]See Section H.10 for a more refined estimation of σ_i.

which allows us to compute $\check{u}_{it}$ by minimizing a loss function

$$\min_{(a_i,b_t,\check{\lambda}_i,\eta_t)} \sum_{i,t} L\left(\frac{y_{it}-a_i-b_t-\check{\lambda}_i\eta_t}{\sigma_i}\right) \tag{36}$$

where the loss function L is such that $L'(x) = x + r(x)$, for a (intuitively, small) residual $r(x)$. For the loss function, we select the Huber (1964) loss function, $L^{\text{Huber}}(x,\delta) = \frac{x^2}{2}1_{|x|\leq\delta} + \left(|x| - \frac{\delta}{2}\right)\delta 1_{|x|>\delta}$, where the tails (i.e., $|x| > \delta$) have less influence than with a quadratic loss function $L^{\text{Gaussian}}(x) = \frac{x^2}{2}$ (which corresponds to $\delta \to \infty$).[37] For the GIV procedure in Section 3.1, we use a quadratic loss function, $r^{\text{Gaussian}}(x) = 0$. In case of fat-tailed idiosyncratic shocks, we use the Huber loss function, see also Fan et al. (2019). In this case, we set

$$r^{\text{Huber}}(x,\delta) = -\max(|x|-\delta, 0)\, sign(x), \tag{37}$$

where we specify δ below. This adjustment limits the impact of extreme observations.

We follow the same baseline GIV procedure while adjusting the loss function. We initialize the algorithm from $(a_i^{(0)}, b_t^{(0)}, \check{\lambda}_i^{(0)}, \eta_t^{(0)})$, and we provide a recommendation for these starting values after outlining the algorithm. The steps of the generalized algorithm are as follows, which recursively solve the first-order conditions of the loss function with respect to $(a_i, b_t, \check{\lambda}_i, \eta_t)$:

1. In case of the Huber loss function, compute $r_{it}^{(n)} = r^{\text{Huber}}(y_{it} - a_i^{(n)} - b_t^{(n)} - \check{\lambda}_i^{(n)}\eta_t^{(n)}, \delta_i^{(n)})$. For the quadratic loss function, set $r_{it}^{(n)} = 0$. We define $y_{it}^{w,(n)} := y_{it} + r_{it}^{(n)}$, where the w alludes to a form of $w$insorization, provided by the term $r_{it}^{(n)}$.

2. $b_t^{(n+1)}$: The FOC for the time fixed effect is solved by

$$b_t^{(n+1)} = \frac{\sum_i (y_{it}^{w,(n)} - a_i^{(n)} - \check{\lambda}_i^{(n)}\eta_t^{(n)})\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}. \tag{38}$$

3. $a_i^{(n+1)}$: The FOC for the entity fixed effect is solved by

$$a_i^{(n+1)} = \frac{1}{T}\sum_{t=1}^{T}(y_{it}^{w,(n)} - b_t^{(n+1)} - \check{\lambda}_i^{(n)}\eta_t^{(n)}). \tag{39}$$

4. $(\check{\lambda}_i^{(n+1)}, \eta_t^{(n+1)})$: Define $\check{y}_{it}^{w} := \frac{y_{it}^{w,(n)} - a_i^{(n+1)} - b_t^{(n+1)}}{\sigma_i}$. We run PCA (uncentered) $\check{y}_i^w = \frac{\check{\lambda}_i^{(n+1)\prime}}{\sigma_i}\eta_t^{(n+1)} + \frac{u_{it}}{\sigma_i}$, and recover $(\check{\lambda}_i^{(n+1)}, \eta_t^{(n+1)})$.

[37] One interpretation is the following. Suppose that $\frac{u_{it}}{\sigma_i}$ has a density $ke^{-L\left(\frac{u}{\sigma}\right)}$ for a constant k. Then (36) is the maximum likelihood estimator of the various parameters. With the Huber loss function, that density has fatter tails than a Gaussian.

5. $\delta_i^{(n+1)}$: We set it equal to the 90%-th percentile of the distribution of $\left|y_{it} - a_i^{(n+1)} - b_t^{(n+1)} - \check{\lambda}_i^{(n+1)}\eta_t^{(n+1)}\right|$ by entity i.[38]

Upon convergence of the algorithm, we form $Z_t = y_{St} - b_t^{(\infty)}$ and estimate for instance M in (26) as

$$s_t = MZ_t + c + \beta^{s,\eta}\eta_t^e + \varepsilon_t^{s,OLS}.$$

The estimates of μ, ϕ^d, and ϕ^s follow analogously.

In case of heteroskedasticity, we first run the homoskedastic procedure until convergence. We then compute the residuals, $u_{it}^e = y_{it} - a_i^{(\infty)} - b_t^{(\infty)} - \check{\lambda}_i^{(\infty)}\eta_t^{(\infty)}$. We re-estimate the variance, for instance, $\sigma_i^{e,2} = \frac{1}{T}\sum_t u_{it}^{e,2}$. We then run the procedure once again until convergence, which provides the final estimates.

We complete this extension by discussing how we initialize the algorithm:

1. We set $b_t^{(0)} = median(y_{it})$, the median across i, and $a_i^{(0)} = median(y_{it} - b_t^{(0)})$, the median across t. We then run the PCA algorithm on $y_{it} - a_i^{(0)} - b_t^{(0)}$ to initialize $\check{\lambda}_i^{(0)}$ and $\eta_t^{(0)}$.

2. We initialize the variance weights with $\sigma_i^{e,2} = var(y_{it})$. Even in the homoskedastic case, we allow σ_i^2 to differ by entity to allow for some small misspecification.

3. We initialize $\delta_i^{(0)} = \Phi^{-1}(0.95)\sigma_i^e$, where $\Phi^{-1}(0.95) \simeq 1.64$.

Missing data Missing data is generally not a problem for GIV estimation. If the panel is unbalanced, we can estimate the relevant parameters using the data available. For instance, using the equivalence between PCA estimation and alternating least squares,[39] we simply minimize $\sum_{i,t}(y_{it} - \lambda_i\eta_t)^2$ in estimating the simple factor model $y_{it} = \lambda_i'\eta_t + u_{it}$, summing only over the non-missing values. The same reasoning can be applied to any other step.[40]

[38] We can allow for asymmetric cutoff points if the distribution of u_{it} is skewed.

[39] In alternating least squares estimation of $y_{it} = \lambda_i'\eta_t + u_{it}$, we alternate between OLS estimation of η_t given λ_i, and of λ_i given η_t.

[40] In practice, one can fill in the missing data with the predicted value $\check{\lambda}_i^{(n)}\check{\eta}_t^{(n)}$ in the n-th iteration. Upon convergence, the predicted value equals the imputed data, which is equivalent to omitting the observation from the summation. The advantage is that each iteration can typically be computed very quickly using standard PCA algorithms.

Online Appendix for
"Granular Instrumental Variables"

Xavier Gabaix and Ralph S.J. Koijen

June 6, 2022

Table of contents

C Proofs omitted in the paper

C.1 Variance facts

We will repeatedly use a number of facts that we record here.

For two vectors of dimensions $n \times 1$, defining $u_X := X'u$ and $u_Y := Y'u$, we have

$$\mathbb{E}\left[u_X u_Y\right] = \mathbb{E}\left[\left(X'u\right)\left(Y'u\right)\right] = X'\mathbb{E}\left[uu'\right]Y = X'V^uY \tag{40}$$

Suppose now that $(u_i)_{i=1 \ldots N}$ is a series of uncorrelated random variables with mean 0 and common variance σ_u^2, then $V^u = \sigma_u^2 I$ and

$$\mathbb{E}\left[u_X u_Y\right] = X'Y\sigma_u^2 \tag{41}$$

and with $\Gamma = S - E$ (with $E_i = \frac{1}{N}$) we have

$$\mathbb{E}\left[u_\Gamma u_E\right] = 0, \qquad \mathbb{E}\left[u_\Gamma^2\right] = \mathbb{E}\left[u_S u_\Gamma\right] = h^2\sigma_u^2, \tag{42}$$

with $h=\sqrt{\sum_{i=1}^{N}S_i^2-\frac{1}{N}}$. Indeed, $\Gamma' E=\frac{1}{N}\sum_i\left(S_i-E_i\right)=0$, and

$$\Gamma'\Gamma=\sum_{i=1}^{N}\left(S_i-\frac{1}{N}\right)^2=\sum_{i=1}^{N}\left(S_i^2-\frac{2}{N}S_i+\frac{1}{N^2}\right)=\sum_{i=1}^{N}S_i^2-\frac{2}{N}\sum_{i=1}^{N}S_i+\frac{1}{N}=\sum_{i=1}^{N}S_i^2-\frac{1}{N}=h^2.$$

In the general heteroskedastic case for V^u, the quasi-equal weight vector is $\tilde{E}=\frac{(V^u)^{-1}\iota}{\iota(V^u)^{-1}\iota}$. Then, for any Γ such that $\iota'\Gamma=0$, we have:

$$\mathbb{E}\left[u_\Gamma u_{\tilde{E}}\right]=0. \tag{43}$$

Indeed, as $\tilde{E}=k\left(V^u\right)^{-1}\iota$ for $k=\frac{1}{\iota(V^u)^{-1}\iota}$, we have

$$\mathbb{E}\left[u_\Gamma u_{\tilde{E}}\right]=\tilde{E}'\mathbb{E}\left[uu'\right]\Gamma=\tilde{E}'V^u\Gamma=k\iota'\left(V^u\right)^{-1}V^u\Gamma=k\iota'\Gamma=0.$$

C.2 Proof of Proposition 1

It is proven in the main text.

C.3 Proof of Proposition 2

Derivation of asymptotic variance This part is elementary, and uses well-known ingredients. We have

$$\phi_T^{s,e}-\phi^s=\frac{\mathbb{E}_T\left[s_t z_t\right]}{\mathbb{E}_T\left[p_t z_t\right]}-\phi^s=\frac{\mathbb{E}_T\left[\left(s_t-\phi^s p_t\right)z_t\right]}{\mathbb{E}_T\left[p_t z_t\right]}=\frac{\mathbb{E}_T\left[\varepsilon_t u_{\Gamma t}\right]}{\mathbb{E}_T\left[p_t u_{\Gamma t}\right]}=\frac{A_T}{D_T}.$$

Next, the law of large number gives:

$$D_T=\mathbb{E}_T\left[p_t u_{\Gamma t}\right]\rightarrow^{a.s.}D,$$

with, using $p_t=\mu u_{St}+\varepsilon_t^p$ (see (3)):

$$D=\mathbb{E}\left[p_t u_{\Gamma t}\right]=\mathbb{E}\left[\left(\mu u_{St}+\varepsilon_t^p\right)u_{\Gamma t}\right]=\mu\mathbb{E}\left[u_{St}u_{\Gamma t}\right].$$

For the numerator, the central limit theorem gives the convergence in distribution:

$$\sqrt{T}A_T\rightarrow^d\mathcal{N}\left(0,\sigma_A^2\right),$$

where:

$$\sigma_A^2=\mathbb{E}\left[\varepsilon_t^2u_{\Gamma t}^2\right]=\mathbb{E}\left[\varepsilon_t^2\right]\mathbb{E}\left[u_{\Gamma t}^2\right]=\sigma_\varepsilon^2\sigma_{u_\Gamma}^2,$$

so that

$$\frac{\sigma_A}{D}=\frac{\sigma_\varepsilon\sigma_{u_\Gamma}}{\mu\mathbb{E}\left[u_{St}u_{\Gamma t}\right]}=:\sigma_{\phi^s}\left(\Gamma\right).$$

which is the announced expression for $\sigma_{\phi^s}(\Gamma)$. We showed that $\sqrt{T}\left(\phi_T^{s,e}-\phi^s\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\phi^s}^2\right)$.

Optimal GIV weights Γ Next, we solve for the optimal Γ, which minimizes $\sigma_{\phi^s}(\Gamma)$ subject to $\Gamma'\Lambda = 0$. Slightly more intuitively, we want to maximize the squared correlation $C(\Gamma) \coloneqq corr\left(u_{St}, u_{\Gamma t}\right)^2$:

$$\max_{\Gamma} corr\left(u_{St}, u_{\Gamma t}\right)^2 \text{ subject to } \Gamma'\Lambda = 0. \tag{44}$$

We next solve this problem. By (41), we have $\mathbb{E}\left[u_{St}u_{\Gamma t}\right] = \sigma_u^2\Gamma'S$, hence:

$$C(\Gamma)^2 = \frac{\mathbb{E}\left[u_{St}u_{\Gamma t}\right]^2}{var\left(u_S\right)var\left(u_{\Gamma t}\right)} = \frac{\left(\Gamma'S\right)^2}{\left(S'S\right)\left(\Gamma'\Gamma\right)}.$$

The problem is invariant to changing Γ into $\lambda\Gamma$ for a non-zero λ. So, we can fix say $S'\Gamma$ at some value. Given this, we want the minimum value of $\Gamma'\Gamma$. So, we minimize over Γ the Lagrangian

$$\mathcal{L} = \frac{1}{2}\Gamma'\Gamma - \Gamma'\Lambda b - c\Gamma'S \tag{45}$$

with some Lagrange multipliers b (of dimension $r \times 1$), and c (a scalar). The first order condition in Γ' is: $0 = \Gamma - \Lambda b - cS$, i.e.

$$\Gamma = cS + \Lambda b. \tag{46}$$

We next use the projection operator $Q = I - \Lambda\left(\Lambda'\Lambda\right)^{-1}\Lambda'$ (see (11)) which satisfies $Q\Lambda = 0$. As $\Lambda'\Gamma = 0$,we have $Q\Gamma = \Gamma$. So

$$\Gamma = Q\Gamma = cQS + Q\Lambda b = cQS \tag{47}$$

The factor c doesn't affect the results, (as Γ and $c\Gamma$ give the same estimator $\phi_T^{s,e}$), so we may choose $c = 1$, and conclude $\Gamma = QS$.

Regression interpretation. The procedure (15) collects residuals $\check{u}_t = Qy_t$. This means that the $\check{u}_t$ are the residual of the regression of y_t on the factors, gathered in the matrix Λ. So, $z_t = \Gamma'u_t = S'Qu_t = S'\check{u}_t$. This shows (14).

C.4 Proof of Proposition 3

In this example $\Lambda = \iota$, where ι is an N-dimensional vector of ones. So, with $E = \frac{\iota}{N}$ (a vector with entries equal to $\frac{1}{N}$)

$$Q = I - \iota\left(\iota'\iota\right)^{-1}\iota' = I - \frac{\iota\iota'}{N} = I - \iota E'$$

so that for a vector u, $\check{u} \coloneqq Qu = u - \iota u_E$, which means $\check{u}_{it} = u_{it} - u_{Et}$. So, $z_t = \sum_i S_i\check{u}_{it} = \sum_i S_i\left(u_{it} - u_{Et}\right) = u_{St} - u_{Et}$. Finally, $\sigma_{u_\Gamma} = h\sigma_u$ comes from (42).

Generalization For instance, if $\lambda_i = (1, x_i)$ with $x_E = 0$, the variance of the GIV is $\sigma^2_{u_\Gamma} = \sigma^2_u S'QS$ i.e.

$$\sigma^2_{u_\Gamma} = \sigma^2_u \left(h^2 - \frac{1}{N} \frac{x_S^2}{\sigma^2_{x_i}} \right). \tag{48}$$

This illustrates how controlling for more factors reduces the standard deviation of the GIV, hence (as in Proposition 3) a lower precision of the estimator, especially if x_S^2 is large and N is small. An advantage of having lots of small firms (large N) is that they make the estimation of the common shocks η_t easier, and hence increase the precision of the GIV estimator (that is, increase $\sigma^2_{u_\Gamma}$ by shrinking the last term in (48), $\frac{1}{N}\frac{x_S^2}{\sigma^2_{x_i}}$).

To derive (48), first note that $\Lambda'\Lambda = \begin{pmatrix} N & 0 \\ 0 & \sum_k x_k^2 \end{pmatrix}$, so that

$$\frac{1}{N}\Lambda'\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2_{x_i} \end{pmatrix}$$

where $\sigma^2_{x_i} = \frac{\sum_k x_k^2}{N}$. We let $\Omega = \Lambda \left(\Lambda'\Lambda\right)^{-1} \Lambda'$ so that $Q = I - \Omega$. Simple calculations show that, with $X = (x_i)_{i=1\dots N}$,

$$\Omega = \frac{1}{N}\iota\iota' + \frac{XX'}{N\sigma^2_{x_i}}$$

This implies:

$$S'QS = S'S - S'\Omega S = S'S - \frac{1}{N} - \frac{(S'X)^2}{N\sigma^2_{x_i}} = h^2 - \frac{x_S^2}{N\sigma^2_{x_i}},$$

in which $h^2 = S'S - N^{-1}, x_S = S'X$. So

$$\frac{\sigma^2_{u_\Gamma}}{\sigma^2_u} = S'QS = h^2 - \frac{x_S^2}{N\sigma^2_{x_i}}.$$

C.5 A canonical decomposition

We derive a "canonical representation" that, at the cost of some overhead and concepts and notations, will make the identifiability reasoning very crisp, and will be useful starting at the proof of Proposition 4.

C.5.1 Position of the problem: a motivating example

To illustrate the issue, consider the very simple factor model (with $i = 1, \dots, N$)

$$y_{it} = \eta_t + u_{it} \tag{49}$$

where the aggregate shock η_t is uncorrelated with the idiosyncratic shocks u_{it}, which are i.i.d.

$$\mathbb{E}\left[\eta_t u_{it}\right] = 0 \tag{50}$$

We are given the y_{it}, for a given t.

With finite N, one cannot estimate η_t exactly. But this is not a real problem for our GIV purpose, which can proceed with finite N. The "trick" is the following. Define

$$\check{\eta}_t := \eta_t + u_{Et}, \qquad \check{u}_{it} := u_{it} - u_{Et} \tag{51}$$

Then, we trivially have

$$y_{it} = \check{\eta}_t + \check{u}_{it} \tag{52}$$

But in addition we can recover $\check{\eta}_t$ and $\check{u}_{it}$, via

$$\check{\eta}_t = y_{Et}, \qquad \check{u}_{it} = y_{it} - y_{Et} \tag{53}$$

In addition, we have[41]

$$\mathbb{E}\left[\check{\eta}_t \check{u}_{it}\right] = 0. \tag{54}$$

Hence, $\check{\eta}_t$ and $\check{u}_{it}$ have properties "almost as good" as η_t and u_{it} —compare (50) and (54). But $\check{\eta}_t$ and $\check{u}_{it}$ easier to work with, as they can be exactly recovered from data (i.e., from y_{it}). The next subsection amplifies this idea to more general factor model, and shows a Proposition 9 that is useful in GIV analysis.

C.5.2 Canonical representation

Using (1) and (3), the model can be written as:

$$y_t = \Lambda \eta_t + \iota\left(\left(M-1\right) u_{St} + \varepsilon_t^y\right) + u_t, \qquad p_t = \mu u_{St} + \varepsilon_t^p \tag{55}$$

where ε_t^p, $\varepsilon_t^y := \phi^d \varepsilon_t^p$ and η_t are generically correlated, but are uncorrelated with u_t. We take the constant Λ case to alleviate notation, but we could have a time-varying Λ_t. We use the Q and R matrices of (31), reproduced here, and with respective dimensions $N \times r$ and $N \times N$:

$$R^{\Lambda,W} := \left(\Lambda' W \Lambda\right)^{-1} \Lambda' W, \qquad Q^{\Lambda,W} := I - \Lambda R^{\Lambda,W}$$

[41] Indeed,

$$\mathbb{E}\left[\check{\eta}_t \check{u}_{it}\right] = \mathbb{E}\left[\left(\eta_t + u_{Et}\right)\left(u_{it} - u_{Et}\right)\right] = \mathbb{E}\left[u_{Et}\left(u_{it} - u_{Et}\right)\right] = \mathbb{E}\left[u_{Et} u_{it}\right] - \mathbb{E}\left[u_{Et}^2\right] = \frac{\sigma_u^2}{N} - \frac{\sigma_u^2}{N} = 0.$$

where W will be kept implicit here to alleviate notations. In the homoskedastic case, $W = I$, and the reader is encouraged to think about this case first. In the heteroskedastic case, we take $W = c\,(V^u)^{-1}$ for some $c > 0$.

We decompose $\Lambda = \left[\iota : \check{\Lambda}\right]$, where $\check{\Lambda} = Q^\iota \Lambda$ is orthogonal to ι —using the scalar product modulated by W, $\langle A, B\rangle = A'WB$ for two vectors A, B: for instance, $\langle \iota, \check{\Lambda}\rangle = 0$ by (32). Recall that the dimensions of Λ, ι, $\check{\Lambda}$ are respectively $N \times r$, $N \times 1$, $N \times (r-1)$. This allows to decompose the space $\Omega = \mathbb{R}^N$ into three orthogonal components, $\Omega = \Omega^\iota \oplus \Omega^{\check{\Lambda}} \oplus \Omega^\perp$, where $\Omega^{\check{\Lambda}}$ (resp. Ω^ι) is the subspace generated by the $\check{\Lambda}$ (resp. ι), and $\Omega^\perp$ is the subspace orthogonal to Λ. Indeed, Q^ι is the projection on the space $\Omega^{\check{\Lambda}} \oplus \Omega^\perp$, the set of vectors orthogonal to ι, Q^Λ is the projection on $\Omega^\perp$, the vector space orthogonal to Λ.

We also decompose $\eta_t = (\eta_{1t}, \eta_{2t}')'$ where $\eta_t, \eta_{1t}, \eta_{2t}$ have dimensions r, 1, and $r-1$.

This gives the following decomposition.

Proposition 9 (Canonical representation) *When $W = c\,(V^u)^{-1}$ for some constant $c > 0$, we define $\Gamma = \Gamma^* := \left(Q^\lambda\right)' S$, and we have the following "canonical representation"*

$$y_t = \check{\Lambda}\check{\eta}_t + \iota y_{Et} + \check{u}_t, \qquad y_{Et} = (M-1)\,u_{\Gamma t} + \dot{\varepsilon}_t^y, \qquad p_t = \mu u_{\Gamma t} + \dot{\varepsilon}_t^p \tag{56}$$

with

$$\check{\eta}_t = R^{\check{\Lambda}} y_t, \qquad \check{u}_t = Q^\Lambda y_t = Q^\Lambda u_t, \qquad y_{Et} = R^\iota y_t, \qquad u_{\Gamma t} = S'\check{u}_t \tag{57}$$

where R and Q are the projection matrices defined in (31). Hence, we can recover $\check{\eta}_t$ and $\check{u}_t$ without error, by simple projections. Importantly, $\check{\eta}_t, \dot{\varepsilon}_t^y, \dot{\varepsilon}_t^p$ can be correlated between themselves but they are uncorrelated with $\check{u}_t$ and $u_{\Gamma t}$; and $\check{\Lambda}$, ι and $\check{u}_t$ are all orthogonal (using the scalar product $\langle A, B\rangle = A'WB$ with $W = (V^u)^{-1}$). Finally we have the following ancillary relations, with $\varepsilon_t^u := R^\Lambda u_t$ (with dimensions $r \times 1$), which is uncorrelated with $\check{u}_t, u_{\Gamma t}$:

$$\dot{\eta}_t := \eta_t + \varepsilon_t^u = \left(\dot{\eta}_{1t}, \check{\eta}_t'\right)', \qquad \dot{\varepsilon}_t^y := \varepsilon_t^y + (M-1)\,\lambda_S \varepsilon_t^u + \dot{\eta}_{1t}, \qquad \dot{\varepsilon}_t^p := \varepsilon_t^p + \mu\lambda_S \varepsilon_t^u \tag{58}$$

Variables η_t, $\dot{\eta}_t$, $\dot{\eta}_{1t}$, and $\check{\eta}_t$ have dimensions r, r, 1 and $r-1$ respectively.

