GRANULAR INSTRUMENTAL VARIABLES

Xavier Gabaix
Ralph S. J. Koijen

## ABSTRACT

We propose a new way to construct instruments in a broad class of economic environments: "granular instrumental variables" (GIVs). 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 in order to create 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 first illustrate the idea in a basic supply and demand framework: we achieve a novel identification of both supply and demand elasticities based on idiosyncratic shocks to either supply or demand. We then show how the procedure can be enriched to work in many situations. We provide illustrations of the procedure with two applications. First, we measure how "sovereign yield shocks" transmit across countries in the Eurozone. Second, we estimate short-term supply and demand multipliers and elasticities in the oil market. Our estimates match existing ones that use more complex and labor-intensive (e.g., narrative) methods. 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
Çambridge, 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 general way to construct instruments: "granular instrumental variables" (GIVs). These instruments in turn allow researchers to establish causal relations in a wide variety of economic contexts.

In the economies we study, many decisions are taken by a few large actors, such as firms, industries, or countries, whose idiosyncratic shocks (e.g., 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 GIVs. The GIVs then allow us to estimate parameters of interest.

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 demand change – 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 "optimally extract" idiosyncratic shocks, thus constructing optimal GIVs.

Once the ideas are in place, we show in Section 3 how the procedure can be broadened to handle many extensions, such as feedback loops, heterogeneity, and several exogenous factors. Section 4 gives a practical user's guide.

**Empirical illustrations** We provide empirical results for two applications: sovereign yield spillovers and the equilibrium of global crude oil markets.

First, we study sovereign yield spillovers in the Eurozone in 2009-2018. 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 present a simple model that allows us to think about that and delivers a theoretically-grounded functional form for the shape of the spillovers. The modeling device we use is partial mutualization of debt, and we argue that other devices are likely to give a similar reduced form. Then, we use GIVs to estimate that spillover. Specifically, we extract idiosyncratic sovereign yield shocks and trace the impact of an idiosyncratic shock of one country on the other countries' yields. We find a "multiplier" of 1.5 (where unity would signify the absence of spillovers). To interpret this, suppose

---

[1]Hence, economies are "granular": their shocks are made of incompressible "grains" of economic activity, at the firm, industry, or country level. This theme is laid out in Gabaix (2011), and developed in Acemoglu et al. (2012), di Giovanni and Levchenko (2012), and Carvalho and Grassi (2019).

that Italy suffers a bad idiosyncratic shock that makes its debt likelier to default, so that the market value of Italy's debt falls by one 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 is an extra 0.5 billion euros of losses in European sovereign debt markets.

Second, we use GIVs to estimate the short-term impact of demand and supply shocks in global crude oil markets as a validation exercise. These multipliers, which depend on demand and supply elasticities, have been studied extensively in the literature using structural VARs. To identify the parameters of interest, the literature has relied on ordered VARs, sign restrictions with informative priors, and narrative methods. We use country-level oil supply growth to construct the GIV, after removing common factors using principal components and OPEC membership to construct an OPEC-specific factor. We find that the multipliers identified by GIV fall within the range documented in the literature.

This application also 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).

**Uses of GIVs**   GIVs allow to "democratize" and "automatize" instruments, especially in macro-finance where they have been rare. In standard practice, finding an instrument is a heroic and very ingenious affair. For instance, work on the "China shock" (the entry of China in the World Trade Organization, studied in Autor et al. (2013)) depends on detailed historical knowledge and applies only to a specific time period. With GIVs, we can have a more systematic way to obtain instruments, that can apply more generally and over many time periods.

Once one thinks about causality and GIV procedures, the answers to many interesting questions suddenly feel within reach. 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. (2020) study the impact of idiosyncratic shocks to fund-level rebalancing on exchange rates. Galaasen et al. (2020) 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 (2020), 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. (2020) study the impact of concentration on wages and use idiosyncratic firm-level shocks to instrument for concentration. Kundu and Vats (2020) estimate how idiosyncratic firm-level shocks in one state affect economic activity in other states via their transmission through the banking system.