One important message of Proposition 9 is that $\check{\eta}_t$ and $\check{u}_t$ can be recovered without any error, even though η_t and u_t can never be recovered exactly with our maintained assumption of fixed N. This is why the analysis is cleaner and easier in that space of $\check{\eta}_t$ and $\check{u}_t$. All this holds with the maintained assumption of parametric factors. We suspect that in future non-parametric analyses of the problem, this decomposition could still be useful.

In the main text, we use the term η_t^e for $\check{\eta}_t$, to signify that this is the estimated part of $\check{\eta}_t$ and to avoid the too burdensome notation $\check{\eta}_t^e$.[42]

[42] In the simplest example of Section C.5.1, with $\phi^d = 0$, η_t and $\check{\eta}_t$ have dimension 1. In the more general case with $\phi^d \neq 0$, one can recover only a $\check{\eta}_t$ with dimension equal to that of η_t, minus 1.

Proof of Proposition 9. In the proof, for simplicity we call $R = R^{\Lambda,W}$ and $Q = Q^{\Lambda,W}$,unless we explicitly state otherwise. We decompose $u_t = Qu_t + (I - Q)\, u_t$, so:

$$u_t = \check{u}_t + \Lambda \varepsilon_t^u, \qquad \varepsilon_t^u := Ru_t \tag{59}$$

Taking $W = c\,(V^u)^{-1}$ has the useful consequence that $\check{u}_t$ (an N dimensional vector) and ε_t^u (a r dimensional vector) are uncorrelated. Indeed:

$$\mathbb{E}\left[\check{u}_t \varepsilon_t^{u\prime}\right] = \mathbb{E}\left[(Qu_t)\,(Ru_t)'\right] = \mathbb{E}\left[Qu_t u_t' R'\right] = Q\mathbb{E}\left[u_t u_t'\right] R' = QV^u R' = c^{-1} Q W^{-1} R' = 0$$

using (32). This is a generalization of our basic case that assumed uniform loadings on the aggregate shock, $R^\lambda u_t = u_{Et}$, we had that u_{Et} and $\check{u}_t$ were uncorrelated. We have

$$u_{St} = S' u_t = S' Q u_t + S' \lambda \varepsilon_t^u = u_{\Gamma t} + \lambda_S \varepsilon_t^u$$

We start from (55):

$$p_t = \mu u_{St} + \varepsilon_t^p = \mu\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t^p = \mu u_{\Gamma t} + \dot{\varepsilon}_t^p$$

with $\dot{\varepsilon}_t^p = \varepsilon_t^p + \mu \lambda_S \varepsilon_t^u$.

In the same way, (55) gives

$$\begin{aligned}
y_t &= \Lambda \eta_t + \iota\left((M-1)\, u_{St} + \varepsilon_t^y\right) + u_t \\
&= \Lambda \eta_t + \iota\left[(M-1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t^y\right] + \check{u}_t + \lambda \varepsilon_t^u \\
&= \Lambda \dot{\eta}_t + \iota\left[(M-1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t^y\right] + \check{u}_t \text{ with } \dot{\eta}_t := \eta_t + \varepsilon_t^u \\
&= \check{\Lambda} \check{\eta}_t + \iota\left[(M-1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t^y + \dot{\eta}_{1t}\right] + \check{u}_t \text{ as } \Lambda \dot{\eta}_t = \iota \dot{\eta}_{1t} + \check{\Lambda} \check{\eta}_t \\
&= \check{\Lambda} \check{\eta}_t + \iota y_{Et} + \check{u}_t
\end{aligned}$$

by observing that $y_{Et} = R^\iota y_t$, so that

$$y_{Et} = (M-1)\, u_{\Gamma t} + \dot{\varepsilon}_t^y, \qquad \dot{\varepsilon}_t^y := (M-1)\, \lambda_S \varepsilon_t^u + \varepsilon_t^y + \dot{\eta}_{1t}$$

□

C.6 Proof of Proposition 4

Estimation of μ We distinguish between the Γ in this proposition (a general non-zero vector Γ satisfying $\Gamma' \Lambda = 0$), and the optimal $\Gamma^* := \left(Q^\lambda\right)' S$ from Proposition 9. That proposition showed that we have $p_t = \mu u_{\Gamma^* t} + \dot{\varepsilon}_t^p$, with $\dot{\varepsilon}_t^p$ uncorrelated with $\check{u}_t$, hence with $u_{\Gamma^* t} = S' \check{u}_t$. Hence, when $\Gamma = \Gamma^*$,the assumptions of OLS in (20) are satisfied, so that the estimator of μ is consistent, unbiased, and the OLS standard errors are correct.

For a more general Γ (satisfying $\Gamma'\Lambda = 0$, which in turns implies $\Gamma' Q = \Gamma'$), we observe first that

$$u_{\Gamma t} = \Gamma' u_t = \Gamma' Q u_t = \Gamma' \check{u}_t \tag{60}$$

Next, we can also project

$$u_{\Gamma^* t} = \beta u_{\Gamma t} + \zeta_t$$

for a regression coefficient $\beta = \frac{\mathbb{E}[u_{\Gamma^* t} u_{\Gamma t}]}{\mathbb{E}[u_{\Gamma t}^2]}$ and a random variable ζ_t uncorrelated with $u_{\Gamma t}$:

$$\mathbb{E}\left[u_{\Gamma t}\zeta_t\right] = 0.$$

We next show that

$$\beta = 1.$$

Indeed, from Proposition 9 we have $u_{St} = u_{\Gamma^* t} + \lambda_S \varepsilon_t^u$, so

$$\mathbb{E}\left[u_{\Gamma^* t} u_{\Gamma t}\right] = \mathbb{E}\left[\left(u_{St} - \lambda_S \varepsilon_t^u\right) u_{\Gamma t}\right] = \mathbb{E}\left[u_{St} u_{\Gamma t}\right] - \lambda_S \mathbb{E}\left[\varepsilon_t^u u_{\Gamma t}\right] = \mathbb{E}\left[u_{St} u_{\Gamma t}\right]$$

because ε_t^u is uncorrelated with $\check{u}_t$, hence with $u_{\Gamma t} = \Gamma' \check{u}_t$, which implies $\mathbb{E}\left[\varepsilon_t^u u_{\Gamma t}\right] = 0$. We record

$$\mathbb{E}\left[u_{\Gamma^* t} u_{\Gamma t}\right] = \mathbb{E}\left[u_{St} u_{\Gamma t}\right] \tag{61}$$

Hence, thanks to Assumption 2, $\beta = \frac{\mathbb{E}[u_{\Gamma^* t} u_{\Gamma t}]}{\mathbb{E}[u_{\Gamma t}^2]} = \frac{\mathbb{E}[u_{St} u_{\Gamma t}]}{\mathbb{E}[u_{\Gamma t}^2]} = 1$.

So, we have $p_t = \mu u_{\Gamma^* t} + \dot{\varepsilon}_t^p$ so

$$p_t = \mu u_{\Gamma t} + \left(\mu \zeta_t + \dot{\varepsilon}_t^p\right) \tag{62}$$

We have $\mathbb{E}\left[u_{\Gamma t}\zeta_t\right] = 0$. Also, as $\Gamma\Lambda = 0$, we have $\Gamma' = \Gamma' Q$, so $u_{\Gamma t} = \Gamma' u_t = \Gamma' Q u_t = \Gamma' \check{u}_t$ and as $\check{u}_t$ is uncorrelated with $\dot{\varepsilon}_t^p$ (from Proposition 9), we have $\mathbb{E}\left[u_{\Gamma t}\dot{\varepsilon}_t^p\right] = 0$. So, $\mathbb{E}\left[u_{\Gamma t}\left(\mu\zeta_t + \dot{\varepsilon}_t^p\right)\right] = 0$, and and the assumption of OLS are satisfied in (62). So that as above, the estimator of μ is consistent, unbiased, and the OLS standard errors are correct.

Estimation of M Relation (56) implies that we can write $s_t = y_{St} = M u_{\Gamma^* t} + \dot{\varepsilon}_t^s$ with $\dot{\varepsilon}_t^s := \check{\lambda}_S \check{\eta}_t + \dot{\varepsilon}_t^y$ uncorrelated with $\check{u}_t$. So, the arguments are the same as for the estimation of μ.

C.7 Proof of Proposition 5

Most of the proof is straightforward, given the work done before. We have

$$s_t = \phi^s p_t + \varepsilon_t \tag{63}$$

We can project $\varepsilon_t = \beta\eta_t^e + \varepsilon_t^\perp$ for some β; and by Proposition 9, η_t^e is uncorrelated with $\check{u}_t$. Hence $\varepsilon_t^\perp$ is uncorrelated with η_t^e and with $\check{u}_t$, hence with $z_t = \Gamma' u_t = \Gamma' \check{u}_t$ (see (60)). So we have:

$$s_t = \phi^s p_t + \beta\eta_t^e + \varepsilon_t^\perp, \tag{64}$$

with $\varepsilon_t^\perp$ uncorrelated with z_t. So, we have: $\mathbb{E}\left[\left(s_t - \phi^s p_t - \beta\eta_t^e\right)\left(z_t, \eta_t^e\right)\right] = 0$. This validates the generalization of Proposition 1 to the case with controls for η_t^e.

The proof of Proposition 2 extends to the case with controls, to show that $\sigma^2_{\phi^s}(\Gamma) = \frac{\sigma^2_{\varepsilon^\perp}\mathbb{E}\left[u^2_{\Gamma t}\right]}{\mu^2\mathbb{E}\left[u_{\Gamma^* t}u_{\Gamma t}\right]^2}$, for Γ^* the optimal weight.[43] Then, by Cauchy-Schwarz, $\frac{\mathbb{E}\left[u^2_{\Gamma t}\right]\mathbb{E}\left[u^2_{\Gamma^* t}\right]}{\mathbb{E}\left[u_{\Gamma^* t}u_{\Gamma t}\right]^2} \geq 1$, and the minimum is reached for $\Gamma = \Gamma^*$. This proves that $\Gamma = \Gamma^*$ achieves the maximum precision.

Finally, for the generalization of Proposition 4 to controlling for η_t^e, we repeat the same arguments, replacing (63) by (64).

C.8 Proof of Proposition 6

The earlier results proved the proposition in the case where there are no controls $C_{it}m$ and zero mean values (so that m and the mean values do not have to be estimated).

We treat the case where there are controls C_{it}^y. A very similar argument applies when there are other controls and additive constants (so that terms do not need to have a zero mean) in all equations.

We first treat the case where we use z_t as an instrument (rather than Z_t). We call $\check{y}_t = Q^{\iota,W} y_t$, i.e. $\check{y}_{it} = y_{it} - y_{\tilde{E}t}$ is the cross-sectionally demeaned value with quasi-equal weights $\tilde{E}$ (here we anticipate the generalization to heteroskedastic weights of Proposition 8). We use notations and concepts from Section C.5. Preparing the terrain for the heteroskedastic case of Proposition 8, we consider weighted regressions with weights W, see the notations of Proposition 8. The homoskedastic case corresponds to $W = I$.

Given a candidate value m, we construct $\check{\eta}_t(m) := R^{\check{\lambda}}\left(y_t - C_t^y m\right)$ and the associated GIV $z_t(m) := S' Q^\lambda\left(y_t - C_t^y m\right)$, as we do in Section 3.1. Define θ to be (m, M, μ) and θ^e to be the GMM estimator of θ associated with the following moments:

$$\mathbb{E}\left[\left(\check{y}_t - \check{x}_t\check{\eta}_t(m) - \check{C}_t^y m\right) W \check{C}_t^y\right] = \mathbb{E}\left[g_1(m)\right] = 0, \tag{65}$$

$$\mathbb{E}\left[\left(p_t - \mu z_t(m) - \beta^p\check{\eta}_t(m)\right)\left(z_t(m), \check{\eta}_t(m)\right)\right] = \mathbb{E}\left[g_2(\theta)\right] = 0 \tag{66}$$

$$\mathbb{E}\left[\left(s_t - M z_t(m) - \beta^s\check{\eta}_t(m)\right)\left(z_t(m), \check{\eta}_t(m)\right)\right] = \mathbb{E}\left[g_3(\theta)\right] = 0. \tag{67}$$

Under the regularity conditions we assumed, $\sqrt{T}\left(\theta^e - \theta\right)$ converges in distribution a normal distribution with mean 0. Now, we notice that $\mathbb{E}\left[\frac{\partial g_2(\theta)}{\partial m}\right] = \mathbb{E}\left[\frac{\partial g_3(\theta)}{\partial m}\right] = 0$ (as $\frac{\partial\check{\eta}_t(m)}{\partial m}$ and $\frac{\partial z_t(m)}{\partial m}$ are proportional to C_t^y). Then, by Theorem 6.1 of Newey and McFadden (1994), the standard errors

[43] Note also that $\mathbb{E}\left[u_{\Gamma^* t}u_{\Gamma t}\right] = \mathbb{E}\left[u_{St}u_{\Gamma t}\right]$ by (61), so this is the same problem as in the proof of Proposition 2.

on μ, M are the same as if we did not have to estimate m. This is precisely the case we worked out above, which proved that the standard errors on μ and M returned by OLS in this procedure are valid when we do not have estimate m (or equivalently that m was known so that all $C_{it}m$ terms were removed). So this proves that this claim (that the standard errors on μ and M returned by OLS in this procedure are valid) holds even when m needs to be estimated.[44]

All this holds when using z_t. When using $Z_t := y_{St} - y_{\tilde{E}t} = z_t + (\lambda_S - \lambda_{\tilde{E}})\, \eta_t = z_t + (\lambda_S - \lambda_{\tilde{E}})\,(0, \check{\eta}_t)$, the same holds with a slight redefinition of β^y and β^p (subtracting $(\lambda_S - \lambda_{\tilde{E}})$ times respectively M and μ).

C.9 Proof of Proposition 7

In vector notation, $y_t = \phi^d p_t + \Lambda \eta_t + u_t$, i.e.

$$y_t = X\left(\dot{\phi}^d p_t + \dot{\lambda}\eta_t\right) + u_t.$$

Hence, we have $\check{u}_t := Q y_t = Q u_t$ and

$$\dot{y}_t := R y_t = \dot{\phi}^d p_t + \dot{\lambda}\eta_t$$

We identify ϕ^s using

$$\mathbb{E}\left[\left(s_t - \phi^s p_t\right) z_t\right] = 0, \tag{68}$$

which is valid for any weighing matrix W (here we provide the proof both in the homoskedastic case and also in the heteroskedastic case, anticipating Proposition 8). Next, we suppose that we have $W = c\,(V^u)^{-1}$ for some constant c. Then we can also estimate $\dot{\phi}^d$ using the moment (which is k-dimensional):

$$\mathbb{E}\left[\left(\dot{y}_t - \dot{\phi}^d p_t\right) z_t\right] = 0. \tag{69}$$

To see this, we form $R y_t = \dot{\phi}^d p_t + \dot{\lambda}\eta_t + R u_t$, and we have $\mathbb{E}\left[(R u_t)\, \check{u}_t'\right] = 0$, per Proposition 9.[45]

C.10 Proof of Proposition 8

The decomposition of Proposition 9 already considered the heteroskedastic case. It shows how to revise all the arguments above, with heteroskedasticity. To generalize the optimality of Γ^* in Proposition 2, the arguments just repeat the Cauchy-Schwarz arguments of the proof of Proposition 5 for the optimality of Γ^*.

[44] Note that the estimation of m need not be efficient: it is enough for the argument that the estimator of m be consistent.

[45] If we take W as the identity rather than the "ideal" W above, the error in (69) is typically quite moderate: for instance, with $X = \iota$, it is only of order $\frac{1}{N}$.

C.11 Dealing with fat tails: Justification of the procedure in Appendix B

Here we provide a justification of the procedure in Appendix B to dampen the influence of outliers.For clarity, we consider the following problem first — our main problem is just a more complex variant. Suppose that we want to estimate β in a regression:

$$y_i = \beta x_i + u_i \tag{70}$$

with x_i independent of u_i, and the u_i's are i.i.d. with density $p(u)$ with mean 0. Suppose that there are outliers, e.g. fat-tailed u_i. What to do?

Background: Traditional winsorization yields biased estimates With outliers, a common procedure is to winsorize y_i, e.g. replace y_i by

$$y_i^W := sign(y_i) \min(|y_i|, \delta), \tag{71}$$

a winsorization at δ for some $\delta \geq 0$. This can be equivalently rewritten as:

$$y_i^W := y_i + r(y_i) \tag{72}$$

with

$$r(u) := r_\delta(u) = -\max(|u| - \delta, 0) \operatorname{sign}(u). \tag{73}$$

While common, there are difficulties with this procedure. The OLS estimator is biased, as in general

$$\mathbb{E}\left[\left(y_i^W - \beta x_i\right) x_i\right] \neq 0 \tag{74}$$

In addition, there is no clear micro foundation of this procedure, e.g. via MLE.

Winsorization of the residual, not of outcome variables Instead, we use a simple variant that solves both those difficulties, following Huber (1964) and e.g. Sun et al. (2020). It uses the following "winsorization of the residual", by defining:

$$y_i^w := y_i + r(y_i - \beta x_i) \tag{75}$$

instead of the traditional (72), and then to run OLS of $y_i^w = \beta x_i + \varepsilon_i$.

This is a fixed point problem, which leads following algorithm. We initialize $\beta^{(0)}$, e.g. setting it to the plain OLS value. The two steps are as follows:

1. Define $y_i^{w,(n)} := y_i + r\left(y_i - \beta^{(n)} x_i\right)$.

2. Run the OLS of

$$y_i^{w,(n)} = \beta x_i + \varepsilon_i. \tag{76}$$

which yields an update $\beta^{(n+1)}$, and we iterate until convergence.

We next justify this. Define $L(u) = -\ln p(u)$, the log likelihood of y is $\sum_i L(y_i - \beta x_i)$, so the maximum likelihood estimator is

$$\min_\beta \sum_i L(y_i - \beta x_i) \tag{77}$$

whose first order condition is

$$\sum_i L'(y_i - \beta x_i) x_i = 0 \tag{78}$$

If the residuals u_i are Gaussian distributed, we have $L(u) = \frac{1}{2}ku^2 + k'$ for some constants k and k', so $L'(u) = ku$, and we obtain the familiar OLS estimator. But otherwise, we have a nonlinear equation, which is a bit painful to solve. In general, we express:

$$L'(u) = k(u + r(u)) \tag{79}$$

where intuitively the residual term $r(u)$ is "small". For instance, for the log density

$$L^{\text{Huber}}(u) = \frac{u^2}{2} 1_{|x| \leq \delta} + \left(|u| - \frac{\delta}{2}\right) \delta 1_{|u| > \delta} \tag{80}$$

then we have $L^{\text{Huber}\prime}(u) = u + r(u)$ with $r(u)$ exactly as in (73). This is why we take this value of $r(u)$ in practice. But we continue the discussion for a general $r(u)$.

Then, the FOC (78) becomes

$$\sum_i (y_i - \beta x_i + r(y_i - \beta x_i)) x_i = 0 \tag{81}$$

To get more intuition, we define $y_i^w := y_i + r(y_i - \beta x_i)$ as in (75), which has the interpretation of sort of "winsorized" y_i, hence the w superscript. Then FOC is

$$\sum_i (y_i^w - \beta x_i) x_i = 0 \tag{82}$$

Hence, we can estimate β by OLS, once we have an estimate of y_i^w.

We next state a simple proposition.

Proposition 10 *Suppose that $\mathbb{E}[r(u_i) x_i] = 0$, for instance, because u_i and x_i are independent, and $\mathbb{E}[r(u_i)] = 0$. Then, at the correct value β we have*

$$\mathbb{E}[(y_i^w - \beta x_i) x_i] = 0 \tag{83}$$

with $y_i^w = \beta x_i + r(y_i - \beta x_i)$.

The proof is almost a tautology: the statement is equivalent to saying that $\mathbb{E}[r(u_i)x_i] = 0$, which is which exactly the main assumption of the proposition. But the advantage is that it lays out a simple procedure to "winsorize" outliers: run the OLS (76), $y_i^w = \beta x_i + \varepsilon_i$. If u_i is e.g. non-symmetric, it shows a simple criterion for other residual functions, $\mathbb{E}[r(u)] = 0$, e.g. by choosing a function $r(u)$ that is non-symmetric.

Link with the procedure in Appendix B. With the more complex factor model of Appendix B, the arguments are exactly the same. We can state the following proposition, and prove it exactly the same way. Hence moment conditions used in Appendix B are valid.

Proposition 11 *Suppose that* $\mathbb{E}[r(u_i)(\lambda_i, a_i, \eta_t, b_t)] = 0$. *Then at the correct values, the following moments hold*

$$\mathbb{E}[(y_{it}^w - (a_i + b_t + \lambda_i \eta_t))(\lambda_i, a_i, \eta_t, b_t)] = 0$$

where $y_{it}^w = y_{it} + r(y_{it} - (a_i + b_t + \lambda_i \eta_t))$.

D Detailed links with previous literature

Procedures containing elements of GIVs A few papers have explored the idea of using idiosyncratic shocks as instruments to estimate spillover effects, such as Leary and Roberts (2014b) in the context of firms' capital structure choice and Amiti et al. (2019) in the context of firms' price setting decisions. The structure of the estimating equations in these papers is similar to the model that we consider here:[46]

$$y_t = \lambda y_{wt} + mC_t + u_t,$$

where $y_{wt} = w' y_t$ can be equally-weighted (Leary and Roberts (2014b)) or size-weighted (Amiti et al. (2019)), depending on the weights w. Both papers use industry and/or year fixed effects, which can be viewed as a choice of controls or exogenous factors, η_t, to which all firms in a given industry have the same exposure.

There are two main differences compared to GIV. First, both papers use idiosyncratic shocks to another variable than y_t, say g_t, to construct an instrument for y_{wt}. Leary and Roberts (2014b) use idiosyncratic stock returns and Amiti et al. (2019) use shocks to competitors' marginal cost, exchange rates, or export prices. We, instead, propose to use idiosyncratic shocks to y_t rather than another instrument (this way requiring fewer times series). Second, and related, we control for heterogeneous exposures to common factors to extract the idiosyncratic shocks, which is important in asymptotic theory and in practice in realistic samples (see Section F).

[46] Amiti et al. (2019) study the price setting decision of firms. In their model, the pricing equation features two endogenous variables, namely the same firm's marginal cost and the size-weighted average of competitors' prices. We focus on the spillover effects of competitors' prices in our discussion in this section.

A third difference is specific to Leary and Roberts (2014b). GIVs crucially depend on the difference between size- and equal-weighted averages of variables. If the estimating equation depends on equal-weighted averages, GIV cannot be applied. In most models, however, not all competitors receive equal weight and larger firms, or perhaps firms that are closer in product space, receive a larger weight.

Lastly, the use of model-based idiosyncratic shocks has some similarities with Amiti and Weinstein (2018), who extract bank supply shocks from Japanese data using a panel of fixed effects, and then estimate the sensitivity of aggregate investment to these shocks. However, unlike our model, Amiti and Weinstein (2018) assume a uniform sensitivity to the aggregate shocks ($\lambda_i \eta_t$ with $\lambda_i = 1$ for all i), and do not allow for general equilibrium effects: shocks to banks affect aggregate investment, but aggregate investment does not circle back around to affect individual bank behavior. This is the key source of endogeneity in many of the models we consider, and by tackling it we are able to estimate a richer set of parameters.

In a tangentially related recent paper, Sarto (2018) uses factor analysis to extract values of η_t (much as we do when we "recover" a factor η_t). Take the basic example in our paper. Then, Sarto does not identify ϕ^s: even if η_t (the aggregate shock to demand) were perfectly identified, that would not allow to estimate p_t. In the supply and demand example, Sarto's approach would identify the demand elasticity ϕ^d, but not the supply elasticity ϕ^s.

Other methods to estimate aggregate elasticities Rigobon (2003) introduces another method that can be used to estimate spillover effects and aggregate multipliers using time-variation in second moments. If shocks are heteroskedastic and the structural parameters are stable across regimes, then the different volatility regimes add additional equations to the system so that the structural parameters can be identified. GIV does not require heteroskedasticity, but can accommodate it, and is therefore complementary to identification methods that rely on heteroskedasticity.

Influence and the "reflection problem" The "reflection problem" (Manski (1993); Kline and Tamer (2020)) studies a related form of contagion. The traditional literature does not use idiosyncratic shocks in its identification strategies. Hence a GIV approach can be useful to complement existing approaches. Section H.17 develops this.[47]

Spatial econometrics In some applications of GIVs we have considered separately, growth in a region affects that of the other regions. So there is a similarity between our setup and that of spatial econometrics (e.g. Kelejian and Prucha (1999)). However, the estimators are quite different. The reason is that spatial econometrics studies the "local" influence (e.g. of neighboring cities on a city), while GIVs study the global influence. Hence, the sources of variation, identifiability conditions and

[47] Somewhat related, Graham (2008) explores the identification of peer effects using conditional variance restrictions on the outcomes by exploiting differences in the sizes of the peer group. Intuitively, smaller peer group sizes lead to a larger contribution of each individual peer on the peer component.

methods are quite different. Certainly, the spatial literature has not identified, as we do, the GIVs as a simple way to estimate elasticities in contexts such as supply and demand problems, and models with general equilibrium effects as opposed to local effects. Still, some of the sophisticated techniques of the spatial literature might be used one day to enrich a GIV-type analysis.

Quasi-experimental instruments and identification by functional form A large literature explores identification by functional form, where consistency of the estimator depends on functional form or distributional assumptions. Classic examples include the Heckman (1978) selection model, identification via heteroskedasticity, as in Rigobon (2003) and Lewbel (2012), and Arellano and Bond (1991) and Blundell and Bond (1998) in the context of dynamic panel data models. The typical concern with these approaches, compared to quasi-experimental instruments that are outside of the model, is that the estimators are inconsistent when the model is misspecified.

In the case of GIVs, we generally start from a structural model that motivates the estimating equation, as in our empirical example. This prescribes the definition of the size vector S and, in some cases, the characteristics that determine the exposures x_{it}. To extract idiosyncratic shocks, we rely on statistical factor models.[48]

Instead of viewing this last step as a merely statistical exercise that is hard to validate externally, GIVs provide an empirical strategy to understand the economic drivers of the instrument by screening the top shocks narratively. By understanding the nature of the shock based on news coverage (as in the narrative examination we just discussed), for instance, we can ensure that the shocks are truly idiosyncratic and interpretable. For instance, a large negative return associated with a failed stress test of a bank in the context of doom loops, a negative supply shock in Kuwait and Iraq during the First Gulf War, or a positive demand shock in China in the early 2000s in the context crude oil markets, are all valid instruments. While alternative identification methods might rely on functional form assumptions only, GIVs, by being able to screen the shocks economically, provide a systematic way to construct instruments more in the spirit of quasi-experimental instruments.

E Microfoundations for the model of sovereign spillovers

We provide a microfounded model for the empirical application of Section 5. In this model, spillovers happens because debt defaults are partially mutualized. This is a stand-in for potentially much richer economics. For instance, contagion might work via GDP spillovers, or the limited risk capacity of specialized financiers. Still, the specification that this model delivers might be broadly similar, as we shall see.

[48] We discuss the robustness of GIVs to various forms of misspecification in Section 3.2.

E.1 Model setup

We make a number of simplifying assumptions. The safe interest rate is normalized to 0, and pricing is risk neutral. Time is continuous in $[0, T]$. We neglect the $O(dt)$ terms, which are irrelevant for the regression analysis we are interested in, i.e. will write $df(X_t) = f'(X_t)\, dX_t$.[49,50]

Payoffs are realized at a date T, which should be thought about as faraway. In particular, countries can default at date T only. Country i's outstanding debt is B_i, and the value of the debt (per unit of face value) is thus:

$$Q_{it} = \mathbb{E}_t \left[1 - L_{iT}^{+}\right] = e^{-(T-t)y_{it}}, \tag{84}$$

where $x^{+} := \max(x, 0)$, y_{it} is the yield spread over the safe interest rate (which we normalized to 0), and L_{iT} is the relative "vulnerability" of the government's bonds, defined as

$$L_{iT} = \frac{F_{iT}}{B_i}, \tag{85}$$

where F_{iT} is the value of potential losses from government default (in euros). We assume that F_{iT} follows:

$$F_{iT} = \psi_{iT} G_{iT}, \tag{86}$$

where $\psi_{iT} \in [0, 1]$ is a propensity to pass on raw government fiscal losses G_{iT} to bondholders. A financially virtuous country (say Germany) has ψ_{iT} close to 0, and a laxer country has a high ψ_{iT}. To gain intuition, it is useful to think that most variation in yield spreads comes from the political willingness to not pay bondholders, ψ_{iT}.

This raw position G_{iT} is in turn:

$$G_{iT} = V_{iT} - \phi F_{iT}^{+} + \phi m_i F_T, \tag{87}$$

where V_{iT} is a stochastic "latent loss", and the total amount lost on bonds is:

$$F_T = \sum_i F_{iT}^{+}. \tag{88}$$

Debts are partially mutualized with intensity $\phi \in [0, 1]$: a fraction ϕ of the loss F_{it}^{+} is passed on to other countries, with a share m_i to country i ($\sum_i m_i = 1$, $m_i \geq 0$). This mutualization creates the sovereign yield spillovers.

[49] Formally, we write all the differential expressions $dY_t = a_t dZ_t$ modulo an equivalence by terms $b_t dt$ (or, to be pedantic, we quotient by the ring of expressions of the type $b_t dt$ where b_t is an adapted function). So, $df(X_t) = f'(X_t)\, dX_t$ modulo dt, where we keep the "modulo dt" implicit.

[50] We only care, for the regressions, about the "dZ_t" terms, that depends on innovations to underlying Brownian shocks dZ_t, as those are the loadings detected by the regressions.

To simplify the analysis, we assume that V_{iT} is strictly positive with probability 1, so that F_{iT}, G_{iT} and L_{iT} are all strictly positive with probability 1. This is less restrictive that it may appear: losses could be very small. This is simply to make the analysis very tractable.

E.2 Model solution

Solving the model,

$$\begin{aligned} L_{iT} &= \frac{\psi_{iT}}{B_i}\left(V_{iT} - \phi F_{iT} + \phi m_i F_T\right) \\ &= \frac{\psi_{iT}}{B_i}\left(V_{iT} - \phi B_i L_{iT} + \phi m_i B L_T\right), \end{aligned}$$

with $B = \sum_i B_i$ and $L_T = \frac{F_T}{B}$, i.e.

$$L_T = \sum_i \frac{B_i}{B} L_{iT}. \tag{89}$$

We call $\rho_i = \frac{m_i}{B_i/B}$, the ratio between country i's mutualization share m_i and its debt share.[51,52]. This leads to:

$$L_{iT} = \frac{\psi_{iT}}{1+\phi\psi_{iT}}\left(\frac{V_{iT}}{B_i} + \phi\frac{m_i}{B_i} B L_T\right).$$

So, if we define

$$\Psi_{iT} = \frac{\psi_{iT}}{1+\phi\psi_{iT}}, \tag{90}$$

we have:

$$L_{iT} = \Psi_{iT}\left(\frac{V_{iT}}{B_i} + \phi\rho_i L_T\right). \tag{91}$$

This shows the "contagion" in the space of vulnerabilities, L_{iT}.

To move to yields, we do a Taylor expansion for small yield spreads, so that (84) gives:

$$y_{it} = a_t \mathbb{E}_t\left[L_{iT}\right], \tag{92}$$

where

$$a_t = \frac{1}{T-t} \tag{93}$$

is a slowly-varying parameter (as T is far from the interval of times t under study – so we'll take the approximation $da_t \simeq 0$). We define $\Psi_{it} = \mathbb{E}_t\left[\Psi_{iT}\right]$, $v_{it} = a_t \mathbb{E}_t\left[\frac{V_{iT}}{B_i}\right]$. Also, we place ourselves in the

[51] The ECB's capital key, which defines the equity shares of member states in the ECB, is defined using 50% of GDP shares and 50% of population shares. However, we do not focus exclusively on spillovers that operate via the ECB and there may be other effects via trade linkages, demand shocks from investors, et cetera. We maintain the assumption that the losses, or exposures, to Eurozone-wide losses are proportional to GDP. Alternatively, we could change the measure m_i to be a function of both population and GDP shares.

[52] One can imagine $\rho_i \simeq 1$ as a simple baseline where most variations come from the political willingness ψ_{it}.

"quasi-static" regime, where all noises are small—see Section E.3 for details. Hence, (91) becomes, in yield space:

$$y_{it} = \Psi_{it} \left(v_{it} + \phi \rho_i y_{St} \right), \tag{94}$$

where

$$y_{St} = \frac{\sum_i B_i y_{it}}{B}. \tag{95}$$

This shows that the yield spread depends on a country-specific fundamental v_{it} and a "spillover" proportional to ϕ. At the same time, for a very financially virtuous country with $\Psi_{it} \simeq 0$, the yield spread is close to 0, so that yield contagion is close to 0: as the country is quite safe anyway, external disruptions cannot move the yield much away from 0.

We have

$$\frac{dy_{it}}{y_{it}} = \frac{d\Psi_{it}}{\Psi_{it}} + \frac{dv_{it}}{v_{it} + \phi \rho_i y_{St}} + \frac{\phi \rho_i y_{St}}{v_{it} + \phi \rho_i y_{St}} \frac{dy_{St}}{y_{St}},$$

hence

$$\frac{dy_{it}}{y_{it}} = dw_{it} + \gamma_{it} \frac{dy_{St}}{y_{St}} \tag{96}$$

for $dw_{it} := \frac{d\Psi_{it}}{\Psi_{it}} + \frac{dv_{it}}{v_{it}+\phi\rho_i y_{St}}$ and for a coefficient $\gamma_{it} := \frac{\phi\rho_i y_{St}}{v_{it}+\phi\rho_i y_{St}} \in [0,1]$. In the simple benchmark where all countries have a similar v_{it} (fundamental government finances) but differ mostly in Ψ_{it} (the propensity to absorb the shocks rather than pass it on to debt holders by defaulting) and $\rho_i = 1$, we have $\gamma_{it} = \frac{\phi y_{St}}{v_t + \phi y_{St}}$.

Written another way, call

$$\tilde{y}_{it} := \ln y_{it}. \tag{97}$$

Then, we have

$$d\tilde{y}_{it} = dw_{it} + \gamma_{it} d\tilde{y}_{\tilde{S}t}, \tag{98}$$

where

$$\tilde{S}_{it} = \frac{B_i y_{it}}{\sum_j B_j y_{jt}}, \tag{99}$$

$$d\tilde{y}_{\tilde{S}t} = \sum_i \tilde{S}_{it} d\tilde{y}_{it} = \sum_i \frac{B_i y_{it} \frac{dy_{it}}{y_{it}}}{\sum_j B_j y_{jt}} = \frac{dy_{St}}{y_{St}}. \tag{100}$$

Hence, if we reason in "log yield spread" space, the proper weights are proportional to $B_i y_{it}$, i.e. debt value times yield spread. This is the formulation that motivates our empirical specification (33). In particular, if $\Psi_{it} = 0$, then the change is always $dy_{it} = 0$. The importance of the spillovers is given by $\sum_j B_j dy_{jt}$, the change in the yield weighted by debt value, summed over all countries.

E.3 Quasi-static regime of stochastic processes

Suppose a stochastic process, governed by some noise size σ, as in $dY_t = \mu(Y_t)\,dt + \sigma v(Y_t)\,dB_t$, where B_t is a Brownian motion. The "quasi-static" regime is the one where σ is very close to 0. Then, things are much simpler to analyze, especially for non-linear processes, provided we accept $O(\sigma^2)$ approximations.

Indeed, consider that vector-valued process Y_t (for $t \leq T$)

$$X_t = \mathbb{E}_t[F(Y_T)] \tag{101}$$

where F is a C^2 function. Then, in the quasi-static regime, we can write

$$X_t = F(\mathbb{E}_t[Y_T]) + O\left(\sigma^2\right) \tag{102}$$

i.e. we swap $\mathbb{E}_t$ and F.[53] So, assume now that Y_t is a martingale,

$$X_t = F(Y_t) + O\left(\sigma^2\right) \tag{103}$$

and

$$dX_t = F'(Y_t)\,dY_t + O\left(\sigma^2\right) \tag{104}$$

or, more informally (as we do in the economic part of this section),

$$dX_t \simeq F'(Y_t)\,dY_t. \tag{105}$$

To work out a concrete example, take $Y_t = \sigma B_t$, and $X_t = \mathbb{E}_t\left[e^{Y_T}\right]$. The exact values are:

$$X_t = e^{Y_t + \frac{\sigma^2}{2}(T-t)}, \qquad dX_t = X_t dY_t \tag{106}$$

and the quasi-static approximation gives

$$X_t = e^{Y_t} + O\left(\sigma^2\right), \qquad dX_t = e^{Y_t} dY_t + O\left(\sigma^2\right). \tag{107}$$

E.4 Details on the data

We use data on general government gross debt for each country from Eurostat (series identifier: *teina225*). The tickers that we use for different countries, and the countries included, are the ones used by European Insurance and Occupational Pensions Authority (EIOPA) to construct the regulatory yield curves of insurance companies and pension funds (EIOPA (2017)). We use their

[53] We do not formally prove this, as this is purely mathematical as opposed to economic. One could do it, e.g. using the Clark-Ocone formula from the Malliavin calculus.

Table E.3: Bloomberg identifiers of countries included in the sovereign yield model.

Country	Government bond ticker ID	Country	Government bond ticker ID
Austria	G0063Z 10Y BLC2 Curncy	Ireland	G0062Z 10Y BLC2 Curncy
Belgium	G0006Z 10Y BLC2 Curncy	Italy	G0040Z 10Y BLC2 Curncy
Finland	G0081Z 10Y BLC2 Curncy	Netherlands	G0020Z 10Y BLC2 Curncy
France	G0014Z 10Y BLC2 Curncy	Portugal	G0084Z 10Y BLC2 Curncy
Germany	G0016Z 10Y BLC2 Curncy	Slovenia	G0259Z 10Y BLC2 Curncy
Greece	G0156Z 10Y BLC2 Curncy	Spain	G0061Z 10Y BLC2 Curncy

government bond ticker ID (see their Table 1). We select the EZ13 Eurozone, which was established in 2007, before the 2008 financial crisis.Table E.3 describes the tickers of the yields that we use in our empirical analysis. We use Bloomberg's price variable PX_LAST.

F Simulations

We use simulations to illustrate the extended GIV algorithm in Section B that can handle (i) non-parametric factors, (ii) heteroskedasticity, and (iii) fat-tailed idiosyncratic shocks.

We describe the data generating process in Section F.1. In Section F.2, we discuss bootstrap methods to compute confidence intervals and coverage rates. We then conclude this section by presenting the results in Section F.3.

F.1 Data generating process

We consider the following data generating process, which is a special case of the model in Section 2:

$$\begin{aligned} y_{it} &= \gamma y_{St} + \lambda_i \eta_t + u_{it}, \\ \eta_t &\sim N(0,1), \\ u_{it} &\sim D(0,1), \\ \lambda_i / c &\sim U[0,1]. \end{aligned}$$

where $D(0,1)$ is a distribution with zero mean and unit variance that we will specify soon — it is either Gaussian or fat-tailed. We choose c so that the share of variance in y_{St} due to idiosyncratic shocks is equal to θ in each run,

$$\theta = \frac{S'S}{c^2 \lambda_S^2 + S'S}, \tag{108}$$

and thus

$$c = \sqrt{(\theta^{-1} - 1)\frac{S'S}{\lambda_S^2}}.$$

We vary $h = \sqrt{S'S - 1/N} \in \{0.2, 0.5\}$ and $\theta \in \{0.2, 0.4\}$. To explore the impact of fat-tailed idiosyncratic shocks, we vary the distribution of u_{it}. We start from $u_{it} \sim N(0, 1)$. We then also consider $u_{it} = \sqrt{\frac{\nu-2}{\nu}}\epsilon_{it}$ with $\epsilon_{it} \sim t_\nu$ and select $\nu = 3$. We assume that there are 50 entities and $T = 100$ time periods. For each set of parameters, we use 10,000 replications in the simulations and we start from the same seed for each of the parameter configurations.

Heteroskedasticity We also explore the impact of heteroskedastic shocks. In this case, we multiply u_{it} by σ_i, where σ_i^2 is specified as

$$\sigma_i^2 = b \exp(-0.3 \ln S_i),$$

where b is chosen to ensure that $S' diag(\sigma^2) S = S'S$ and thus

$$b = \frac{S'S}{S' diag(\exp(-0.3 \ln S))S}, \tag{109}$$

which makes sure that (108) is satisfied for the same c.

F.2 Bootstrap methods

To compute confidence intervals and coverage rates, we consider two algorithms, the non-parametric bootstrap and the wild bootstrap, by adjusting methods discussed in Goncalves and Perron (2014) and Horowitz (2019) to our setting.

F.2.1 Non-parametric bootstrap

In case of the non-parametric bootstrap (i.e., the basic bootstrap), we resample the vector y_t from the original sample with replacement. We consider 1,000 bootstrap samples.

F.2.2 Wild bootstrap

Given an estimate M^e, we compute the structural parameter γ^e, $\gamma^e = \frac{M^e - 1}{M^e}$. In rare cases where $|\gamma^e| > 1$, we ensure that the estimated parameter is in $[-1, 1]$ to ensure that the model is stable by setting $\gamma^e = \max(\min(\frac{M^e-1}{M^e}, 1), -1)$. We then construct the bootstrap samples using the following procedure:

1. Compute $\tilde{y}_{it} = y_{it} - \gamma^e y_{St} = \lambda_i \eta_t + u_{it}$.

2. Winsorize $\tilde{y}_{it}$ at the 2.5% and 97.5% level to mitigate the impact of outliers, and refer to the winsorized data as $\tilde{y}_{it}^{W,0.05}$.[54]

3. Compute time fixed effect $b_t^e = \frac{1}{N}\sum_i \tilde{y}_{it}^{W,0.05}$, and the entity fixed effect $a_i^e = \frac{1}{T}\sum_t \left(\tilde{y}_{it}^{W,0.05} - b_t^e\right)$.

4. Run PCA on $\tilde{y}_{it}^{W,0.05} - a_i^e - b_t^e$ and call the estimated factor η_t^e and loadings λ_i^e.

5. Recover the residuals as $u_{it}^e = \tilde{y}_{it} - a_i^e - b_t^e - \lambda_i^e \eta_t^e$.

6. Construct the n-th bootstrap sample $y_{it}^{(n)} = (I - \gamma^e \iota S')^{-1} (\lambda_i^e \eta_t^e + a_i^e + b_t^e + \nu_{it} u_{it}^e)$, where ν_{it} are independently drawn from a standard normal distribution, $v_{it} \sim N(0,1)$. This last step is the essence of the wild bootstrap.

F.2.3 Confidence intervals and coverage rates

The confidence interval and coverage rate are computed following the procedure outlined in Horowitz (2019).[55]

1. Let $\hat{s}$ be the OLS standard error of $\hat{M}$ estimated from the regression in the final step of the GIV algorithm
$$y_{St} = MZ_t + c + \beta\eta_t^e + \epsilon_t.$$

2. Obtain the estimates $\hat{M}^{(n)}$ and $\hat{s}^{(n)}$ for each bootstrapped sample of sample size T indexed by n.

3. Compute the pivotal statistic $t^{(n)} = \sqrt{T}\frac{\hat{M}^{(n)} - \hat{M}}{\hat{s}^{(n)}}$.

4. Across all bootstrapped samples, compute $z_{1-\alpha}$ as the $1-\alpha$ quantile of the distribution of $|t^{(n)}|$.

5. The $1-\alpha$ confidence interval for M is computed as $(\hat{M} - \frac{1}{\sqrt{T}}\hat{s}z_{1-\alpha}, \hat{M} + \frac{1}{\sqrt{T}}\hat{s}z_{1-\alpha})$.

6. The $1-\alpha$ coverage rate is defined as the probability of true M falling in the $1-\alpha$ confidence interval across simulations.

[54] We winsorize the data in this step to mitigate the impact of outliers in case of fat-tailed idiosyncratic shocks. As an alternative, we can use the Huber loss function to estimate $(a_i, b_t, \lambda, \eta_t)$, but this is computationally slower than the procedure outlined here.