We sketch a few further potential applications here, hoping that they will inspire other researchers to investigate these and related topics with the help of GIVs. In some cases, creative

instruments have been uncovered and GIVs may be useful to sharpen the estimates or to extend the estimates to other settings (e.g., other time periods, industries or countries).

The notion of doom loops is that when banks do badly, this will hamper the financial health of the state (as the state may need to bail out banks), and hence will increase the yield on the sovereign's debt. This in turn will create a fall in bank returns and result in a "doom loop." How important are those doom loops quantitatively? Using the idiosyncratic returns of large banks, GIVs allow us to estimate both channels, from banks to state and from state to bank. We plan to pursue this application.

GIVs can be used to estimate the pass-through from exchange rates to real economic activity. To take a concrete example, if the Turkish Lira appreciates, how does that affect Turkey's exports and borrowing? One could handle that via idiosyncratic demand shocks by large investment funds for the Turkish Lira.[2]

If there is an export boom, then what is the impact on the exchange rate, and the rest of the economy? Idiosyncratic shocks to large exporters will be useful to answer that question, as recent research has shown them to be very large (di Giovanni et al. (2014); Gaubert and Itskhoki (2020); Kramarz et al. (2020)).

Do firm-specific hiring, investment, and innovations spill over to peer firms operating in the same product market? That is, what is the sign and magnitude of strategic complementaries? Idiosyncratic innovation shocks to some large players will help construct the GIV.

How much do the constraints of financial intermediaries (such as broker-dealers) matter for asset prices? The GIV will rely on idiosyncratic shocks to intermediary wealth, which may be related to shocks to other parts of the banks.

How do international macroeconomic shocks propagate? For instance, how does a boom in Germany transmit to the rest of Europe? Using idiosyncratic shocks (differentiating between productivity and demand shocks) to countries will help us answer that question.

Likewise, how do regional "micro" shocks propagate into macro shocks? GIVs allow us to measure that and to estimate a micro-to-macro multiplier.[3]

**Related literature**   We relate to a number of literatures. We offer some brief pointers here, while offering a longer discussion in Section 8.2.

*Instruments for macro.* 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 systematically.

*Origins of aggregate fluctuations.* 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

---

[2]See Gabaix and Maggiori (2015) and Caballero and Simsek (2020) for models along those lines and Koijen and Yogo (2019) for a methodology for asset demand systems.

[3]We are pursuing this last question in ongoing work.

(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); Pasten et al. (2020)). We provide tools that can tease that apart in the presence of feedback loops. Datasets used in this literature can be revisited, forming GIVs which allow us to investigate causal relations. Gabaix (2011) introduces the notion of the "granular residual", a weighted sum of proxies for idiosyncratic shocks, and shows how idiosyncratic shocks to firms appear to explain about one third of GDP fluctuations. But that paper does not take the crucial step of using this kind of concept as an instrument to measure causal relations, for instance in a demand and supply setting.

The idea that we propose is in retrospect natural, so that we suspected it may have been already introduced in the literature, perhaps in a forgotten paper from the 1950s. However, after searching the literature and consulting with many experts, we did not find an earlier statement of it. We are quite sure that this idea has not been systematically implemented in mainstream economic applications. 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 8.2, such as Leary and Roberts (2014b), Amiti and Weinstein (2018), and Amiti et al. (2019). 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 presents a gentle introduction to the GIV framework, centered around a very simple model of supply and demand. It highlights the core intuition. Section 3 contains the general procedure. Section 4 gives a practical user's guide. Section 5 presents simulations. Sections 6 and 7 give empirical applications. Section 8 presents a number of extensions and robustness checks, and discusses more extensively the link with the rest of the literature. Section 9 concludes. Long proofs are in the online appendix.