[55] The intuitive justification is the following. Suppose that the standard error $\hat{s}$ on M is underestimated by a factor of 2, for some reason like a small sample bias. Then $t^{(n)}$ and $z_{1-\alpha}$ are doubled, and nicely the standard error $\frac{1}{\sqrt{T}}\hat{s}z_{1-\alpha}$ is then corrected by the procedure.

F.3 Simulation results

We present the baseline results in Table F.4. In Panel A of the table, we present the results for homoskedastic shocks and we use the homoskedastic GIV algorithms. In Panel B, we simulate heteroskedastic idiosyncratic shocks and use GIV algorithms that adjust for (unknown) heteroskedasticity, as discussed in Section B.

In each panel, there are four sets of results. In the top part of the panel ("Baseline GIV algorithm"), we use the Gaussian loss function. In the bottom part of the panel ("Huber GIV algorithm"), we use the Huber loss function that is robust to fat-tailed idiosyncratic shocks. In the left part of each panel (the first 13 columns), we simulate idiosyncratic shocks that are thin-tailed and normally distributed. In the right part of each panel (the last 11 columns), we simulate the idiosyncratic shocks are drawn from a t_3-distribution.

The true value of $M = 2$ in all simulations. Focusing on the top-left segment of Panel A, we see that the baseline GIV algorithm is unbiased and that the standard errors decrease when concentration increases. In this case, the OLS standard errors are close to the simulated standard errors, which is not directly obvious as we estimate the factor loadings using PCA. As a result, the OLS coverage rates are close to 95%, also for both bootstrap methods.

Turning to the top-right segment of Panel A, where we simulate fat-tailed shocks, we see that the Baseline GIV algorithm remains close to unbiased, but the OLS standard errors overstate the precision of the estimator. The reason is that the factor estimates are severely distorted by the outliers, which is missed by the OLS standard errors. The bootstrap standard errors are close to the simulated standard errors and the boostrapped coverage rates are close to 95%.

If we turn to the bottom segments of Panel A, where we use the Huber GIV algorithm, we first note that the performance is virtually identical to the Baseline GIV algorithm when shocks are thin-tailed (bottom-left segment). Most importantly, in the presence of fat-tailed shocks (bottom-right segment), the Huber GIV algorithm produces tight simulated standard and OLS standard errors are reliable. Interestingly, when comparing the top-left and bottom-right segments of the table, we see that the Huber GIV algorithm yields smaller standard errors in the presence of fat-tailed shocks than the Baseline GIV algorithm in the presence of thin-tailed shocks. This implies that the GIV procedure benefits from fat-tailed shocks and that the Huber GIV algorithm avoids distortions in estimating the factors.

In Panel B, we repeat the analysis but now we simulate heteroskedastic shocks and use GIV algorithms that estimate the shocks' heteroskedasticity. By comparing the top-left segments of Panel A and Panel B, we see that estimating the volatilities of the shocks increases the standard errors, even though the algorithm is close to unbiased. OLS standard errors are too tight in this case and the bootstrap procedures improve the coverage rates, and in particular the non-parametric bootstrap. Using the Baseline GIV algorithm in the presence of fat-tailed shocks leads to large standard errors (top-right segment), but, as before, the Huber GIV algorithm results in unbiased estimates and fairly tight standard errors. The coverage rates of the bootstrap procedure are also

Table F.4: Baseline simulation results. In Panel A of the table, we present the results for homoskedastic shocks and we use the homoskedastic GIV algorithms. In Panel B, we simulate heteroskedastic idiosyncratic shocks and use GIV algorithms that adjust for (unknown) heteroskedasticity as discussed in Section B. In each panel, there are four sets of results. In the top part of the panel ("Baseline GIV algorithm"), we use the Gaussian loss function. In the bottom part of the panel ("Huber GIV algorithm"), we use the Huber loss function that is robust to fat-tailed idiosyncratic shocks. In the left part of each panel (the first 13 columns), we simulate idiosyncratic shocks that are thin-tailed and normally distributed. In the right part of each panel (the last 11 columns), we simulate the idiosyncratic shocks are drawn from a t_3-distribution. There are 50 entities and $T = 100$ time periods. We use 10,000 simulation samples and 1,000 bootstrap samples for each draw. We report the 2.5%, 50%, and 97.5% quantiles across simulations, the mean, as well as the standard deviation ("Sim" under standard errors). Then we report the average OLS standard error ("OLS"), and the standard deviation of the estimate across bootstrap samples ("NP" for non-parametric (i.e. basic) and "W" for wild). We then compute the coverage rates using OLS standard errors and the two bootstrap methods.

Panel A: Homoskedastic data and homoskedastic estimators

Baseline GIV algorithm

							Standard errors			Coverage rate								Standard errors			Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	1.40	2.02	2.64	2.02	0.32	0.28	0.38	0.38	0.92	0.97	0.94	0.22	2.02	3.59	2.02	0.76	0.35	0.61	0.60	0.76	0.96	0.94
0.2	0.4	1.36	2.03	2.68	2.02	0.34	0.27	0.36	0.33	0.89	0.96	0.90	0.66	2.00	2.95	1.96	0.57	0.32	0.42	0.41	0.75	0.86	0.93
0.5	0.2	1.76	2.01	2.24	2.00	0.12	0.12	0.13	0.19	0.94	0.95	0.96	1.61	2.01	2.76	2.01	0.35	0.13	0.23	0.33	0.85	0.96	0.97
0.5	0.4	1.76	2.01	2.25	2.01	0.12	0.11	0.14	0.17	0.93	0.97	0.95	1.44	2.02	2.72	2.02	0.32	0.14	0.25	0.29	0.79	0.96	0.96

Huber GIV algorithm

							Standard errors			Coverage rate								Standard errors			Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	1.41	2.01	2.63	2.01	0.31	0.27	0.34	0.32	0.92	0.97	0.92	1.57	2.03	2.53	2.04	0.24	0.22	0.28	0.27	0.92	0.97	0.94
0.2	0.4	1.42	2.01	2.61	2.02	0.31	0.25	0.29	0.27	0.89	0.94	0.89	1.58	2.02	2.53	2.03	0.24	0.21	0.29	0.24	0.91	0.97	0.91
0.5	0.2	1.69	1.97	2.25	1.97	0.14	0.12	0.13	0.17	0.90	0.92	0.92	1.76	1.99	2.22	1.99	0.12	0.10	0.11	0.16	0.92	0.93	0.94
0.5	0.4	1.70	1.96	2.22	1.96	0.14	0.11	0.13	0.15	0.89	0.93	0.89	1.77	1.98	2.20	1.98	0.11	0.10	0.12	0.14	0.92	0.94	0.93

Panel B: Heteroskedastic data and heteroskedastic estimators

Baseline GIV algorithm

							Standard errors			Coverage rate								Standard errors			Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	0.84	1.95	3.11	1.96	0.58	0.44	0.62	0.53	0.86	0.96	0.89	-0.81	1.99	4.04	1.96	1.10	0.48	0.79	0.68	0.71	0.93	0.87
0.2	0.4	0.72	1.92	2.94	1.90	0.57	0.39	0.46	0.43	0.82	0.89	0.82	0.19	1.94	3.13	1.86	0.74	0.40	0.49	0.45	0.74	0.84	0.83
0.5	0.2	1.39	1.98	2.66	1.99	0.32	0.24	0.32	0.32	0.85	0.94	0.88	0.46	2.00	3.73	2.07	0.77	0.26	0.53	0.55	0.70	0.92	0.88
0.5	0.4	1.33	2.00	2.71	2.01	0.36	0.23	0.30	0.28	0.77	0.87	0.79	0.06	2.01	3.00	1.94	0.69	0.25	0.42	0.37	0.61	0.82	0.78

Huber GIV algorithm

							Standard errors			Coverage rate								Standard errors			Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	0.85	1.92	3.09	1.93	0.57	0.43	0.61	0.48	0.87	0.96	0.89	1.28	1.99	2.80	2.01	0.39	0.32	0.50	0.38	0.90	0.98	0.93
0.2	0.4	0.79	1.92	2.89	1.90	0.55	0.39	0.45	0.39	0.83	0.90	0.84	1.24	2.00	2.86	2.01	0.41	0.31	0.41	0.33	0.87	0.95	0.89
0.5	0.2	1.38	1.97	2.63	1.98	0.32	0.24	0.32	0.30	0.86	0.94	0.87	1.56	2.00	2.52	2.01	0.24	0.20	0.29	0.29	0.89	0.96	0.92
0.5	0.4	1.35	1.99	2.67	2.00	0.34	0.23	0.29	0.26	0.79	0.89	0.80	1.55	2.02	2.63	2.05	0.27	0.19	0.30	0.25	0.84	0.94	0.87

close to 95% when using the Huber GIV algorithm.

We conclude that the Huber GIV algorithm performs well in the presence of non-parametric factors, heteroskedasticity, and outliers.

We extend the simulation results in Table F.5 by exploring the robustness to misspecification. In Panel A of the table, we present the results for homoskedastic shocks and we use the heteroskedastic GIV algorithms as discussed in Section B. In Panel B, we simulate heteroskedastic idiosyncratic shocks and use GIV algorithms that do not adjust for heteroskedasticity. In each panel, there are four sets of results. In the top part of the panel ("Baseline GIV algorithm"), we use the Gaussian loss function. In the bottom part of the panel ("Huber GIV algorithm"), we use the Huber loss function that is robust to fat-tailed idiosyncratic shocks. In the left part of each panel (the first 13 columns), we simulate idiosyncratic shocks that are thin-tailed and normally distributed. In the right part of each panel (the last 11 columns), we simulate the idiosyncratic shocks are drawn from a t_3-distribution.

By comparing Panel A of Table F.5 to Panel A of Table F.4, we find that the results are very comparable. In fact, for the Baseline GIV algorithm and in the presence of fat-tailed shocks (the top-right segment of Panel A), the procedure that adjusts for heteroskedasticity produces somewhat smaller standard errors as it limits the impact of outliers. The important practical takeaway is that using algorithms that adjust for heteroskedasticity do not lead to worse estimates when the underlying data are homoskedastic.

Panel B of Table F.5 shows, however, that estimating the model using homoskedastic procedures when the underlying data are heteroskedastic can lead to biased estimates. This effect is more pronounced for the Baseline GIV algorithm compared to the Huber GIV algorithm, as the latter is close to unbiased (bottom segments of Panel B).

These results strengthen our earlier conclusion that the Huber GIV algorithm performs well in the presence of non-parametric factors, heteroskedasticity, and outliers.

G General setup and multipliers

We now propose a more general setup with potentially several factors and rich heterogeneity.

G.1 Framework

Consider the following model of outcome variables y_{it} (such as employment, investment, TFP shocks, returns, and so on)for "actor" i (e.g., a firm or industry i in a closed-economy setting, or a country i in an international setting):

$$y_{it} = \sum_f \lambda_{it}^f F_t^f + u_{it} + C_{it}^y m, \tag{110}$$

Table F.5: Simulation results: The impact of model misspecification. In Panel A of the table, we present the results for homoskedastic shocks and we use the heteroskedastic GIV algorithms as discussed in Section B. In Panel B, we simulate heteroskedastic idiosyncratic shocks and use GIV algorithms that do not adjust for heteroskedasticity. In each panel, there are four sets of results. In the top part of the panel ("Baseline GIV algorithm"), we use the Gaussian loss function. In the bottom part of the panel ("Huber GIV algorithm"), we use the Huber loss function that is robust to fat-tailed idiosyncratic shocks. In the left part of each panel (the first 13 columns), we simulate idiosyncratic shocks that are thin-tailed and normally distributed. In the right part of each panel (the last 11 columns), we simulate the idiosyncratic shocks are drawn from a t_3-distribution. There are 50 entities and $T = 100$ time periods. We use 10,000 simulation samples and 1,000 bootstrap samples for each draw. We report the 2.5%, 50%, and 97.5% quantiles across simulations, the mean, as well as the standard deviation ("Sim" under standard errors). Then we report the average OLS standard error ("OLS"), and the standard deviation of the estimate across bootstrap samples ("NP" for non-parametric (i.e. basic) and "W" for wild). We then compute the coverage rates using OLS standard errors and the two bootstrap methods.

Panel A: Homoskedastic data and heteroskedastic estimators

Baseline GIV algorithm

						Standard errors				Coverage rate							Standard errors				Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	1.36	2.00	2.65	2.00	0.33	0.28	0.39	0.37	0.91	0.98	0.93	0.44	2.01	3.52	2.03	0.66	0.31	0.58	0.57	0.82	0.98	0.94
0.2	0.4	1.33	2.01	2.67	2.01	0.35	0.27	0.36	0.32	0.88	0.95	0.89	0.72	2.01	2.92	1.98	0.54	0.30	0.42	0.40	0.78	0.92	0.92
0.5	0.2	1.76	2.00	2.25	2.00	0.13	0.12	0.13	0.18	0.93	0.96	0.95	1.75	2.00	2.37	2.02	0.29	0.12	0.19	0.29	0.91	0.96	0.96
0.5	0.4	1.76	2.01	2.26	2.01	0.13	0.11	0.15	0.16	0.92	0.97	0.94	1.69	2.01	2.72	2.04	0.30	0.13	0.23	0.28	0.86	0.97	0.96

Huber GIV algorithm

						Standard errors				Coverage rate							Standard errors				Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	1.32	1.96	2.61	1.96	0.33	0.28	0.39	0.34	0.91	0.97	0.92	1.51	1.99	2.52	2.00	0.25	0.22	0.31	0.29	0.92	0.97	0.93
0.2	0.4	1.30	1.97	2.62	1.97	0.34	0.27	0.35	0.29	0.88	0.94	0.88	1.50	1.99	2.54	2.00	0.26	0.22	0.33	0.25	0.90	0.97	0.91
0.5	0.2	1.74	1.99	2.24	1.99	0.13	0.12	0.13	0.18	0.93	0.95	0.94	1.80	2.00	2.23	2.00	0.11	0.10	0.11	0.17	0.93	0.95	0.95
0.5	0.4	1.74	2.00	2.25	1.99	0.13	0.12	0.15	0.15	0.92	0.96	0.93	1.79	2.00	2.23	2.00	0.11	0.10	0.13	0.14	0.93	0.96	0.94

Panel B: Heteroskedastic data and homoskedastic estimators

Baseline GIV algorithm

						Standard errors				Coverage rate							Standard errors				Coverage rate		
h	θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	0.60	1.62	2.73	1.62	0.55	0.36	0.48	0.44	0.69	0.82	0.80	-0.80	1.63	3.43	1.58	1.04	0.44	0.62	0.53	0.58	0.79	0.82
0.2	0.4	0.57	1.61	2.46	1.58	0.50	0.33	0.36	0.36	0.67	0.73	0.77	0.04	1.58	2.60	1.51	0.66	0.35	0.41	0.38	0.64	0.71	0.78
0.5	0.2	0.90	1.64	2.53	1.66	0.44	0.21	0.25	0.27	0.48	0.55	0.68	-0.15	1.66	3.14	1.68	0.82	0.24	0.42	0.40	0.39	0.67	0.75
0.5	0.4	0.87	1.63	2.35	1.63	0.41	0.20	0.23	0.23	0.50	0.57	0.69	-0.05	1.63	2.47	1.53	0.66	0.22	0.32	0.27	0.50	0.65	0.74

Huber GIV algorithm

						Standard errors				Coverage rate							Standard errors				Coverage rate		
h	\θ	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W	p2.5	p50	p97.5	Mean	Sim	OLS	NP	W	OLS	NP	W
0.2	0.2	1.01	1.95	3.00	1.97	0.51	0.37	0.47	0.41	0.85	0.94	0.85	1.34	2.01	2.77	2.02	0.36	0.29	0.42	0.35	0.90	0.97	0.91
0.2	0.4	1.05	1.99	2.75	1.96	0.45	0.34	0.35	0.35	0.87	0.90	0.86	1.39	2.05	2.80	2.06	0.36	0.28	0.34	0.30	0.87	0.93	0.88
0.5	0.2	1.08	1.76	2.55	1.78	0.40	0.21	0.24	0.25	0.58	0.66	0.65	1.25	1.85	2.47	1.86	0.31	0.18	0.24	0.25	0.68	0.79	0.77
0.5	0.4	1.17	1.81	2.44	1.81	0.35	0.19	0.22	0.22	0.63	0.69	0.68	1.34	1.88	2.46	1.89	0.29	0.17	0.23	0.21	0.71	0.82	0.76

where each F_t^f is a factor, λ_{it}^f is factor loading, u_{it} is an idiosyncratic shock, and C_{it}^y is a vector of controls that may include lagged demands and other characteristics. We could also add constants, but we omit them for notational simplicity. Factor f follows:

$$F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f. \tag{111}$$

It depends on an exogenous shock η_t^f, and potentially on the mean action y_{St}, and on a set of controls C_t^f (potentially different from C_{it}^y). Those controls may include, for instance, lagged values. We assume that the "size" weights have been normalized to add to one, $\sum_i S_i = 1$.

We use the structure (110)-(111) because many economic models of interest follow this structure, at least after linearization, so that the GIV allows to estimate some of their parameters.

We partition the factors into "exogenous factors", where we know $\alpha^f = 0$, and "endogenous" factors where α^f may be non-zero. As in the rest of the paper, we make the mild assumption that all our variables (e.g. η_t^f, u_t) have finite second moments

In the baseline case here we study the parametric case. We have some characteristics x_{it} of actors: for instance, depending on the application we know that the loading is an affine function of log market capitalization, or the stock market beta of a bank, or OPEC membership. We also have a priori knowledge that for some parameter $\dot{\lambda}^f$ to be estimated we have:

$$\lambda_{it}^f = \dot{\lambda}_0^f + \dot{\lambda}_1^f x_{it}^f, \tag{112}$$

This is consistent with the practice in modern finance in which risk exposures (betas) align with characteristics (see e.g. Fama and French (1993)), so that parametric approaches are preferred, in particular because they are more stable than non-parametric approaches.

We make the following identifying assumptions. For all f, i, the shocks u_{it} are idiosyncratic:

$$\mathbb{E}\left[u_{it}\left(\eta_t^f, C_t^y, C_t^f, x_t^f\right)\right] = 0, \tag{113}$$

but the η_t^f may be correlated across f's, and η_t^f may be correlated with the controls, C_t^y and C_t^f. The u_{it} may have some correlation across i's and can be heteroskedastic, as we discuss later. For expositional simplicity we assume that all dates are i.i.d.

We rewrite model (110) in vector form:

$$y_t = \Lambda_t F_t + u_t + C_t^y m, \qquad F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f, \tag{114}$$

with Λ_t a $N \times r$ matrix, F_t a $r \times 1$ vector, C_t^y an $N \times c$ matrix, m is $c \times 1$, where c is the dimension of the controls.[56]

[56] Our initial examples are particular cases of the general procedure.

G.2 Multipliers

Solving the model gives $y_{St} = \Lambda_{St} F_t + u_{St} + C^y_{St} m$, that is, $y_{St} = \Lambda_{St} \alpha y_{St} + u_{St} + \varepsilon^y_t$, where α is the vector stacking the α^f's and ε^y_t satisfies $\varepsilon^y_t \perp u_t$. So, we can solve for the aggregate outcome y_{St} as $y_{St} = \frac{u_{St}+\varepsilon^y_t}{1-\Lambda_{St}\alpha}$, that is,

$$y_{St} = M_t \left(u_{St} + \varepsilon^y_t \right), \tag{115}$$

where the multiplier M_t measures the total impact of shocks, after going through all general equilibrium effects (where we assume that the denominator is not 0):

$$M_t = \frac{1}{1 - \Lambda_{St}\alpha} = \frac{1}{1 - \sum_f \Lambda^f_{St} \alpha^f}. \tag{116}$$

Hence, an idiosyncratic shock has an impact on the aggregate action y_{St} that is M_t times bigger than its direct effect. Also, the total impact of an idiosyncratic shock on factor f is:

$$F^f_t = M_t \alpha^f u_{St} + \varepsilon^f_t, \tag{117}$$

where it again holds that $\varepsilon^f_t \perp u_{St}$. This shows intuitively, and we will prove formally below, that our regressions will allow to identify M_t and $M_t \alpha^f$.

In some cases, we may not observe all endogenous factors, F^f_t. In this case, we still recover the correct multiplier, M_t, and it should be interpreted as accounting for all general equilibrium effects in the economy, including those operating via the unobservable, endogenous factors. However, we can obviously not estimate α^f for those unobserved factors.

G.3 A formal identifiability result

We provide here formal conditions for identification, completing the simpler case of Section 2. We study the parametric case. Section H.4 develops the full non-parametric version, estimating the factors. We don't have a priori information about the η_t, nor their variance V^η.

Assumption 3 (Condition for identification with GIV) *The vector $V^u S$ is not spanned by the factors loadings λ^f (where V^u is the covariance matrix of u_t).*

Assumption 3 ensures that the GIV is not identically 0 (as $z_t := S' Q^{\Lambda,(V^u)^{-1}} u_t$, as in (**??**) and (31)). Economically, this assumption seems like a mild restriction. It is generically satisfied.[57,58]

For simplicity, we shall make here a strong further Assumption 4, which can be relaxed.

Assumption 4 (Known form of the variance matrix of the idiosyncratic shocks) *The u_{it}'s are homoskedastic, or, more generally, the econometrician knows the matrix V^u up to a proportionality*

[57] See Footnote 20.

[58] This also suggest that to control for size, one wants to use log size, but not absolute size.

factor. In addition, V^u is invertible.

Though this would be easy to relax, we assume that all shocks are i.i.d. over time, that Λ_{St} is constant, so that M_t is constant. We assume that second moments are finite for all random variables $\left(\eta_t, u_{it}, \eta_t^f, C_{it}^y, C_t^f, x_{it}\right)$.

We next state a formal identification result, which is proven by method very similar to those of Section C.8. The complete proof is in Section C.5 of the NBER Working Paper version of this paper.

Proposition 12 (Sufficient condition for identification with GIV) *Consider the factor model above, when N is fixed but $T \to \infty$, and make Assumption 3 and 4. We assume the parametric case, where we know the actor (e.g., firm or country) characteristics X_{it}. Then, we can identify α^f and M by GIV. Furthermore, the standard errors on M and $\alpha^f M$ returned by OLS (using the GIV) in this procedure are valid.*

Proposition 12 shows identification in the case with parametric factors. We conjecture that it also holds for the case of non-parametric factors. As a partial substitute, we provide numerical simulations that support the view that the procedure also works in the latter case. But given the complexity of that case, we defer it to future research as one of the several interesting extensions of the GIV.

H Complements

H.1 Multi-dimensional actions

Suppose now that the outcome or action y_{it} and idiosyncratic noise u_{it} are $q-$dimensional, for some $q \geq 1$. For instance, y_{it}'s components might be the growth rate and the labor share of firm i, and then $q = 2$. Then, the general GIV procedure extends well, as we shall now see.

We call $a \in \{1, \dots, q\}$ (as in action) a component of y. We consider the model

$$\begin{aligned} y_{S^a t}^a &= \sum_f \lambda_{S^a, f}^a F^f + u_{S^a t}^a, \\ F_t^f &= \eta_t^f + \sum_a \alpha_a^f y_{S^a, t}^a, \end{aligned}$$

Here u_{it} is q dimensional, α is a $r \times q$ dimensional matrix, and λ_S is a $q \times r$ dimensional matrix.

We can also estimate M (hence $\sum_f \lambda^f \alpha^f$), the α^f. Indeed, for ε_t a composite of aggregate shocks,

$$y_{St} = H y_{St} + u_{St} + \varepsilon_t,$$

where

$$H = \Lambda A = \sum_f \lambda^f \alpha^f,$$

with $\Lambda_{af} = \lambda^a_{S^a,f}$ and $A_{fa} = \alpha^f_a$ matrices with dimensions $q \times r$ and $r \times q$ respectively, so that H is $q \times q$, and

$$u_{St} = \left(u^a_{S^a t}\right)_{a=1,\dots,q}.$$

This implies

$$y_{St} = M\left(u_{St} + \varepsilon_t\right), \tag{118}$$

where the multiplier M is now a $q \times q$ matrix:

$$M = (I - H)^{-1}.$$

We will form a GIV:

$$z_t = u_{\Gamma t},$$

which is $q-$dimensional: $u_\Gamma = \left(u^a_{\Gamma^a}\right)_{a=1,\dots,q}$. We want, with $E^a = S^a - \Gamma^a$,

$$\mathbb{E}\left[u_{Et} u'_{\Gamma t}\right] = 0$$

i.e., for all a, b, $\Omega^{ab} = 0$, where

$$\Omega^{ab} := \mathbb{E}\left[u^a_{E^a t} u^b_{\Gamma^b t}\right].$$

Let us focus on the case where u_{it}, u_{jt} are uncorrelated for $i \neq j$, but for a given i, u^a_{it}, u^b_{it} can be correlated. (If a firm has an investment boom, it will likely hire more labor, so that the components of its idiosyncratic shock in $y_{it} \in \mathbb{R}^q$ will be correlated.)

We have:

$$\Omega^{ab} = \sum_i E^a_i \Gamma^b_i v^{ab}_i, \qquad v^{ab}_i := \mathbb{E}\left[u^a_{it} u^b_{it}\right]. \tag{119}$$

For simplicity, we will suppose that that there are v^{ab} and σ^2_i such that

$$v^{ab}_i = \sigma^2_i v^{ab}. \tag{120}$$

Hence, we can simply take $E_i = \frac{k}{\sigma^2_i}$ with $k = \frac{1}{\sum_j 1/\sigma^2_j}$ and set, for all a, $E^a_i = E_i$ and $\Gamma^a = S^a - E^a$. Then,

$$\Omega^{ab} = \sum_i \frac{k}{\sigma^2_i} \Gamma^b_i \sigma^2_i v^{ab} = k v^{ab} \sum_i \Gamma^b_i = 0,$$

so that we have achieved our goal that $\mathbb{E}\left[u_{Et} u'_{\Gamma t}\right] = 0$. In the more general case, other Γ^a_i can probably be found.

Given (118), we have

$$y_{St}=M\left(u_{St}+\varepsilon_t\right)=M\left(u_{\Gamma t}+u_{Et}+\varepsilon_t\right),$$

so

$$\mathbb{E}\left[y_{St}z_t'\right]=M\mathbb{E}\left[z_tz_t'\right],$$

hence our estimator is

$$M=\mathbb{E}\left[y_{St}z_t'\right]\mathbb{E}\left[z_tz_t'\right]^{-1}. \tag{121}$$

Finally, we can also estimate $\alpha^f M$ by regressing on z_t:

$$F_t^f=\eta_t^f+\sum_a \alpha_a^f y_{S^a,t}^a=\eta_t^f+\alpha^f y_{St}=\eta_t^f+\alpha^f M\left(u_{\Gamma t}+u_{Et}+\varepsilon_t\right),$$

so $\beta^f=\alpha^f M$ (a row vector) obtains by simply regressing

$$F_t^f=\beta^f z_t+\varepsilon_t^f,$$

and get $\beta^f=\alpha^f M$, $\beta^f=\mathbb{E}\left[F_t^f z_t'\right]\mathbb{E}\left[z_tz_t'\right]^{-1}$.

Extension: causal estimation of the actor-specific multiplier The following is a refinement. We can also identify causally $\mu_i:=\lambda_i\alpha=\sum_f \lambda_i^f\alpha^f$. Indeed, use

$$u_{\Gamma t,-i}:=u_{\Gamma t}-S_i^u u_{it}, \tag{122}$$

which is the granular shock purged of a correlation with u_{it}. Then, a shock u_{St} creates an impact $\frac{dF_t}{du_{St}}=M\alpha$, hence an impact

$$\frac{dy_{it}}{du_{St}}=\lambda_i M\alpha.$$

Hence, we can identify μ_i, by regression

$$y_{it}=\mu_i M u_{\Gamma t,-i}+\phi^i\mathcal{C}_t+\varepsilon_{it}^y, \tag{123}$$

with some noise ε_{it}^y. This is the average impact of a causal impact of idiosyncratic shocks of the other entities on entity i.

H.2 Nonlinear GIV

We imagine a nonlinear GIV. Suppose that instead of the simple $s_t=\phi^s p_t+\varepsilon_t$ (equation (2)) we have a more complex

$$s_t=\Phi\left(p_t,\phi^s\right)+\varepsilon_t \tag{124}$$

for Φ a nonlinear function. We can use the moment:

$$\mathbb{E}\left[\left(s_t - \Phi\left(p_t, \phi^s\right)\right) z_t\right] = 0 \tag{125}$$

and can still identify a one-dimensional ϕ^s. For a higher-dimensional ϕ^s, we might add z_t^2 as instrument, though the instrument becomes weaker.

H.3 When the researcher assumes too much homogeneity

Take the supply and demand example, and imagine that the econometrician assumes a homogeneous elasticity of demand ϕ^d, even though there are in fact heterogeneous elasticities ϕ_i^d. What happens then?

The model (1)-(2) becomes, for the demand:

$$y_{it} = \phi_i^d p_t + \lambda_i \eta_t + u_{it},$$

and for the supply

$$s_t = \phi^s p_t + \varepsilon_t.$$

As supply equals demand, $y_{St} = s_t$, which gives the price

$$p_t = \frac{u_{St} + \lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi_S^d}. \tag{126}$$

In this thought experiment, the econometrician assumes identical elasticities of demand across countries, $\phi_i^d = \phi^d$. He runs a panel model for $y_{it} - y_{Et}$, and we assume that it's large enough that he can extract η_t successfully.[59] The GIV (we use the notation Z_t rather than z_t to denote the GIV before controls by η_t) is then

$$Z_t := y_{\Gamma t} = \phi_\Gamma^d p_t + \lambda_\Gamma \eta_t + u_{\Gamma t} = \left(1 + \frac{\phi_\Gamma^d}{\phi^s - \phi_S^d}\right) u_{\Gamma t} + \lambda^Z \tilde{\eta}_t = \frac{1}{\psi} u_{\Gamma t} + \lambda^Z \tilde{\eta}_t,$$

so

$$Z_t = \frac{1}{\psi} u_{\Gamma t} + \lambda^Z \tilde{\eta}_t, \qquad \frac{1}{\psi} = \frac{\phi^s - \phi_E^d}{\phi^s - \phi_S^d}, \tag{127}$$

where $\frac{1}{\psi} = 1$ in the homogeneous-elasticity case, $\tilde{\eta}_t = (\eta_t, \varepsilon_t, u_{Et})$ gathers the common shocks, and λ^Z is a vector of loadings.

Hence, when we run the first stage

$$p_t = b^p Z_t + \beta^p \eta_t + \varepsilon_t^p,$$

[59] One of the factors, formally, will be p_t. We assume that it is not included in the vector of factors η_t.

we will gather

$$b^p = \frac{1}{\phi^s - \phi^d_E}.$$

If we run

$$s_t = b^s Z_t + \beta^s \eta_t + \varepsilon^s_t,$$

we will estimate

$$b^s = \frac{\phi^s}{\phi^s - \phi^d_E}.$$

The ratio of the two coefficients still gives ϕ^s. Likewise, the IV on the elasticity of demand will give ϕ^d_E.

In the polar opposite case where η_t cannot be estimated or controlled for, then the simple procedure becomes biased, however, as (127) shows. To fix it, one can estimate the model with non-parametric coefficients (Section H.4).

H.4 Heterogeneous demand elasticities: Non-parametric extension

Non-parametric version for ϕ^d We present a variant of the procedure in Section 4.2, but now with non-parametric heterogeneous demand elasticities ϕ^d_i. The model is

$$y_t = \phi^d p_t + \lambda \eta_t + u_t, \tag{128}$$

We still assume parametric loading of unobserved factors η.

We propose two procedures to estimate ϕ^d.

H.4.1 First procedure for the nonparametric estimation of heterogeneous demand elasticities

Recall the model (128). We replace ϕ^d by ϕ for simplicity:

$$y_t = \phi p_t + \lambda \eta_t + u_t, \tag{129}$$

Unlike earlier, we now do not assume parametric knowledge of λ. We propose the following procedure.

1. Guess a candidate for ϕ, called ϕ^c, and $W = (V^u)^{-1}$ (initially, as it's enough to know all those up to a multiplicative factor, we might take $\phi^c = \iota$, and $W = I$, or $W = Diag\left(1/var\left(y_{it}\right)\right)$). We define $Q^\phi := Q^{\phi^c, W}$, keeping W implicit in this step and the next. If $\phi^c = \phi$, then

$$Q^\phi y_t = \left(Q^\phi \lambda\right) \eta_t + Q^\phi u_t \tag{130}$$

2. We can apply the "singular factor analysis" procedure of Section H.11 to $Q^\phi y_t$ (so, in the

notation of that section, $G = Q^{\phi}$). This returns: $\check{\lambda} := Q^{\phi}\lambda$, η_t^e, V^u, $\check{u}_t = Q^{\check{\lambda},\phi}u_t$. We form $z_t := S'\check{u}_t$, and $\Gamma := \left(Q^{\check{\lambda},\phi}\right)' S$.

3. We estimate the vector of sensitivities ϕ. We use a specific instrument z_{it} for each entity i. We proceed as follows:

   (a) We define the debiasing vector a^i. As $\check{u}_t = Qu_t$, we have $V^{\check{u}} = QV^uQ'$, and we define

$$a_j^i := \frac{V_{ij}^{\check{u}}}{V_{ii}^{\check{u}}} \tag{131}$$

   (b) We define the instrument for entity i,

$$z_{it} := S'\left(\check{u}_t^e - a^i \check{u}_{it}^e\right) \tag{132}$$

   Morally, it's the size weighted sum idiosyncratic shocks of the entities different from i.[60]

   (c) We use the following moment to identify ϕ_i:

$$\mathbb{E}\left[\left(y_{it} - \phi_i p_t - \lambda_i \eta_t^e\right) z_{it}\right] = 0 \tag{135}$$

4. Given this new estimates of ϕ and V^u, we go back to step 1-3, and loop until convergence.

This algorithm also applies to the parametric case where we know that $\phi_{it} = X_i\dot{\phi}$ (Section 4.2), but keep the loadings λ non-parametric. Then, in steps 1-2 we replace ϕ by X, and in the last step we replace ϕ by $X\dot{\phi}$ and estimate $\dot{\phi}$.

Proposition 13 (Moment conditions to identify non-parametric elasticities). *Define $\check{u}_t = Q^{\check{\lambda},\phi}u_t$ in the notations above, and the entity-i specific GIV z_{it} defined in* (132). *Then the moment condition* (135) *holds.*

Proof of Proposition 13. Definition (131), together with $\check{u}_t = Qu_t$ and $V^{\check{u}} = QV^uQ'$, implies

$$a_j^i = \frac{\mathbb{E}\left[\check{u}_{jt}\check{u}_{it}\right]}{\mathbb{E}\left[\check{u}_{it}^2\right]} \tag{136}$$

[60] Indeed, if we had no common shocks, we'd have

$$a_j^i = 1_{i=j} \tag{133}$$

so that $z_{it} = \sum_{j\neq i} S_j \check{u}_{jt}^e$ is a "leave one out" estimator. In the general case, z_{it} is in some sense a refined quasi-leave one out estimator, refined so that (137) holds. In the simple case of Section 2.3 with an additive shock ($\lambda_i \equiv 1$), we just have $\check{u}_{it} = u_{it} - u_{Et}$, i.e. $Q_{ij} = 1_{i=j} - \frac{1}{N}$ and

$$a_j^i := \frac{1_{i=j} - \frac{1}{N}}{1 - \frac{1}{N}}. \tag{134}$$

This may be useful as a starting point numerically.

so that

$$\mathbb{E}\left[z_{it}\check{u}_{it}\right]=\mathbb{E}\left[S'\left(\check{u}_t-a^i\check{u}_{it}\right)\check{u}_{it}\right]=0 \tag{137}$$

As z_{it} is also uncorrelated with η_t^e, moment (135) holds. □

H.4.2 Second procedure for the nonparametric estimation of heterogeneous demand elasticities

We premultiply (128) by $Q=Q^\lambda$ and set $\check{x}_t:=Qx_t$. So $\check{y}_t=\check{\phi}^d p_t+\check{u}_t$. With $\Gamma=Q'S$, we have $y_{\Gamma t}=\phi_\Gamma^d p_t+u_{\Gamma t}$. To ease on notations, we call $\psi:=\phi_\Gamma^d$. Given a candidate estimate ψ^c of ψ we form the associated GIV: $z_t\left(\psi^c\right):=y_{\Gamma t}-\psi^c p_t$.

If we have the correct z_t, the following moments hold[61,62], with $b^p=\frac{1}{\phi^s-\phi_S^d}$ the coefficient of the first stage regression (20), $p_t=b^p z_t+\varepsilon_t^p$,

$$\begin{aligned}
\mathbb{E}\left[\left(y_t-\phi^d p_t\right)z_t\right] &= V^u\Gamma, \qquad \mathbb{E}\left[\left(p_t-b^p z_t\right)z_t\right]=0 &(140)\\
\mathbb{E}\left[\left(\check{y}_{it}-\check{\phi}_i^d p_t\right)^2\right] &= V_{ii}^{\check{u}}, \qquad \mathbb{E}\left[z_t^2\right]=\Gamma' V^u\Gamma &(141)
\end{aligned}$$

which potentially allow to estimate, respectively, ϕ^d, b^p (hence ϕ^s), V_{ii}^u and ϕ_Γ^d. Indeed, if we know z_t, we know ϕ^d and b^p.[63]

We examine in more detail how to estimate $\psi:=\phi_\Gamma^d$. Calling the true value $z_t(\psi)=u_{\Gamma t}$, we have $\mathbb{E}\left[z_t^2\right]=\sigma_{u_\Gamma}^2$, where $\sigma_{u_\Gamma}^2=\Gamma' V^u\Gamma$ is the theoretical variance of z_t given in (141). So, we solve for ψ^c (a *c*andidate answer for ψ) so that the empirical variance of the GIV is equal to its theoretical variance:

$$\mathbb{E}\left[z_t\left(\psi^c\right)^2\right]-\sigma_{u_\Gamma}^2=0$$

i.e. $\mathbb{E}\left[p_t^2\right]\left(\psi^c\right)^2-2\mathbb{E}\left[y_{\Gamma t}p_t\right]\psi^c+\mathbb{E}\left[y_{\Gamma t}^2\right]-\sigma_{u_\Gamma}^2=0$. This is a quadratic equation in ψ^c, which yields two roots:[64] a good (i.e. correct) root, $\psi^G=\psi$, and a bad root, $\psi^B=\psi+2\frac{\mathbb{E}\left[z_t^* p_t\right]}{\mathbb{E}\left[p_t^2\right]}$. Fortunately,

[61] Indeed, we should have $\mathbb{E}\left[\left(y_t-\phi^d p_t\right)z_t\right]=\mathbb{E}\left[u_t z_t\right]=\mathbb{E}\left[u_t\left(u_t'\Gamma\right)\right]=V^u\Gamma$. Also, as $\check{u}=\check{y}-\check{\phi}p$, and $V^{\check{u}}=QV^uQ'$.

[62] As a variant, we decompose into the equal weighted version, which gives ϕ_E (we premultiply by $\tilde{E}'$):

$$\mathbb{E}\left[\left(y_{\tilde{E}t}-\phi_{\tilde{E}}p_t\right)z_t\right]=0 \tag{138}$$

and the deviation from the mean, which gives $\check{\phi}_i$ via:

$$\mathbb{E}\left[\left(\check{y}_{it}-\check{\phi}_i p_t\right)z_t\right]=\left(QV^u\Gamma\right)_i \tag{139}$$

[63] We recommend starting from the parametric estimates of Section 4.2, which gives potentially good starting values for ϕ^d, z_t and V^u.

[64] Indeed, calling $\psi^\Delta:=\psi^c-\psi$ the error, we have

$$0=\mathbb{E}\left[z_t\left(\psi^c\right)^2\right]-\sigma_{u_\Gamma}^2=\mathbb{E}\left[\left(z_t^*-\psi^\Delta p_t\right)^2\right]-\mathbb{E}\left[z_t^{*2}\right]=-2\psi^\Delta\mathbb{E}\left[z_t^* p_t\right]+\left(\psi^\Delta\right)^2\mathbb{E}\left[p_t^2\right]$$

there is an economic way to determine which is the correct root. Calling G (resp. B) the estimation with the good (resp. bad) root, one can show that:

$$b^{p,B} = -b^{p,G}, \tag{142}$$

Hence, if we have a prior on the sign of of the first stage coefficient b^p (e.g. we know that $b^p > 0$ in a demand and supply model), we can choose the correct root as the one yielding a positive b^p in the first stage.

Justification of the proposed procedure Consider an econometrician who would use the bad root:

$$z_t^B = y_{\Gamma t} - \phi_\Gamma^B p_t = u_{\Gamma t} + \phi_\Gamma p_t - \phi_\Gamma^B p_t = z_t - \beta p_t, \qquad \beta = 2\frac{\mathbb{E}[z_t p_t]}{\mathbb{E}[p_t^2]}$$

This bad root satisfies $\mathbb{E}\left[z_t^B p_t\right] = \mathbb{E}\left[(z_t - \beta p_t) p_t\right] = -\mathbb{E}[z_t p_t]$, so:

$$\mathbb{E}\left[z_t^B p_t\right] = -\mathbb{E}[z_t p_t], \qquad \mathbb{E}\left[\left(z_t^B\right)^2\right] = \mathbb{E}\left[z_t^2\right] \tag{143}$$

Hence, when estimating b^p in the "first stage" via $\mathbb{E}\left[(p_t - b^p z_t) z_t\right] = 0$, the econometrician will find:

$$b^{p,B} = \frac{\mathbb{E}\left[p_t z_t^B\right]}{\mathbb{E}\left[\left(z_t^B\right)^2\right]} = \frac{-\mathbb{E}[p_t z_t]}{\mathbb{E}[z_t^2]} = -b^{p,eG} \tag{144}$$

Hence, the coefficient in the first stage will have the wrong sign. This allows to find the correct root.

A more general argument We show how even with other procedures there are two roots for a nonparametric model with heterogeneous elasticities, and that fortunately (as in our outlined procedure) there is a simple economic way to identify the correct root. The model is, in vector form:

$$y_t = \lambda \eta_t + \phi p_t + u_t, \qquad p_t = \alpha y_{St} + \dot{\varepsilon}_t$$

with $\alpha = \frac{1}{\phi^s}$, and we use notation $\dot{\varepsilon}_t$ as we wish to keep the simpler notation ε_t for later. So solving for $y_{St} = M\left(\lambda_S \eta_t + \phi_S \dot{\varepsilon}_t + u_{St}\right)$, $M = \frac{1}{1-\phi_S \alpha}$, we get, for a properly defined $\ddot{\varepsilon}_t$ (an unimportant linear combination of $\dot{\varepsilon}_t$ and η_t), $p_t = \alpha M u_{St} + \ddot{\varepsilon}_t$, hence:

$$y_t = \lambda \eta_t + \phi \ddot{\varepsilon}_t + \alpha M \phi u_{St} + u_t, \qquad p_t = \ddot{\varepsilon}_t + \alpha M u_{St}$$

We wish to estimate ϕ and αM.

We consider the vector $Y_t = (y_t', p_t)'$ stacking together y_t and p_t. Then, with $U_t = (u_t', 0)'$, $\Phi = (\phi', 1)'$, $\Lambda = (\lambda', 0)'$, and adding a weight "0" to the last component of the vector S (extended here to have 1 more component, with a mild abuse of notations) we have[65]

$$Y_t = \Lambda\eta_t + \Phi\ddot{\varepsilon}_t + \alpha M\Phi u_{St} + U_t \tag{145}$$

i.e., with $\Psi := \alpha M\Phi$, and $\varepsilon_t := \frac{1}{\alpha M}\ddot{\varepsilon}_t$,

$$Y_t = \Lambda\eta_t + \Psi\varepsilon_t + \Psi u_{St} + U_t = \Lambda\eta_t + \Psi\varepsilon_t + (I + \Psi S')\, U_t \tag{146}$$

All the information is in $V^Y = \mathbb{E}\left[Y_t Y_t'\right]$:

$$\begin{aligned} V^Y &= \sigma_\eta^2\Lambda\Lambda' + \sigma_\varepsilon^2\Psi\Psi' + \sigma_{\eta\varepsilon}\left(\Lambda\Psi' + \Psi\Lambda'\right) + (I + \Psi S')\, V^U\left(I + S\Psi'\right) && (147)\\ &= \sigma_\eta^2\Lambda\Lambda' + \gamma\Psi\Psi' + \Psi b' + b\Psi' + V^U && (148)\\ b &= \sigma_{\eta\varepsilon}\Lambda + V^U S && (149)\\ \gamma &= \sigma_\varepsilon^2 + S'V^U S && (150)\end{aligned}$$

The idea for the multiplicity of roots in Ψ is that we have a second degree equation in Ψ, so that we can have multiple roots – like in the one-dimensional case. Let us next calculate the roots, which will lead to a procedure to identify the correct root. Forming the vector $a = \frac{-1}{\gamma}b$, we have

$$(\Psi - a)\,(\Psi - a)' = C := \frac{1}{\gamma}\left(V^Y - V^U - \sigma_\eta^2\Lambda\Lambda'\right) + aa' \tag{151}$$

Suppose that we have estimated all the parameters, and it remains to estimate Ψ, i.e. solve for Ψ^c (as in a candidate value for Ψ) the equation:

$$(\Psi^c - a)\,(\Psi^c - a)' = C$$

We know that this identity holds under the correct root, so that $C = (\Psi - a)\,(\Psi - a)'$. Now, there are two solutions to the equation $XX' = DD'$, with X the unknown vector and D a known vector: $X = D$ and $X = -D$. Hence, the two solutions are $\Psi^c - a = \Psi - a$ and $\Psi^c - a = -\,(\Psi - a)$. The first one is the good root, $\Psi^G = \Psi$, and the second one is the bad root:

$$\Psi^B = 2a - \Psi \tag{152}$$

Now, because Λ_p and S_p (i.e., the component of those vectors on the last coordinate, corresponding to p) are both 0, we have $b_p = 0$ (see 149) and thus $a_p = 0$. So the component of the bad root on the price is $\Psi_p^B = 2a_p - \Psi_p = -\Psi_p$:

$$\Psi_p^B = -\Psi_p \tag{153}$$

[65] This idea of stacking together then y_t and p_t, with a "size 0" for the innovations to the price, could be fruitfully used more generally.