**Notations** We will use the following notations. For a vector $X = (X_i)_{i=1...N}$ and a series of relative weights $S_i$ with $\sum_{i=1}^{N} S_i = 1$, we let

$$X_E := \frac{1}{N} \sum_{i=1}^{N} X_i, \qquad X_S := \sum_{i=1}^{N} S_i X_i, \qquad X_\Gamma := X_S - X_E, \qquad (1)$$

so that $X_E$ is the equal-weighted average of the vector's elements, $X_S$ is the size-weighted average, and $X_\Gamma$ is their difference.

We also commonly use the notation $u_i$ for shocks that are uncorrelated and with variance $\sigma_{u_i}^2$. Then, we will define the "inverse variance weights" or "pseudo-equal weights" (using that term to

highlight that they are a small variant of equal weights):

$$\tilde{E}_i := \frac{1/\sigma_{u_i}^2}{\sum_j 1/\sigma_{u_j}^2}, \tag{2}$$

which satisfy $\sum_i \tilde{E}_i = 1$, and are equal to $\tilde{E}_i = \frac{1}{N}$ when all the $\sigma_{u_i}^2$ are equal. Then $X_{\tilde{E}} := \sum_{i=1}^N \tilde{E}_i X_i$. We also define $\tilde{\Gamma}_i := S_i - \tilde{E}_i$. Then, $X_{\tilde{\Gamma}} = X_S - X_{\tilde{E}}$ will be the "granular residual" in a number of settings. It is the size-weighted sample average of $X$ minus the "inverse-variance" weighted sample average of $X$. It will be an optimal proxy for idiosyncratic shocks.

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. Lastly, 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; $\iota$ for a vector of 1's; $I$ for the identity matrix, 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); and $X \perp Y$ to say that the random variables $X$ and $Y$ are uncorrelated.

## 2 The basics of GIV

We introduce the GIV idea by considering a very simple example with supply and demand. We progressively extend it throughout the present section. The full GIV procedure is then outlined and synthesized in Section 3. The step-by-step approach we take in this section is helpful, we think, to progressively build an understanding of the GIV idea from its core to its more arduous extensions. We discuss extensions in Section 8.

### 2.1 A very simple example with no feedback loop

#### 2.1.1 Basic model

For clarity, we lay out a concrete economic model of the equilibrium in, for instance, the oil market. 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 disturbance term, and $S_i$ is country $i's$ share of demand, normalized to follow $\sum_{i=1}^N S_i = 1$. The demand does not depend on the price for now: this is for simplicity, and because it is a useful example of an economy without feedback loops. Once the basic ideas of the GIV are in place, we will easily extend this simple model to the case of a non-zero demand elasticity in Section 2.2. The demand disturbance is assumed to be the sum of a common shock $\eta_t$ and an idiosyncratic shock $u_{it}$:

$$y_{it} = \lambda_i \eta_t + u_{it}. \tag{3}$$

For now we consider the case with uniform loadings, $\lambda_i = 1$, but we will relax that soon.

All shocks are i.i.d. across dates for now (it is easy to relax that, see Section 4.2). 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 $Q_t = \bar{Q}\left(1 + \frac{p_t - \varepsilon_t}{\alpha}\right)$, where $p_t = \frac{P_t - \bar{P}}{\bar{P}}$ is the proportional deviation from $\bar{P}$, which is thus the average price of oil. Then, to equilibrate supply and demand $(D_t = Q_t(p_t))$, we must satisfy $\bar{Q}(1 + y_{St}) = \bar{Q}\left(1 + \frac{p_t - \varepsilon_t}{\alpha}\right)$. That is, the deviation of the price from its average satisfies:

$$p_t = \alpha y_{St} + \varepsilon_t. \tag{4}$$

It depends on the size-weighted average demand shock, $y_{St} = \sum_i S_i y_{it}$.

Throughout this paper we will make the mild assumption that all our variables (such as $\eta_t, \varepsilon_t, u_t = (u_{it})_{i=1\ldots N}$) have finite second moments.