This allows to distinguish between the two roots, as the right one has $\Psi_p = \alpha M$ and the other one has $\Psi_p = -\alpha M$. Hence, if economic reasoning tells us the sign of αM (e.g., it is positive in a supply and demand context), we can pick the good root by inspecting the sign of Ψ_p.

H.5 Relation between Bartik instruments and GIVs

H.5.1 Relating the Bartik setup to the GIV setup

Bartik instruments are widely used to estimate parameters of interest. In this appendix, we compare the assumptions under which Bartik instruments are valid to those under which GIVs are valid. This comparison is useful also to highlight settings in which GIVs can and cannot be used (and vice versa for Bartik).

As a general matter, in a number of cases where a cross-section is used (e.g. Autor et al. (2013)), Bartik applies, but GIV does not apply, for instance because there is no large idiosyncratic shock that one can use.

Next, to study the difference between GIV and Bartik more analytically, we start from the setup in Borusyak et al. (2022), and then map it to our model.[66] Their model can be summarized as

$$y_l = \beta x_l + \epsilon_l,$$

where we omit observable controls, $w_l'\gamma$. In this specification, l corresponds to locations. The endogeneity concern is that $\mathbb{E}[x_l \epsilon_l] \neq 0$. The endogenous variable can be written in terms of industry-location shares, where industries are indexed by n,

$$x_l = \sum_n s_{ln} g_{ln},$$

and $\sum_n s_{ln} = 1$. To connect Bartik instruments to GIV, we assume a simple factor model in g_{ln},

$$g_{ln} = g_n + \tilde{g}_{ln},$$

that is, the loadings on the common factor, g_n, are equal to one. In Bartik applications, a concern is typically that $\mathbb{E}[\tilde{g}_{ln} \epsilon_l] \neq 0$, for instance, when local economic conditions in location l are correlated with the idiosyncratic growth rate of industry n in location l.

To express the identifying assumption in Borusyak et al. (2022), they write the model at the industry level

$$\bar{y}_n = \alpha + \beta \bar{x}_n + \bar{\epsilon}_n, \tag{154}$$

where $\bar{b}_n = \frac{\sum_l s_{ln} b_l}{\sum_l s_{ln}}$, for some variable b_l. The shares, s_{ln}, are assumed to be non-stochastic, and the main identifying assumption is that $\mathbb{E}[g_n \bar{\epsilon}_n] = 0$.

[66] We are grateful to a referee for suggesting this connection.

H.5.2 Defining and comparing the Bartik and GIV instruments

The Bartik instrument is defined as

$$z_n^{Bartik} = \frac{1}{L}\sum_l g_{ln}.$$

The GIV is defined as

$$z_n^{GIV} = \sum_l \tilde{s}_{ln} g_{ln} - \frac{1}{L}\sum_l g_{ln},$$

where $\tilde{s}_{ln}$ is the location share of industry n so that $\sum_l \tilde{s}_{ln} = 1$. Hence, we have $\lim_{L\to\infty} z_n^{Bartik} = g_n$ and $\lim_{L\to\infty} z_n^{GIV} = \lim_{L\to\infty} \sum_l \tilde{s}_{ln}\tilde{g}_{ln} = \tilde{g}_{\tilde{S}n}$. In the remainder of this section, we work with the large L version of the instruments to simplify the exposition.

For Bartik instruments to be valid in (154), we need $\mathbb{E}[g_n \bar{\epsilon}_n] = 0$. For the GIV to be valid, we need $\mathbb{E}[\sum_l \tilde{s}_{ln}\tilde{g}_{ln}\bar{\epsilon}_n] = \mathbb{E}[\sum_l \tilde{s}_{ln}\tilde{g}_{ln}\frac{\sum_l s_{ln}\epsilon_l}{\sum_l s_{ln}}] = 0$. As $\mathbb{E}[\tilde{g}_{ln}\epsilon_l]$ may not be zero in cross-sectional settings, as discussed before, GIV is not the most natural instrument in those circumstances.

By the same logic, the identifying assumption of the Bartik instrument may be less appealing in settings where the identifying assumption of the GIV is more plausible. To connect the Borusyak et al. (2022) setup to the one we consider in this paper, we relabel l to $i = 1, ..., N$ and n to $t = 1, ..., T$. In addition, we set $\lambda_i = 1$, g_{ln} to y_{it}, g_n to η_t, $\tilde{g}_{ln}$ to u_{it}, which implies $y_{it} = \eta_t + u_{it}$. For simplicity, we assume that the shares do not vary across time (and thus across n in the Bartik setup). We use S_i, with $S_{it} = S_i$, to denote the relative size such that $\sum_i S_i = 1$.

We redefine the Bartik instrument and the GIV using these definitions:

$$z_t^{Bartik} = \frac{1}{N}\sum_i g_{it} = g_{Et},$$

and

$$z_t^{GIV} = \sum_i S_i g_{it} - \frac{1}{N}\sum_i g_{it}.$$

Hence, we have $\lim_{N\to\infty} z_t^{Bartik} = \eta_t$ and $\lim_{N\to\infty} z_t^{GIV} = u_{St}$. As before, we work with the large N version of the instruments to simplify the exposition. To provide a simple example where the Bartik instrument may be less appealing, we consider a trivial version of our baseline model in Section 2 (with $\phi^d = 0$)

$$y_{it} = \eta_t + u_{it},$$
$$s_t = \phi^s p_t + \epsilon_t,$$

where $p_{St} = (\eta_t + u_{St} - \epsilon_t)/\phi^s$. If we average the model in the cross-section (again using the limit when $N \to \infty$)

$$
\begin{aligned}
y_{Et} &= \eta_t, \\
s_t &= \phi^s p_t + \epsilon_t,
\end{aligned}
$$

To estimate ϕ^S, we can use two instruments. First, we can use GIV, which requires

$$\mathbb{E}\left[\epsilon_t u_{St}\right] = 0.$$

Alternatively, we can use the Bartik instrument and assume

$$\mathbb{E}\left[\epsilon_t \eta_t\right] = 0.$$

In this example, the Bartik instrument requires assumptions that are too strong in models in which η_t and ϵ_t are correlated. These are the settings that we focus on in this paper, and GIV is well suited to estimate the parameters of interest.

H.6 Complements to the general procedure

The procedure can be simplified in some cases. *When we have a long time-series.* Recall that

$$y_{St} = \sum_f \lambda_{St}^f F_t^f + u_{St}. \tag{155}$$

Hence, if all factors with λ_{St}^f possibly non-zero are observables and exogenous, we can measure the λ_{St}^f by OLS with the regression above, and get u_{St} to be the residual. This is useful when we have high-frequency data (e.g. daily financial returns), which can give an acceptably small error.[67]

We can aggregate entities into categories. For this discussion, we replace "entity" by "firm". We could aggregate the firms into $K > 1$ sub-categories (e.g. industries – or even an arbitrary categorization like "blue firms" and "red firms") — then the above works, but interpreting the partition i as "aggregate firm category i" rather than "firm i". Indeed, (110) aggregates without problem: if aggregate k is made of firm $i \in I_k$, we just define the aggregate size of category k as $S_{[k]} := \sum_{i \in I_k} S_i$, the relative weight of firm i in category k as $\omega_{[k]i} = \frac{S_i 1_{i \in I_k}}{S_{[k]}}$, and the action factor loading as value-weighted averages ($y_{[k],t} = \sum_i \omega_{[k]i,t} y_{it}$, $\alpha_{[k]}^f = \sum_i \omega_{[k]i} \alpha_i^f$). Then, the model works, using those aggregated categories. What we do need is that categories have non-trivial idiosyncratic shock (so that a "very small firms" category would not be valid, as it would have $var\left(u_{it}\right) \simeq 0$).

[67]Indeed, this time-series regressions gives an $O\left(\frac{1}{\sqrt{T}}\right)$ error, which is good enough for large T. Using the cross section, as in the basic procedure, gives an $O\left(\frac{1}{\sqrt{TN}}\right)$ error.

H.7 When only some shocks are kept in the GIV

If we truncate the residuals, i.e. use

$$z_t = \sum_i \tau \left(S_i \left(u_{it} - u_{Et} \right) \right)$$

for the hard thresholding function

$$\tau (x) = x 1_{|x| \geq b}$$

for some $b > 0$, then everything works too. Indeed, we have that $\check{u}_{it} := u_{it} - u_{Et}$ is orthogonal to u_{Et}. Let us assume that it is independent. In our basic example of Section 2.1, we still have $\mathbb{E}\left[\left(p_t - \alpha y_{St} \right) z_t \right] = 0$, so that the IV procedure of Proposition 1 still works.

Furthermore, the OLS estimates still hold. The key is that we can write:

$$u_{\Gamma t} = z_t + z_t^{<},$$

where $z_t^{<} = \sum_i \tau^{<} \left(S_i \check{u}_{it} \right)$, using $\tau^{<} (x) = x 1_{|x| < b}$, so that $z_t \perp z_t^{<}$. Hence, regressing $u_{\Gamma t}$ on this truncated z_t gives a coefficient of 1, and all the analysis goes through.

H.8 Sporadic factors

A potential issue is that of a "sporadic factor", i.e. a factor η_t that affects a few actors special ways, but is not recurrent. An example would be a one-off policy announcement by the European Central Bank that they will buy both Italian and Spanish bonds, so that the truth is not that Italy is affecting Spain or vice-versa, but rather the ECB affecting both.

One solution, besides the narrative check that we just detailed, would be to filter out days with a high "sporadicity statistic" $\mathcal{S}_t$ that we now propose. Suppose that for each date we filter out the idiosyncratic shocks $\check{u}_{it}$. For each date and actor i we form $b_{it} = \frac{\check{u}_{it}^2}{\sigma_{u_i, t-1}^2}$, where a high b_{it} is an indicator of extra activity, and $\sigma_{u_i, t-1}^2$ is a predictor of the volatility of u_{it}. We may allow that one entity has a large idiosyncratic shock, but if two (or more) do, this is suspicious, and possibly the sign of a sporadic factor. So, calling $b_{(2)t}$ the activity of the second more active actor, we form $\mathcal{S}_t = b_{(2)t}$.[68] Over the entire sample, we might remove the days with anomalously high sporadicity statistics, e.g. in the top 5% by that metric.

H.9 GIV for differentiated product demand systems

We develop the basic ideas for the logit demand model and extend these ideas to the random-coefficients logit model as in Berry et al. (1995a) in the next subsection.[69]

[68] We could also sum over the most active K entities, excluding the most active one.

[69] We thank Robin Lee, Alex MacKay, and Ariel Pakes for very helpful feedback on this section.

H.9.1 Logit demand

The utility that household h derives from product i, for $i = 0, ..., N$, is given by[70]

$$\begin{aligned} U_{hit} &= \delta_{it} + e_{hit}, \\ \delta_{it} &= -\gamma p_{it} + \beta' x_{it} + \alpha_i + \xi_{it}, \end{aligned}$$

where e_{hit} follows a Type-1 extreme-value distribution, p_{it} denotes the log price, x_{it} observable characteristics, and $\mathbb{E}\left[\xi_{it}\right] = 0$. We refer to $i = 0$ as the outside option and normalize $\delta_{0t} = 0$. This model implies that the market share s_{it} is the probability that a given household selects product i, meaning that $s_{it} = \mathbb{P}\left(U_{hit} > \max_{j \neq i} U_{hjt}\right)$, and can be expressed as

$$s_{it} = \frac{\exp(\delta_{it})}{\sum_{j=0}^{N} \exp(\delta_{jt})}.$$

Firms set prices to maximize profits and we assume that each product is produced by a single firm, which solves

$$\max_{P_{it}} Q_{it} \left(P_{it} - C_{it}\right),$$

where C_{it} equals marginal cost and $Q_{it} = s_{it} Q_t$ with Q_t the total size of the market. The firm optimally sets the price to

$$P_{it} = \left(1 - \frac{1}{\epsilon_{it}}\right)^{-1} C_{it},$$

where $\epsilon_{it} = -\frac{\partial \ln s_{it}}{\partial p_{it}}$, that is, the negative of the price elasticity of demand. The goal is to estimate $\theta = (\beta, \gamma)$.

It is convenient to rewrite the model as

$$\log\left(\frac{s_{it}}{s_{0t}}\right) = -\gamma p_{it} + \beta' x_{it} + \alpha_i + \xi_{it}.$$

To identify β, it is commonly assumed that $\mathbb{E}\left[x_{it}\xi_{it}\right] = 0$ and we maintain this assumption. However, as prices respond to demand shocks, ξ_{it}, we cannot assume $\mathbb{E}\left[p_{it}\xi_{it}\right] = 0$. There are three common approaches to create instrumental variables in the demand estimation literature. First, variables that capture variation in marginal cost, C_{it}, that is unrelated to demand shocks. Second, Berry et al. (1995a) suggest to use the average of characteristics of other firms

$$z_{it}^{BLP} = \frac{1}{N-1} \sum_{j, j \neq i} x_{jt},$$

which results in valid instruments under some assumptions (see Nevo (2000) and the references

[70] We use the log price, p_{it}, instead of the price, P_{it}, in the formulation of δ_{it} to simplify some of the expressions, but the basic logic extends to the case where δ_{it} depends on P_{it}.

therein).[71] The resulting moment is $\mathbb{E}\left[z_{it}^{BLP}\xi_{it}\right] = 0$.[72] Third, one can use panel data for the same firm that operates in different locations. Under the assumption that demand shocks are uncorrelated across locations, prices in other locations of the same firm will be valid instruments. The intuition is that prices across locations share the same marginal cost but the demand shocks are, by assumption, uncorrelated, see Nevo (2001).

GIV provides an alternative by exploiting exogenous variation in markups due to idiosyncratic demand shocks to large firms. We assume that demand shocks follow a factor model,

$$\xi_{it} = \eta_t + u_{it}, \tag{156}$$

which can be extended to allow for heterogeneous exposures, i.e. replacing η_t by $\lambda_i \eta_t = \sum_k \lambda_i^k \eta_t^k$. Also, we assume for simplicity that η_t and u_{it} are i.i.d. over time, but the logic in this section can be extended to persistent demand shocks (see also Sweeting (2013)).

We propose to use the GIV instrument as the weighted sum of idiosyncratic demand shocks of the competitors:

$$z_{it} = \sum_{j:j\neq i} \bar{s}_{j,t-1} u_{jt}, \tag{157}$$

where $\bar{s}_{j,t-1}$ is the average market share for product j up to time $t-1$. This allows us to add a moment condition

$$\mathbb{E}\left[z_{it}\xi_{it}\right] = 0, \tag{158}$$

which identifies γ. Remember that we use $\mathbb{E}\left[x_{it}\xi_{it}\right] = 0$ to identify β.

The intuition for why z_{it} is a meaningful instrument is the following: if there is a high idiosyncratic shock for Tesla cars (high u_{jt}, with j being Tesla), this leads Ford (firm i) to reduce the price of its cars (in this particular model, this is because the positive shock for Tesla cars reduces the demand for Ford, which sees its market share s_{it} fall, so that it wants to lower its price p_{it}).

Generalizing this intuition, we sum over all the demand shocks of the competitors, $z_{it} = \sum_{j:j\neq i} \bar{s}_{j,t-1} u_{jt}$, weighing them by size, i.e. market share. As in our general GIV, even a single shock u_{jt} is a valid instrument (for $j \neq i$). The size-weighted sum is simply a typically useful way to pool those idiosyncratic shocks. It is optimal in our basic GIV, and is likely to be reasonably close to optimal in this IO context. The same idea generalizes: e.g. using a weighted sum of the idiosyncratic cost shocks, rather than demand shocks, of the competitors would also be a valid GIV instrument.

A motivation for the weighting in (157) is as follows. Recall that in this simple model the demand elasticity is

$$\epsilon_{it} = \gamma(1 - s_{it}),$$

[71] For other recent advances to construct instruments, see Sweeting (2013) and MacKay and Miller (2019).

[72] If a firm offers multiple products, the average of characteristics of other products produced by the same firm can be used as well.

and also that $\frac{\partial \log s_{it}}{\partial \delta_{jt}} = -s_{jt}$, so that $\frac{\partial \log s_{it}}{\partial u_{jt}} = -s_{jt}$ (controlling for the price p_{jt}). This implies that the direct impact of all idiosyncratic demand shocks to other companies on s_{it}, and hence ϵ_{it}, is

$$\sum_{j:j\neq i} \frac{\partial \log s_{it}}{\partial u_{jt}} u_{jt} = -\sum_{j:j\neq i} s_{jt} u_{jt}. \tag{159}$$

Hence, shocks to companies with larger market shares have a larger impact.

H.9.2 Random coefficients logit as in BLP

Berry, Levinsohn and Pakes (1995a) extend the standard logit model by allowing for random variation in the preference parameters

$$\theta_h = \theta + \nu_h,$$

where $\nu_h = \left(\nu_h^\beta, \nu_h^\gamma\right)$ and $\nu_h \sim F_\nu(\nu;\Theta)$, for some vector of parameters Θ. The market share equation modifies to

$$s_{it} = \int_\nu s_{hit} dF_\nu(\nu;\Theta),$$

where

$$s_{hit} = \frac{\exp\left(\delta_{it} - \nu_h^\gamma p_{it} + \nu_h^{\beta\prime} x_{it}\right)}{\sum_{j=0}^N \exp\left(\delta_{jt} - \nu_h^\gamma p_{jt} + \nu_h^{\beta\prime} x_{jt}\right)}.$$

To estimate the model, Berry (1994) suggests to recover δ_{it} from the market shares using a contraction mapping (see Nevo (2000) for an introduction). With δ_{it} in hand, we form moment conditions as before to estimate (θ, Θ).

To construct a GIV instrument in this model, one can also use (157) as an instrument.

One can also refine it. For instance, we can recompute the total impact of idiosyncratic shocks to other firms on the demand elasticity, which is now slightly more involved. The negative of the demand elasticity, which enters into the pricing equation via the markup, is given by

$$\epsilon_{it} = \int_\nu \nu_h^\gamma \frac{s_{hit}}{s_{it}} (1 - s_{hit}) dF_\nu(\nu;\Theta).$$

An approximation of the model around $\theta_h = \theta$ yields the same weights as before, although it is feasible to numerically calculate the optimal weights by computing

$$\sum_{j,j\neq i} \frac{\partial \epsilon_{it}}{\partial u_{jt}} u_{jt}.$$

This suggests forming

$$z_{it} := \sum_{j:j\neq i} s_{j,t-1}^i u_{jt}, \tag{160}$$

where $s^i_{j,t}$ is

$$s^i_{j,t} := -\frac{\partial \log s_{it}}{\partial u_{jt}}. \tag{161}$$

Indeed, in the homogeneous elasticity case, $s^i_{j,t} = s_{jt}$. This generalization to heterogeneous elasticity allows to capture that if firms i and j tend to serve the same consumers (e.g., both sell family cars), then the $s^i_{j,t}$ will be high, and u_{jt} receives a high weight in the firm-i specific GIV z_{it}.

H.10 When the variance-covariance matrix of the u's is estimated

H.10.1 Main message

To determine V^u, we propose the following procedure. We first pick some W (e.g. the identity). Then we use the N moments to identify V^u (they come from $V^{\check{u}} = QV^uQ'$):[73]

$$\mathbb{E}\left[\check{u}^2_{ii}\right] = \left(QV^uQ'\right)_{ii}.$$

Now that we have V^u, we form $W = \left(V^u\right)^{-1}$, and use the associated $Q = Q^{X,W}$ and R, we form a new GIV $z_t := S'Q^{X,W}y_t$, and use those to identify $\dot{\phi}^d$ via (69).

We now discuss refinements of that basic theme.

H.10.2 Identification of the variance-covariance matrix

In some cases, we'll want to identify the matrix V^u. We discuss this in Section B. We add some thoughts on extensions here.

A simple sufficient condition is the following.

Assumption 5 (Restriction on the admissible variance-covariance matrix of residual u_t) *(a) The variance-covariance on u_t is diagonal. (b) The function $V \mapsto QVQ'$ from the space of diagonal matrices is injective.*

Assumption 5(a) could be relaxed in number of ways.[74] Other sufficient condition for identification might be that V^u is $k-$sparse, e.g. has at most k non-zero off-diagonal elements, for some k, e.g. $N - r^2$ (see also Zou et al. (2006)). Another is to allow for some correlation that depends on the distance between entities i and j, perhaps via Gaussian processes (Rasmussen and Williams (2005)). We conjecture that this proposition could be generalized in a number of ways, including in the large T, N domain, using material such as Bai and Ng (2006). Doing this would however take us too far afield.

[73] We assume that V^u can be characterized by N moments, e.g. that V^u is diagonal. In the more general version we would use more moments from the identity $\mathbb{E}\left[\check{u}_t\check{u}'_t\right] = QV^uQ'$, and not just its diagonal terms.

[74] However, relaxations of Assumption 5 will still need to ensure some restrictions on the space of variance-covariances allowed.

Assumption 5(b) is equivalent to saying that knowing the variance-covariance matrix of the residuals $\check{u}_t = Qu_t$ allows to get the variance of the u_i's. We have explored sufficient conditions on the X for that to hold, but they are not particularly enlightening.[75,76] We now show how to do that in the most basic (and useful) case.

H.10.3 When the u_i are uncorrelated (but heteroskedastic) and only $u_i - u_E$ is measured

We suppose that the $u_i's$ are uncorrelated, with variance $\sigma_i^2 = var\left(u_{it}\right)$. We only measure $\check{u}_{it} = u_{it} - u_{Et}$. Here's a bit of algebra to recover σ_i^2.

We define $\sigma_E^2 := \frac{1}{N}\sum_i \sigma_i^2$. We have:

$$var\left(u_E\right) = \frac{\sigma_E^2}{N}, \qquad var\left(\check{u}_{it}\right) = \sigma_i^2\left(1 - \frac{2}{N}\right) + \frac{\sigma_E^2}{N} \tag{162}$$

which implies:

$$\sigma_{\check{u},E}^2 := \frac{1}{N}\sum_i var\left(\check{u}_{it}\right) = \sigma_E^2\left(1 - \frac{1}{N}\right) \tag{163}$$

So, we can recover

$$\sigma_i^2 = \frac{var\left(\check{u}_{it}\right) - \frac{\sigma_{\check{u},E}^2}{N-1}}{1 - \frac{2}{N}} \tag{164}$$

H.10.4 A simple recovery procedure

More generally, suppose that we have $\check{u}_t = Qu_t$ for a known square matrix Q (e.g. that in (31)). Suppose that that V^u is diagonal, and we know $V^{\check{u}}$. Call D and $\check{D}$ the vectors with the diagonal elements of V^u and $V^{\check{u}}$, respectively, so that $D_i = var\left(u_i\right)$ and $\check{D}_i = var\left(\check{u}_i\right)$. Then with the matrix $H_{ij} := Q_{ij}^2$, we have:[77]

$$\check{D} = HD \tag{165}$$

[75] If we estimate V^u via $V^{\check{u}}$, we need: $\frac{1}{2}\left(N-r\right)\left(N-r+1\right) \geq N$ as the projection Q on a space of dimensions $N-r$ leaves only $\frac{1}{2}\left(N-r\right)\left(N-r+1\right)$ degrees of freedom. So in some cases with very small N one may want another procedure to estimate V^u, perhaps simply using the whole of V^y.

[76] Note that a necessary (and often sufficient) conditions is that the number of parameters to be estimated is not too big. Take the problem where we mostly are interested in estimating multiplier M, and have no extraneous observable factors. Matrix V^y gives $\frac{1}{2}N\left(N+1\right)$ parameters. We want to estimate: the number of unknowns in V^u, n_{V^u} (equal to N if we assume a diagonal matrix V^u, equal to 1 is we assume homoskedasticity), parameter M, and matrix V^η (which has $\frac{1}{2}r\left(r+1\right)$ degrees of freedom). If the factor model is parametric with X_{it} a $r-$dimensional vector, then we need: $\frac{1}{2}N\left(N+1\right) \geq 1 + \frac{1}{2}r\left(r+1\right) + n_{V^u}$.

[77] Proof: Calling e_i the vector with 1 at coordinate i and 0 elsewhere, we have $V^u = \sum_j e_j e_j' D_j$. As $V^{\check{u}} = QV^uQ'$,

$$\check{D}_i = e_i'V^{\check{u}}e_i = e_i'QV^uQ'e_i = \sum_j e_i'Qe_je_j'Q'e_iD_j = \sum_j Q_{ij}^2D_j = \sum_j H_{ij}D_j.$$

Hence, when H is invertible (which is typically true), we can recover the variance of $\sigma_{u_i}^2 = D_i$ by $D = H^{-1}\check{D}$.[78] One disadvantage of this procedure is that it does not guaranty that D_i is positive (indeed, in (164), the right-hand side is not necessarily positive).

H.10.5 An MLE-based recovery procedure

Here is another procedure, which guaranties to recover positive $\sigma_{u_i}^2$. We suppose that we know $\check{u}_t = Qu_t$ for a known matrix Q (e.g. that in (31), but not necessarily), do not know the underlying u_t. We want to recover V^u, assumed to be diagonal. The dimensions of u_t and $\check{u}_t$ are respectively n and m, with potentially $m \neq n$. So Q has dimensions $m \times n$.

We do a singular value decomposition of Q. We call its rank k. We can write:

$$Q = UDV^* \tag{166}$$

where U and V are unitary matrixes with dimensions $m \times m$ and $n \times n$ respectively, and D is rectangular diagonal with dimensions $m \times n$, and V^* is the conjugate transpose. We order the diagonal elements of D, d_i, so that the first $d_1, \ldots, d_k$ are non-zero, while the remaining d_i's are zero. We set $\pi := \begin{pmatrix} I_{k\times k} & 0_{k\times(m-k)} \end{pmatrix}$ and set $B := \pi U^*$, both with dimension $k \times m$. We define

$$\bar{u}_t := B\check{u}_t \tag{167}$$

which is a k dimensional vector that gathers the useful k degrees of freedom in $\check{u}_t$. While $V^{\check{u}}$ was singular, typically $V^{\bar{u}}$ has full rank. So, $\bar{u}_t = Mu_t$ with $M = BQ$.

To recap, we have $\bar{u}_t = Mu_t$ with M a $k \times n$ matrix of rank $k \leq n$. So, $V^{\bar{u}} = MV^uM'$. We call $W^{\bar{u}} = (V^{\bar{u}})^{-1}$ the theoretical inverse variance, $V^{\bar{u},e} = \frac{1}{T}\sum_{t=1}^T \bar{u}_t'\bar{u}_t$ the variance of $\bar{u}_t$. We assumed that $V^u = Diag\left(\sigma_{u_1}^2, \ldots, \sigma_{u_n}^2\right)$. We assume that

$$\frac{1}{2}k\left(k+1\right) \geq n$$

The left-hand side is the number of degrees of freedom in $V^{\bar{u}}$: it should be higher than the number of parameters n we want to estimate for V^u.

The log likelihood l is (we omit some constants in 2π, and use the notation $|W^{\bar{u}}|$ for the deter-

[78] Numerically, within finite samples, we can get negative $\sigma_{u_i}^2$ (see (164)). So, one can imagine variants that guaranty positivity, e.g. adding a winsorization step, $D_i := \max\left(D_i, \xi \cdot \text{median}\left(D\right)\right)$ for a low ξ such as $\xi = 0.5$. Another procedure is to do

$$\min_D \left\|\check{D} - HD\right\|^2 \text{subject to } \min D_i \geq \xi \cdot \text{median}(D)$$

or another constraint, e.g. $D_i \geq \xi \check{D}_i$. Yet another variant is to minimize $\left\|V^{\check{u}} - QDiag\left(D_i\right)Q'\right\|^2$, subject to the same constraints.

minant of $W^{\bar{u}}$),

$$\begin{aligned}\ell &:= \frac{2L}{T} = -\mathbb{E}_T\left[\operatorname{tr}\left(\bar{u}_t' W^{\bar{u}} \bar{u}_t\right)\right] + \ln\left|W^{\bar{u}}\right| = -\operatorname{tr}\left(\mathbb{E}_T\left[\bar{u}_t \bar{u}_t'\right] W^{\bar{u}}\right) + \ln\left|W^{\bar{u}}\right| \\ \ell &= -\operatorname{tr}\left(V^{\bar{u},e} W^{\bar{u}}\right) + \ln\left|W^{\bar{u}}\right| \qquad (168)\end{aligned}$$

We recover V^u by numerically maximizing ℓ over the $\sigma_{u_i}^2 > 0$.

H.11 Singular Factor Analysis

We propose a tool useful in the advanced parts of this project: A way to do factor analysis with singular matrices.

Suppose that we have an underlying factor model:

$$Y_t^* = \Lambda^* \eta_t + u_t,$$

where Y_t^* and u_t have dimension N, and η_t dimension r, but we only observe:

$$Y_t = G Y_t^*$$

for a known matrix G such that $G^2 = G$ (e.g. G could be a Q matrix as in (31)). But potentially G has less than full rank, so that V^Y is singular. We present a procedure to estimate V^u (assuming that it has some structure, here that it is diagonal), and also recover $\Lambda := G\Lambda^*$ and a proxy for the η_t.

Projecting Y_t into a lower-dimensional y_t We do an eigendecomposition of G. Calling K the rank of G, we can write:

$$G = A^{-1} D A \qquad (169)$$

where $D = \begin{pmatrix} I_K & 0 \\ 0 & 0 \end{pmatrix}$, and A^{-1} is the matrix whose columns are the corresponding eigenvectors of G.[79,80] We introduce $\pi := \begin{pmatrix} I_K & 0 \end{pmatrix}$ and set $B := \pi A$, both with dimension $K \times N$. We have[81]

$$D = \pi'\pi, \qquad \pi D = \pi, \qquad B = \pi A, \qquad BG = B, \qquad A^{-1}\pi' B = G \tag{170}$$

We call

$$y_t := BY_t \tag{171}$$

which is a K dimensional vector that gathers the useful K degrees of freedom in Y_t. While V^Y was singular, typically V^y has full rank. With $\lambda := B\Lambda$ (dimension: $K \times r$) and $v_t := BGu_t = Bu_t$ (dimension: $K \times 1$) we have:

$$y_t = \lambda \eta_t + v_t \tag{172}$$

Doing PCA on y_t We have

$$V^y = \lambda V^\eta \lambda' + V^v,$$

The first PCs in a PCA of V^y will not be λ, unless V^v is proportional to the identity matrix.[82] Ideally, we'd like then to estimate the PCA on $Ly_t = L\lambda\eta_t + Lv_t$, for $L = k\left(V^v\right)^{-1/2}$ (with k a constant) because then the covariance of residuals Lv_t will be proportional to identity, and the PCA will correctly recover $L\lambda$. But we do not know V^v: we need to estimate it.

This motivates the following algorithm. For an invertible matrix V of dimensions $K \times K$, call[83]

$$J(V) := V^{-1/2}\sqrt{\frac{\left|\operatorname{tr}(V)\right|}{K}} \tag{173}$$

[79] We order the columns of A^{-1} with first the eigenvectors with eigenvalue 1, then those with eigenvalue 0.

[80] We found that Matlab could get lost, and return complex eigenvectors, even though $G^2 = G$ ensures that the eigenvalues are 0 and 1, and all eigenvectors are real. One fix to this numerical implementation issue is the following. We observe that in practice the $G = Q^{\lambda,W}$ come from (31). Calling $H = W^{-1/2}$, then $\tilde{G} = H^{-1}GH$ is a symmetric matrix, so that Matlab recognizes that the eigenvectors should all be real (to be numerically safe, we entered it as $\left(\tilde{G} + \tilde{G}'\right)/2$). Then, if e is an eigenvector of $\tilde{G}$ with eigenvalue k, He is an eigenvector of G with the same eigenvalue k. This way, we recover real eigenvectors of G.

[81] Indeed, $BG = (\pi A)\left(A^{-1}DA\right) = \pi DA = \pi A = B$, and

$$A^{-1}\pi' B = A^{-1}\pi'\pi AG = A^{-1}DAG = GG = G.$$

[82] Indeed, if V^v is proportional to the identity matrix, then $V^y\lambda$ is proportional to λ, so that the column vectors of λ are eigenvectors of V^y (and PCA, which extracts the eigenvectors, will successfully recover λ). But this is not the case if V^v is not proportional to the identity matrix.

[83] Given a positive definite matrix V, and α a scalar, V^α is defined as follows. Do an eigendecomposition $V = P\Delta P^{-1}$, where the columns of P are the eigenvectors of V, and $\Delta = Diag\left(\Delta_i\right)$ is the diagonal matrix with V's eigenvalues. Then, we define $V^\alpha = P Diag\left(\Delta_i^\alpha\right) P^{-1}$. We apply this to $\alpha = -\frac{1}{2}$.

The trace (tr) factor is there so that the transformation $V \mapsto J(V)\,VJ(V)$ keeps the "size" of V (as measured by its trace) fixed:

$$\mathrm{tr}\left(J(V)\,VJ(V)\right) = \mathrm{tr}\,(V) \tag{174}$$

Also, if V is proportional to the identity matrix I, then $J(V) = I$.

So we can envisage the following scheme: We start with $L_0 = I_K$.

1. We do a PCA on $y_{nt} := L_n y_t$, so do $y_{nt} = L_n \lambda \check{\eta}_t + \dot{v}_{nt}$ and get the residuals $\dot{v}_{nt}$.
2. We define $\check{u}_t := A^{-1}\pi' L_n^{-1} \dot{v}_{nt}$. We note that $\dot{v}_{nt} = Q^{L_n\lambda, I_K} L_n v_t$ and $v_t = Bu_t$, so that $\check{u}_{nt} = q_n u_t$ with matrix $q_n := A^{-1}\pi' L_n^{-1} Q^{L_n\lambda, I_K} L_n B$. As the rank of λ is r, matrix q_n as dimension $N \times N$ but only rank $K - r$. From $V^{\check{u}_n}$ we obtain V_n^u, using the procedure in Section H.10.4, observing that $\check{u}_{nt} = q_n u_t$.
   - (a) We can use the procedure in Section H.10.4, but it does not guaranty that $\sigma_{u_i}^2$ is always positive. (Then, some winsorization can impose $\sigma_{u_i}^2 > 0$).
   - (b) Alternatively, we can use the MLE procedure of Section H.10.5 which does guaranty that $\sigma_{u_i}^2$ is positive.[84]
3. Set $L_{n+1} := J(V_n^v)$, with $V_n^v := BV_n^u B'$.
4. Iterate steps 1-4 until convergence (e.g. $\left\|L_{n+1}^{-1} L_n - I\right\| < 0.01$). As a check, we note that at convergence, $V^{L_n v_t} = L_n V_n^v L_n$ should be close to proportional to the identity matrix ($\left\|\frac{L_n V_n^v L_n}{\mathrm{tr}(L_n V_n^v L_n)/K} - I\right\|$ should be small).

The procedure returns $V^{u,e} = V_n^u$. It also returns the factors $\check{\eta}_t$ (up to a rotation, as usual). We also obtain an estimator of Λ:

$$\Lambda^e = A^{-1}\pi' L_n^{-1}\left(L_n \lambda\right) \tag{175}$$

where $L_n\lambda$ is estimated from the PCA in Step 1.[85]

H.12 Full recovery when different factors have different "size" weights

In the basic model, we can identify α^f, $M = \frac{1}{1-\sum_f \lambda^f \alpha^f}$, but not λ^f.

We give some conditions under which we can actually also identify the λ^f (in addition to α^f and M). We show here that this is the case if we assume that the size S^f differs across all factors f, and this knowledge is given to us (from a model).

[84] We can alternatively use the same procedure to $\dot{v}_{nt} = Q^{L_n\lambda, I_K} L_n B u_t$ as the observed vector, rather than going through the higher-dimensional $\check{u}_t$ as the measured vector.

[85] Indeed, at convergence $\Lambda^e = A^{-1}\pi'\lambda = A^{-1}\pi' B\Lambda = G\Lambda = GG\Lambda^* = G\Lambda^* = \Lambda$, by (170).

Here we take the basic set up as in Section G.1, in the simplified case where $\lambda_i^f = \lambda^f$ for all "endogenous" factors, i.e. for the factors f such that $\alpha^f \neq 0$, the other exogenous factors η all have an impact of 1:

$$\begin{aligned} y_{it} &= u_{it} + \sum_f \lambda^f F_t^f + \eta_t^y, && (176) \\ F_t^f &= \alpha^f y_{S^f,t} + \eta_t^f. && (177) \end{aligned}$$

This implies

$$y_t = u_t + \iota \sum_f \lambda^f F_t^f + \iota \eta_t^y = u_t + \iota \sum_f \lambda^f \left(\eta_t^f + \alpha^f S^{f'} y_t \right) + \iota \eta_t^y.$$

With "ε^k" denoting some combination of the various η's, and as usual $M = \frac{1}{1 - \sum_f \lambda^f \alpha^f}$,

$$\begin{aligned} y_t &= \left(I - \iota \sum_f \lambda^f \alpha^f S^{f'} \right)^{-1} \left(u_t + \iota \varepsilon_t^1 \right) \\ &= \left(I + M \iota \sum_f \lambda^f \alpha^f S^{f'} \right) \left(u_t + \iota \varepsilon_t^1 \right) \\ y_t &= u_t + M \iota \sum_f \lambda^f \alpha^f u_{S^f,t} + \iota \varepsilon_t^y, && (178) \end{aligned}$$

i.e., since $F_t^f = \eta_t^f + \alpha^f y_{S^f,t}$ this gives:

$$F_t^f = \alpha^f \left(u_{S^f,t} + M \sum_g \lambda^g \alpha^g u_{S^g,t} \right) + \varepsilon_t^f. \tag{179}$$

Hence, suppose that we extracted the $\check{u}_{it} = u_{it} - u_{Et}$ (following our usual procedure). Then, we form

$$z_{\Gamma^f t} := S^{f'} \check{u}_t = u_{S^f t} - u_{Et}. \tag{180}$$

Then, regressing F_t^f on the various $z_{\Gamma^g t}$

$$F_t^f = \sum_g b_g^f z_{\Gamma^g t} + \varepsilon_t^{f,1} \tag{181}$$

(for ε^{f1} some residual noise) yields a regression coefficient:

$$b_g^f = \alpha^f \left(1_{f=g} + M \lambda^g \alpha^g \right). \tag{182}$$

This allows to recover everything, and with several overidentifying restrictions. Indeed,

$$b^f := \sum_g b_g^f = \alpha^f \left(1 + M \sum_g \lambda^g \alpha^g\right) = \alpha^f M,$$

which identifies $\alpha^f M$. Next, for $f \neq g$,

$$\frac{b_g^f}{b^f} = \lambda^g \alpha^g,$$

which gives $\lambda^g \alpha^g$ (and should be equal for all f), thus M. Hence, we obtained $\alpha^f M$, M and $\lambda^g \alpha^g$ — therefore all quantities: α^f, λ^f, M.

H.13 When aggregate shocks are made of idiosyncratic shocks

GIVs extend to economies where aggregate shocks η_t are themselves made of idiosyncratic shocks u_{it}. We summarize the situation here.

Take the basic supply and demand model of Section 2.1. We achieved identification provided that $u_{\Gamma t} \perp \varepsilon_t$; we did not need $u_{\Gamma t} \perp \eta_t$, so aggregate demand shocks can be influenced by idiosyncratic shocks, but not aggregate supply shocks. If aggregate supply shocks are affected by idiosyncratic shocks, the elementary strategy does not work, but a variant does work, with a slightly different identification assumption. We suppose disaggregated supply and demand data (for the commodity in question, e.g. oil) is available, at least for large countries. We model country i's supply and demand with the following factor model:

$$y_{it}^k = \phi^k p_t + \lambda_i^k \eta_t^k + u_{it}^k, \tag{183}$$

where $k = s, d$ indicates supply or demand, respectively. We allow $\mathbb{E}\left[u_{it}^s u_{it}^d\right]$ to be nonzero: for instance, if the US has a positive "fracking shock" that affects both supply and demand, it will be captured by a positive u_{it}^s and u_{it}^d for $i =$ USA. This is a concrete case in which supply and demand shocks are correlated: this happens via the correlations in country-level shocks. At the same time, we impose that the u_{it}^k are uncorrelated with the aggregate shocks $\eta_t^{k'}$ for $k, k' \in \{s, d\}$. Then, Section H.14 shows how to identify the elasticities of supply and demand.

One can also consider an economy as a network (Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019)). Under some assumptions, one can obviate the network structure, for instance via aggregation theorems such as Hulten's theorem. This is developed in Section H.15. It shows that we can identify important multipliers even if we have only crude proxies for the primitive shocks such as TFP. The GIV for a general network is a rich topic, potentially for another paper – Section H.16 lays out some of the basics.

In conclusion, one can often handle cases where aggregate shocks are made of idiosyncratic shocks: then, some more disaggregated data and economic reasoning allows to use a GIV to estimate macro parameters of interest.

H.14 When we have disaggregated data for both the demand and the supply side

To estimate supply and demand elasticities, it is enough to have idiosyncratic shocks to one side of the market — demand in our basic example (Proposition 4). We complete our examination of supply and demand, with disaggregated data for both the demand and the supply side.

We posit that demand and supply disturbances follow:

$$y_{it}^k = \phi^k p_t + \lambda_i^k \eta_t^k + u_{it}^k, \tag{184}$$

for type $k = s, d$ for supply and demand. Total quantity demanded or supplied in side k of the market is (as a disturbance from the average), $y_{S^k t}^k := \sum_i S_i^k y_{it}^k$, where S_i^d (resp. S_i^s) is the average fraction of demand (resp. supply) accounted by country i) The price p_t adjusts so that supply equals demand, $y_{S^s t}^s = y_{S^d t}^d$, i.e.

$$p_t = \frac{u_{S^d}^d - u_{S^s}^s + \lambda_{S^d}^d \eta_t^d - \lambda_{S^s}^s \eta_t^s}{\phi_{S^s}^s - \phi_{S^d}^d} \tag{185}$$

So, the aggregate supply that was $s_t = \phi^s p_t + \varepsilon_t$ (see (2)) in the aggregated model is now

$$s_t = y_{S^s t}^s = \phi^d p_t + \lambda_{S^s}^s \eta_t^s + u_{S^s t}^s$$

so that the supply shock is

$$\varepsilon_t = \lambda_{S^s}^s \eta_t^s + u_{S^s t}^s.$$

We allow $\mathbb{E}\left[u_{it}^s u_{it}^d\right]$ to be nonzero: for instance, if the US has a "fracking shock" that affects both supply and demand, it will be captured by both u_{it}^s and u_{it}^d for $i =$ USA. Then, the initial exclusion restriction $\mathbb{E}\left[u_{it}\varepsilon_t\right] = 0$ (see (5)) fails. A fracking shock in the US both increases idiosyncratic US demand (as the US is richer) and also world supplies (as the US supplies more oil via its fracking technology).

But the situation is not so bleak. We make the following assumption

$$\mathbb{E}\left[u_{it}^k \eta_t^{k'}\right] = 0 \text{ for all } k, k' \in \{s, d\}. \tag{186}$$

For instance, when $k = d$ and $k' = s$, (186) means that idiosyncratic demand shocks are uncorrelated with the aggregate supply shocks η_t — once we control for idiosyncratic supply shocks (i.e. they may be correlated with ε_t but not with η_t^s). For simplicity, we discuss the homoskedastic case (where the $\left(u_{it}^{d},u_{it}^{s}\right)$ are i.i.d. across i, t).

Then, we can still identify the elasticity of supply and demand. Indeed, we can form two GIVs, based on supply and demand respectively:

$$z_t^k := \Gamma^{k\prime} y_t^k = u_{\Gamma^k t}^k, \tag{187}$$

for $k = s, d$ (with $\Gamma^k = Q^{\lambda^k \prime} S^k$ in the general case and $\Gamma^k = S^k - E^k$ in the simple case $\lambda^k = \iota$, as in Proposition 3).

Then we have

$$\mathbb{E}\left[z_t^k \eta_t^{k'}\right] = 0 \text{ for all } k, k' \in \{s, d\}.$$

and one can easily see (as in the main paper) that the following identification moments hold, for $k, k' \in \{s, d\}$

$$\mathbb{E}\left[\left(y_{Et}^k - \phi^k p_t\right) z_t^{k'}\right] = 0. \tag{188}$$

So, we can estimate the demand and supply elasticities.[86][87].