The classic problem is that we cannot estimate $\alpha$ by OLS. Indeed, a direct regression of $p_t$ on $y_{St}$ (that is, a regression of the form $p_t = \alpha y_{St} + \varepsilon_t$) would be biased, as $\varepsilon_t$ and $\eta_t$ (hence $\varepsilon_t$ and $y_{St}$) can be correlated.

However, suppose that we form the GIV:

$$z_t := y_{\Gamma t} = y_{St} - y_{Et} = \sum_{i=1}^{N} S_i y_{it} - \sum_{i=1}^{N} \frac{1}{N} y_{it}. \tag{5}$$

Then, we have, using $u_{St} := \sum_{i=1}^{N} S_i u_{it}$, $u_{Et} := \sum_{i=1}^{N} \frac{1}{N} u_{it}$, that

$$y_{St} = \sum_{i=1}^{N} S_i y_{it} = \eta_t + u_{St}, \qquad y_{Et} = \sum_{i=1}^{N} \frac{1}{N} y_{it} = \eta_t + u_{Et},$$

so $z_t := y_{St} - y_{Et} = (\eta_t + u_{S_t}) - (\eta_t + u_{Et})$ satisfies

$$z_t = u_{St} - u_{Et} =: u_{\Gamma t}. \tag{6}$$

Note that $z_t := y_{St} - y_{Et}$ is just constructed from observables. It is the difference between the size-weighted demand and the equal-weighted demand. Intuitively, it captures the "idiosyncratic demand" by large units, as shown by $z_t = u_{\Gamma t}$.

We assume that the shocks $u_{it}$ are idiosyncratic, in the sense that:

$$\mathbb{E}[u_{it}\varepsilon_t] = 0 \text{ for all } i, t. \tag{7}$$

This "exogeneity" or "exclusion" assumption needs to be discussed in each economic application – as we will below.[4] Then, we have

$$\mathbb{E}[z_t\varepsilon_t] = 0 : \text{ Exogeneity}, \tag{8}$$

---

[4] The $u_{it}, u_{jt}$ could be correlated. If we have disaggregated data for both supply and demand, we can relax condition (7): see Section D.4.

$$\mathbb{E}\left[z_t y_{St}\right] \neq 0 : \text{ Relevance.} \tag{9}$$

Hence, $z_t = u_{\Gamma t}$ is a valid instrument (and as Proposition 3 will show, an optimal one).[5] We call it a "granular instrumental variable" (GIV).

Given that $p_t - \alpha y_{St} = \varepsilon_t$, we have

$$\mathbb{E}\left[\left(p_t - \alpha y_{St}\right) z_t\right] = 0, \tag{10}$$

that is, $\mathbb{E}\left[p_t z_t\right] - \alpha \mathbb{E}\left[y_{St} z_t\right] = 0$, which gives the supply elasticity $\alpha$, by

$$\alpha = \frac{\mathbb{E}\left[p_t z_t\right]}{\mathbb{E}\left[y_{St} z_t\right]}. \tag{11}$$

Indeed, in practice, we might estimate $\alpha$ using sample means:

$$\alpha_T^e := \frac{\frac{1}{T}\sum_t p_t z_t}{\frac{1}{T}\sum_t y_{St} z_t}. \tag{12}$$

We now state a formal proposition.[6]

**Proposition 1** (Consistency of the GIV estimator in this example). *Suppose that $\mathbb{E}\left[u_{it}\varepsilon_t\right] = 0$, although the $u_{it}$ can have an arbitrary distribution, with mean 0 and finite variance. Form the GIV estimator $z_t := y_{\Gamma t}$. Then, $z_t$ identifies the price elasticity, by $\alpha = \frac{\mathbb{E}[p_t z_t]}{\mathbb{E}[y_{St} z_t]}$. In other terms, for fixed $N$ but as $T \to \infty$, the GIV estimator $\alpha_T^e := \frac{\frac{1}{T}\sum_t p_t z_t}{\frac{1}{T}\sum_t y_{St} z_t}$ is consistent for the price elasticity $\alpha$.*