Proposition 14 (Identification with disaggregated supply and demand data). *Suppose that we have disaggregated supply and demand data following* (184). *Suppose that the shock are idiosyncratic in the sense of* (186), *and that we have i.i.d.* $\left(u_{it}^{d},u_{it}^{s}\right)$ *across* i, t. *Then, the GIVs* z_t^d, z_t^s *in* (187) *identify* ϕ^d *and* ϕ^s, *via moments* (188).

In summary, in that example, the GIV fails if aggregate shocks (ε_t) are importantly made of idiosyncratic shocks. However, in the same example, having more disaggregated data (on both the demand and supply side), together with a slightly different exclusion restriction, allow estimation of both elasticities by GIV.

H.15 Identification of the TFP to GDP multiplier in a production network economy

Suppose a two-period model with a production network, as in Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019). There are both idiosyncratic TFP shocks $\hat{\Lambda}_{it}$ and a government reform that creates correlated shocks η_t to TFP and change in labor supply $\hat{L}_t$. Utility is $C_t - e^{\eta_t^L} L_t^{1+1/\phi}$, so that ϕ is the Frisch elasticity of labor supply. We call Λ_t total TFP, which depends on the industry TFPs Λ_{it}. So, as $C_t = \Lambda_t L_t$,

[86] The optimal instrument is $z_t = z_t^d - z_t^s$, as this is the most correlated with the price (185) (this generalizes the reasoning of Proposition 2).

[87] One could imagine variants. For instance, if we assume only that $\mathbb{E}\left[z_t^{\ell} \eta_t^k\right] = 0$ for a given (k, ℓ), we can identify ϕ^k via $\mathbb{E}\left[\left(y_{Et}^k - \phi^k p_t\right) z_t^{\ell}\right] = 0$.

labor supply is $\hat{L}_t = \phi\left(\hat{\Lambda}_t - \eta_t^L\right)$,[88] and GDP is $\hat{Y}_t = \hat{L}_t + \hat{\Lambda}_t$, i.e.

$$\hat{Y}_t = m\hat{\Lambda}_t - \phi\eta_t^L, \qquad m = 1 + \phi \tag{189}$$

We seek to find the "GDP multiplier" $m = 1 + \phi$, so that a TFP increase of 1 percent translates into a GDP increase of m percent.[89]

This is potentially a complicated problem, as for instance, in the Long and Plosser (1983) case with input-output matrix A, output changes are $\hat{Y}_t = (I - A)^{-1}\hat{\Lambda}_t + \hat{L}_t$, so that output changes are correlated in complicated ways. However, one can sidestep using this disaggregated production data. We assume that TFP change in industry i is:

$$\hat{\Lambda}_{it} = \lambda_i \eta_t^\Lambda + u_{it}. \tag{190}$$

In the neoclassical equilibrium, TFP follows Hulten's theorem, so is $\hat{\Lambda}_t = \sum_i s_i \hat{\Lambda}_{it}$ where s_i is the Domar weight (sales of industry i over GDP).

We can identify the multiplier m if we have disaggregated TFP data. In the simplest case, we assume that industry-level productivities are available, and we get the residuals u_{it}^e. Then, we can identify the multiplier m by GIV.

We can identify the multiplier m if we have even crude proxies for disaggregated TFP. The same procedure works (with less efficiency) if our data is made of proxies for productivity growth $\tilde{\hat{\Lambda}}_{it}$ (where the tilde indicates that we deal with a proxy). An example could be growth of sales per employee, or even the growth rate of sales. We assume a factor model

$$\tilde{\hat{\Lambda}}_{it} = \tilde{\lambda}_i \tilde{\eta}_t^\Lambda + \tilde{u}_{it}. \tag{191}$$

The proxy is of better quality when the proxy's idiosyncratic shock $\tilde{u}_{it}$ has a high correlation with the true idiosyncratic shock u_{it}. Then, we extract the $\tilde{u}_{it}^e$ from a factor model, form $z_t = \tilde{u}_{St}^e - \tilde{u}_{Et}^e$ (with $S_i = \frac{s_i}{\sum_j s_j}$), and use the moment $\mathbb{E}\left[\left(\hat{Y}_t - m\hat{\Lambda}_t\right) z_t\right] = 0$, which identifies the TFP to GDP multiplier m.

Using more general models (e.g. taking into account imperfections as in Baqaee and Farhi (2020)) would be very interesting, but would be a new paper by itself. Indeed, even in that case z_t is likely to be a useful instrument, even though it won't be the optimal one. In any case, those examples show how GIV, with some economic reasoning, translate to more complex economies where aggregate shocks can be made of idiosyncratic shocks.

---

[88] The problem is $\max_{L_t} \Lambda_t L_t - e^{\eta_t^L} L_t^{1+1/\phi}$, which leads to $\left(1 + \frac{1}{\phi}\right) L_t^{1/\phi} = \Lambda_t e^{-\eta_t^L}$, hence the announced expression.

[89] If more than one factor changes, m has the broader interpretation of a multiplier between TFP and GDP.

H.16 When the influence matrix is not proportional to size

H.16.1 Position of the problem

Suppose a model

$$y_{it} = \gamma \sum_j G_{ij} y_{jt} + \lambda_i \eta_t + u_{it}, \tag{192}$$

i.e.

$$y_t = \gamma G y_t + \Lambda \eta_t + u_t, \tag{193}$$

with a given "influence" matrix G. For instance, if we have an "industrial similarity" matrix H with entries H_{ij} (for instance $H_{ij} = 1$ iff i and j are in the same industry, and 0 otherwise) we might set

$$G_{ij} = \frac{H_{ij} S_j}{\sum_k H_{ik} S_k}.$$

In our basic setup $G = \iota S'$. We'd like to identify γ.

H.16.2 A simple approach

We study the model (193), where the factor loading Λ (an $N \times r$ matrix) is not necessarily equal to ι (but we keep imposing that the Λ spans ι, i.e. there is a q such that $\iota = \Lambda q$). As before, η_t is a low-dimensional vector of factors.

First, we suppose that we have a first estimate of γ, which we call γ^e. We will later iterate on it. Then, we form:

$$\tilde{y}_t (\gamma^e) \coloneqq (I - \gamma^e G) \, y_t. \tag{194}$$

If $\gamma^e = \gamma$, then $\tilde{y}_t (\gamma) = \Lambda \eta_t + u_t$. Hence, we run a factor analysis on $\tilde{y}_t (\gamma^e)$, which recovers Λ and $W^u = (V^u)^{-1}$. We introduce Q as in (31) so that $Q\Lambda = 0$ and set

$$\check{u}^e \coloneqq Q \tilde{y}_t (\gamma^e),$$

so that at $\gamma^e = \gamma$, $\check{u}^e_t = Q u_t$. We observe that $G y_t = G (I - \gamma G)^{-1} u_t + B \eta_t$ for some B. This suggests the following procedure.

We define the GIV z_t as a vector (with dimension N):

$$z_t \coloneqq G (1 - \gamma^e G)^{-1} \check{u}^e_t = G (1 - \gamma^e G)^{-1} Q (I - \gamma^e G) \, y_t. \tag{195}$$

Indeed that z_t will imitate the movements of the idiosyncratic shocks on y_t.

Our key moment is:[90]

$$\mathbb{E}\left[(y_t - \gamma G y_t)' W^u z_t\right] = \delta, \tag{197}$$

where δ is a discrepancy term

$$\delta := \operatorname{tr}\left(QG(1-\gamma G)^{-1}\right). \tag{198}$$

This yields an estimate of γ.[91]

The discrepancy term δ is often 0. For instance, in our basic example, $G = \iota S'$ and $Q\iota = 0$, so we have $\delta = 0$: the discrepancy was 0. Hence, (197) generalizes our basic GIV. Likewise, take the case of a block-diagonal $G_{ij} = S_j^{(k)}$ if i and j belong to industry k, and $G_{ij} = 0$ otherwise, where $S_j^{(k)}$ is the relative size of firm j in industry k (so $\sum_{j\in k} S_j^{(k)} = 1$). Also, assume that the vector of characteristics have industry dummies. Then, $QG = 0$, and again $\delta = 0$.

H.17 Identification of social interactions and the reflection problem

Superficially, there seems to be a contradiction between Section 4.7's finding that we do achieve identification, and Manski (1993)'s Proposition 2 and Bramoullé et al. (2009)'s Proposition 1, which seem also to state the impossibility of identification. Bramoullé et al. (2009) analyze social interactions of the type:

$$\boldsymbol{y}_t = \beta \boldsymbol{G} \boldsymbol{y}_t + \gamma \boldsymbol{x}_t + \delta \boldsymbol{G} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{199}$$

with $\mathbb{E}[\boldsymbol{\varepsilon}_t | \boldsymbol{x}_t] = 0$. In their main result, they conclude that if the matrices I, G, G^2 are not linearly independent, then the system is not identified. However, in our setup $G = \iota S'$ (where ι is a vector of 1's) so that $G^2 = G$ and we satisfy Bramoullé et al. (2009)'s condition that seems to guarantee the impossibility of identification. However, we can identify the parameters, as we saw in Section 4.7. How do we solve that seeming contradiction?

The short answer is that Manski (1993) and Bramoullé et al. (2009) do not consider anything like a GIV, as they immediately reason on averages based on observables, eschewing any exploration of the noise. In contrast, GIVs are all about exploring some structure in the noise — the idiosyncratic shocks of large entities. For instance Manski (1993) considers something akin to:

$$\mathbb{E}[\boldsymbol{y}_t | \boldsymbol{x}_t] = \beta \boldsymbol{G} \mathbb{E}[\boldsymbol{y}_t | \boldsymbol{x}_t] + \gamma \boldsymbol{x}_t + \delta \boldsymbol{G} \boldsymbol{x}_t, \tag{200}$$

[90] Here is the proof. At the right estimator $\gamma = \gamma^e$,

$$z_t = G(1-\gamma G)^{-1} Q (I - \gamma G) y_t = G(1-\gamma G)^{-1} Q u_t = H u_t, \tag{196}$$

for $H = G(1-\gamma G)^{-1} Q$. We also have $y_t - \gamma G y_t = \Lambda \eta_t + u_t$. This implies that

$$\begin{aligned}
\mathbb{E}\left[(y_t - \gamma G y_t)' W^u z_t\right] &= \mathbb{E}\left[(\Lambda \eta_t + u_t)' W^u H u_t\right] = \mathbb{E}\left[\operatorname{tr}(u_t' W^u H u_t)\right] = \mathbb{E}\left[\operatorname{tr}(W^u H u_t u_t')\right], \\
&= \operatorname{tr}(W^u H V^u) = \operatorname{tr}(H V^u W^u) = \operatorname{tr}(H).
\end{aligned}$$

[91] We have a fixed point: an initial γ^e gives an estimate of γ; that's then the new estimate γ^e, and we re-iterate the process, until convergence.

where all the noise has been averaged out.

Indeed, we do impose some structure, namely:

$$\varepsilon_{it} = \eta_t + u_{it}, \qquad u_{it} \text{ i.i.d., orthogonal to } \eta_t, \tag{201}$$

We could generalize to richer factor models, like in the body of the paper.

Second, we can generalize to the case where $\boldsymbol{G}^2 = \boldsymbol{G}$ (the case where G^2 is a linear combination of G and I is similar[92]), which seems to lead to the impossibility of identification in Bramoullé et al. (2009). This is formalized here.

Proposition 15 (Identification achieved in the *Bramoullé et al. (2009)* setup). *Suppose that $G^2 = G$, which is satisfied in our basic setup, but leads to the impossibility of identification in the Bramoullé et al. (2009) setup without further assumptions. Suppose also the "simple noise structure" assumption* (201). *Suppose also the existence of two n-dimensional vectors S and Γ satisfying*

$$G'S = S, \qquad G'\Gamma = 0, \qquad \iota'S \neq 0, \qquad \Gamma'S \neq 0. \tag{202}$$

Then GIV is possible in that setup, i.e. with the GIV $z_t = \Gamma' y_t$, we can identify the coefficients (β, γ, δ).

In our basic setup, we had S_i the relative sizes, and $G = \iota S'$, $\Gamma = S - \frac{\iota}{N}$. Hence (202) *is an abstract generalization of our concrete conditions.*

Hence, in many situations of interest we can be quite confident that condition (202) is satisfied.[93]

There could be another way to analyze social influence with a matrix of influence, like in Section H.16.

In conclusion: our GIV approach gives some renewed hope for identification in the context of social influence and reflection problems. Indeed, it provides a way to achieve identification where it seemed impossible. Informally, this is by exploiting the idiosyncratic noise of "large players". Formally, and less intuitively, it is by exploiting a little bit of structure in the noise (so that there is a low-dimensional common noise). Future research might profitably firm up the exact necessary and sufficient conditions for this.

Proof of Proposition 15 The identification goes as follows. By rescaling S, we impose $\iota'S = 1$. Define $E := S - \Gamma$ (which is $\frac{1}{N}\iota$ in our framework), and form

$$y_{Et} = E'y_t, \qquad y_{St} := S'y_t,$$

[92]It can be reduced to that case by rescaling $H = b_0 + b_1 G$ with the right coefficient, with $H^2 = H$.

[93]As $G^2 = G$, one can always find vectors Γ, S satisfying the first 3 conditions (provided n is big enough and G is not the identity nor 0), and the last one is rather "generically" easy to satisfy.

which are our generalized "equal weighted" and "value weighted" averages – for more abstract setting. Then, premultiplying (199) by Γ' gives:

$$z_t := \Gamma' y_t = \delta x_{\Gamma t} + u_{\Gamma t}.$$

Hence, estimating this equation by OLS we can obtain δ, $u_{\Gamma t}$, and $var\left(u_{\Gamma t}\right)$, so that we obtain also σ_u^2. Next,

$$y_E = \beta y_S + \gamma x_E + \delta x_S + \eta + u_E,$$

so that

$$\mathbb{E}\left[\left(y_{Et} - \beta y_{St} - \gamma x_{Et} - \delta x_{St}\right)\left(z_t, x_{St}\right)\right] = \left(\mathbb{E} u_{Et} u_{\Gamma t}, 0\right). \tag{203}$$

The right-hand side is known, as $\mathbb{E} u_{Et} u_{\Gamma t} = E'\Gamma \sigma_u^2$. So, we have two unknowns β, γ and two equations: we can solve the system. The condition $\Gamma' S \neq 0$ ensures that $\mathbb{E}\left[u_{St} u_{\Gamma t}\right] \neq 0$. $\square$

H.18 Identification of the elasticity of substitution between capital and labor / Elasticity of demand in partially segmented labor markets

Here we show how GIVs can estimate the elasticity of substitution between capital and labor; and how to estimate the elasticity of demand in partially segmented markets. The first problem uses the second one.

As a motivation, imagine that industry i has the CES production function[94]

$$Q_{it} = B_{it}\left(K_{it}^{\frac{\phi_i - 1}{\phi_i}} + A_{it}^{\frac{1}{\phi_i}} L_{it}^{\frac{\phi_i - 1}{\phi_i}}\right)^{\frac{\phi_i}{\phi_i - 1}} \tag{204}$$

The first order condition of the problem $\max_{K_{it}, L_{it}} Q_{it} - R_t K_t - W_{it} L_{it}$ is $A_{it}^{\frac{1}{\phi_i}}\left(\frac{L_{it}}{K_{it}}\right)^{-\frac{1}{\phi_i}} = \frac{W_{it}}{R_t}$,i.e. a demanded labor / capital ratio:

$$\frac{L_{it}}{K_{it}} = A_{it}\left(\frac{W_{it}}{R_t}\right)^{-\phi_i} \tag{205}$$

We'd like to estimate the elasticity of substitution ϕ_i between capital and labor. This is the wage elasticity of demand. GIVs allow to estimate that, as we shall see.

Let us use our general notations, and define $y_{it}^d = \ln L_{it}$, $p_{it} = \ln W_{it}$, $C_{it} = \ln K_{it}$, and $\phi_i^d = -\phi_i$ (as this is the elasticity of labor demand). Then, we can write (205) as:

$$y_{it}^d = \phi_i^d p_{it} + C_{it} + \lambda_i^d \eta_t + u_{it}^d \tag{206}$$

where C_{it} is a control, and as usual vector η_t is a common shock, and u_{it}^d is a demand shock (those in turn come from the productivity A_{it}). For notational simplicity we will drop C_{it}, but this is not

[94] We thank Julieta Caunedo for prompting us to think about this identification problem.

important.

Now, log labor supply is modeled as:

$$y_{it}^s = \phi_i^s p_{it} - \psi_i p_{St} + \lambda_i^s \eta_t + u_{it}^s \tag{207}$$

It is increasing in wage p_{it} in industry i, and decreasing in the wage in the other industries (p_{St}). One could imagine replacing $\psi_i p_{St}$ by a different average for each industry, and we will examine that in an extension. But for now we keep the simple structure.

As supply equals demand in each market ($y_{it}^d = y_{it}^s$), we obtain the price of labor in each market i:

$$p_{it} = \frac{\psi_i p_{St} + u_{it}^d - u_{it}^s + \left(\lambda_i^d - \lambda_i^s\right)\eta_t}{\phi_i^s - \phi_i^d} \tag{208}$$

i.e.

$$p_{it} = \gamma_i p_{St} + v_{it} + \lambda_i^p \eta_{it} \tag{209}$$

when we define $\gamma_i = \frac{\psi_i}{\phi_i^s - \phi_i^d}$, $v_{it} = \frac{u_{it}^d - u_{it}^s}{\phi_i^s - \phi_i^d}$, $\lambda_i^p = \frac{\lambda_i^d - \lambda_i^s}{\phi_i^s - \phi_i^d}$.

Problem (209) is a standard GIV. In the general case, we can estimate γ_i as in Section (H.4).[95]

So, we obtain γ_i and v_{it}^e (the proxy for v_{it}) in (209). We also form $z_{it} = S'\left(v_t^e - a^i v_{it}^e\right)$ as in (132). This is the GIV formed of the idiosyncratic shock of all industries but industry i. We will use the shock to those other industries, and their impact on the outside wage, as an instrument to estimate labor demand. Indeed, we go back to the labor demand equation (206), and instrument for p_{it} using the z_{it}

$$p_{it} = b_i z_{it} + \varepsilon_{it}^p \tag{210}$$

we estimate b_i, and define $p_{it}^e = b_i z_{it}$ as the price in industry i instrumented by the changes in other industries. We use the estimated η_t^e as controls, and run

$$y_{it}^d = \phi_i^d p_{it}^e + C_{it} + \lambda_i^d \eta_t^e + u_{it}^d \tag{211}$$

which yields a consistent estimate ϕ_i^d of the labor demand.

Extension We can extend (207) to

$$y_{it}^s = \phi_i^s p_{it} - \psi_i \sum_j G_{ij} p_{jt} + \lambda_i^s \eta_t + u_{it}^s \tag{212}$$

where the "influence" matrix G_{ij} captures the influence of the price in market j on market i. This might be proxied by various measures of distance between the market. We can then use the material in Sections 4 and H.16 to handle this case.

[95] This procedure is much simplified if the γ_i and λ_i^p are assumed to be constant. Then, we can just define $z_t := p_{\Gamma t} = p_{St} - p_{Et}$, so that $z_t = v_{\Gamma t}$ and as $p_{St} = \frac{v_{St} + \lambda^p \eta_t^s}{1-\gamma}$, regressing $p_{St} = b z_t + \varepsilon_t^p$ yields $b = \frac{1}{1-\gamma}$.

References for Online Appendix

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

Amiti, Mary and David E Weinstein, "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data," *Journal of Political Economy*, 2018, *126* (2), 525–587.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International shocks, variable markups, and domestic prices," *The Review of Economic Studies*, 2019, *86* (6), 2356–2402.

Arellano, Manuel and Stephen Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 1991, *58* (2), 277–297.

Autor, David H, David Dorn, and Gordon H Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Bai, Jushan and Serena Ng, "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," *Econometrica*, 2006, *74* (4), 1133–1150.

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

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica*, 1995, pp. 841–890.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 1998, *87* (1), 115–143.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

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

Carvalho, Vasco and Xavier Gabaix, "The great diversification and its undoing," *American Economic Review*, 2013, *103* (5), 1697–1727.

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

EIOPA, "Technical Documentation of the methodology to derive EIOPA's risk-free interest rate term structures," *EIOPA-BoS-15/035, May 4, 2017*, 2017.

Fama, Eugene F and Kenneth R French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 1993, *33* (1), 3–56.

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

Goncalves, Silvia and Benoit Perron, "Bootstrapping Factor-augmented Regression Models," *Journal of Econometrics*, 2014, *182* (1), 156–173.

Graham, Bryan S, "Identifying social interactions through conditional variance restrictions," *Econometrica*, 2008, *76* (3), 643–660.

Heckman, James, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–59.

Horowitz, Joel L., "Bootstrap Methods in Econometrics," *Annual Review of Economics*, 2019, *11* (1), 193–224.

Huber, Peter J, "Robust estimation of a location parameter," *Annals of Mathematics Statistics*, 1964, *35*, 73–101.

Kelejian, Harry H and Ingmar R Prucha, "A generalized moments estimator for the autoregressive parameter in a spatial model," *International Economic Review*, 1999, *40* (2), 509–533.

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

Leary, Mark T and Michael R Roberts, "Do peer firms affect corporate financial policy?," *The Journal of Finance*, 2014, *69* (1), 139–178.

Lewbel, Arthur, "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models," *Journal of Business & Economic Statistics*, 2012, *30* (1), 67–80.

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

MacKay, Alexander and Nathan Miller, "Estimating models of supply and demand: Instruments and covariance restrictions," *Available at SSRN 3025845*, 2019.

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

Nevo, Aviv, "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 2000, *9* (4), 513–548.

Nevo, Aviv, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 2001, *69* (2), 307–342.

Newey, KW and D McFadden, "Large sample estimation and hypothesis," *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, 1994, pp. 2112–2245.

Rasmussen, Carl and Christopher Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2005.

Rigobon, Roberto, "Identification through heteroskedasticity," *Review of Economics and Statistics*, 2003, *85* (4), 777–792.

Sarto, Andres, "Recovering macro elasticities from regional data," *Unpublished manuscript, Princeton University*, 2018.

Sun, Qiang, Wen-Xin Zhou, and Jianqing Fan, "Adaptive huber regression," *Journal of the American Statistical Association*, 2020, *115* (529), 254–265.

Sweeting, Andrew, "Dynamic Product Positioning in Differentiated Product Markets: The Effects of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, 2013, *81* (5), 1763–1803.

Zou, Hui, Trevor Hastie, and Robert Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, 2006, *15* (2), 265–286.