**Precision of the GIV estimator**  We define the excess Herfindahl as $h = \sqrt{-\frac{1}{N} + \sum_{i=1}^{N} S_i^2}$. In the context of industries, for example, a higher $h$ means that the industry is more concentrated: an industry where all the firms have the same size features $h = 0$. The quantity $h \in \left[0, \sqrt{1 - \frac{1}{N}}\right]$ will prove to be analytically useful, since if $(u_{it})_{i=1...N}$ is a series of i.i.d. uncorrelated random variables with mean 0 and common variance $\sigma_u^2$, then the volatility of the GIV $z_t = u_{\Gamma t}$ is:

$$\sigma_{u_\Gamma} = h\sigma_u. \tag{13}$$

In particular, under a standard central limit theorem, an appropriately scaled and centered version of the above GIV estimator is asymptotically normally for fixed $N$ as $T \to \infty$:

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

---

[5]The relevance condition is generically true, and otherwise assumes a very mild condition on the shares, see Proposition 2 and Assumption 1.

[6]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.

where $\sigma_\alpha^2$ is the asymptotic variance. The next proposition states the conditions under which the asymptotic variance of the estimator simplifies substantially (its proof is in Section C).

**Proposition 2** (Precision of the GIV estimator in this example). *If $\varepsilon_t$ is homoskedastic conditional on the $u_{it}$'s, then the asymptotic standard deviation of the scaled and centered GIV estimator is $\sigma_\alpha = \frac{\sigma_\varepsilon}{\sigma_{u_\Gamma}}$. If we assume further than the $u_{it}$ are i.i.d. with variance $\sigma_u^2$, then $\sigma_{u_\Gamma} = h\sigma_u$ and*

$$\sigma_\alpha = \frac{\sigma_\varepsilon}{h\sigma_u}, \tag{14}$$

*where $h$ is the excess Herfindahl:*

$$h := \sqrt{-\frac{1}{N} + \sum_{i=1}^{N} S_i^2}. \tag{15}$$

So in order to have a precise estimate (low $\sigma_\alpha$), we need: some large units (in order to have a large excess Herfindahl $h$), and that idiosyncratic shocks are large compared to aggregate shocks (large $\sigma_u/\sigma_\varepsilon$).

This simple example illustrates the basic idea. The reader might at this point have in mind a number of questions and objections: What if the factor structure is non-trivial (for instance, we don't have $\lambda_i = 1$ in (3))? What if the demand is sensitive to price? Is the GIV that we constructed the best instrument we can find? What happens if there are more feedback loops?

The next subsections are devoted to answering them in turn.

### 2.1.2 Optimality of the GIV weights

Above, we have shown that $z_t = y_{\Gamma t}$ allows for identification, for a specific $\Gamma = S - E$. It is easily verified that any GIV with weights $\Gamma$ such that $\sum_i \Gamma_i = 0$ would work. Hence, we can seek an optimal $\Gamma$. The $\Gamma$ we proposed is actually optimal, as we formalize below.

**Proposition 3** (Optimal weights $\Gamma$ for the GIV $y_{\Gamma t}$). *Consider the GIV $z_t = y_{\Gamma t} = \sum_i \Gamma_i y_{it}$, with some weights $\Gamma_i$ with $\sum_i \Gamma_i = 0$. Call $V^u$ the variance-covariance matrix of the idiosyncratic shocks $u_{it}$. Then, in the basic supply and demand model of Section 2.1, the asymptotic variance of the estimator $\alpha_T^e$ in (12), which is $\sigma_\alpha^2 = \lim_{T\to\infty} Tvar(\alpha_T^e - \alpha)$, satisfies $\sigma_\alpha^2(\Gamma) = \frac{\sigma_\varepsilon^2 \mathbb{E}[y_{\Gamma t}^2]}{\mathbb{E}[y_{St}y_{\Gamma t}]^2}$. The value*

$$\tilde{\Gamma} = S - \tilde{E}, \qquad \tilde{E} := \frac{(V^u)^{-1}\iota}{\iota'(V^u)^{-1}\iota}, \tag{16}$$

*gives the optimal GIV estimator, in the sense that for any other $\Gamma$ that is not collinear to $\tilde{\Gamma}$, the asymptotic variance $\sigma_\alpha^2(\Gamma)$ is larger. When the shocks are i.i.d., this implies $\tilde{E}_i = \frac{1}{N}$, and when they are uncorrelated, this implies $\tilde{E}_i := \frac{1/\sigma_{u_i}^2}{\sum_j 1/\sigma_{u_j}^2}$, so that $\tilde{E}$ may be called the "precision-weighted quasi-equal" weights.*

9

Hence, the "essence" of the GIV is not to be "size-weighted minus value-weighted" idiosyncratic shocks, but rather "size-weighted minus precision-weighted" (i.e. inverse-variance weighted when shocks are uncorrelated) idiosyncratic shocks.[7]

There are two more ways in which the GIV is optimal. First, it is the optimally-weighted GMM estimator.[8] This implies that other combinations of idiosyncratic shocks (besides weighing by $\Gamma$) would not help the precision of the estimator. Second, one can show that it is the maximum likelihood estimator, if we assume that all shocks are Gaussian (see Section E). Still, the optimality formulation of Proposition 3 is the simplest to use in other contexts.

### 2.1.3 Time-varying size weights

Suppose that we have a time-varying size $S_{i,t-1}$, so that the demand increase is $\sum_i S_{i,t-1} y_{it}$, with $\mathbb{E}\left[u_{it}\left(\eta_t, \varepsilon_t\right)'\left(1, S_{i,t-1}\right)\right] = 0$. Then everything goes through without problems, replacing $S_i$ by $S_{i,t-1}$ throughout. The basic GIV becomes: $z_t = y_{S_{t-1},t} - y_{Et} = \sum_i \left(S_{i,t-1} - \frac{1}{N}\right) y_{it}$.

### 2.1.4 Model with an enriched factor structure

An important extension is that the shocks might have a richer factor structure, with $r$ factors, so that instead of (3) we have:

$$y_{it} = \sum_{f=1}^{r} \lambda_i^f \eta_t^f + u_{it}, \tag{17}$$

or, in vector form:

$$y_t = \Lambda \eta_t + u_t, \tag{18}$$

where $\Lambda$ is a $N \times r$ matrix, and $\mathbb{E}\left[u_t\left(\eta_t, \varepsilon_t\right)\right] = 0$.

Then, in order to construct a valid GIV we simply run a factor model – for example, via Principal Component Analysis (PCA) – and, in essence, we extract the residuals $u_{it}$ to form the GIV. Calling $\check{u}_{it}$ the residuals from regression (17), we form the GIV as the share-weighted average idiosyncratic shock:

$$z_t := S'\check{u}_t = \sum_i S_i \check{u}_{it}. \tag{19}$$

Let us see more precisely what happens "inside the model" when we do that. Suppose that we know the $\lambda$ vectors (for example, they are country characteristics, or they have been estimated by PCA, as we will detail later). Then, let $Q$ be a $N \times N$ matrix projecting vectors onto a space orthogonal to $\Lambda$, so that $Q\Lambda = 0$. Then, $Qy_t = Qu_t$ (to see that, premultiply (18) by $Q$). Hence,

---

[7]The fact that one wants to precision-weight variables is of course well-appreciated from the literatures on Bayesian inference and generalized least squares.

[8]Any moment $\mathbb{E}_T\left[\left(p_t - \alpha y_{St}\right)\left(u_{it} - u_{Et}\right)\right] = 0$ is a valid GMM moment to identify $\alpha$. 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 $\mathbb{E}_T\left[\left(p_t - \alpha y_{St}\right)\left(u_{St} - u_{Et}\right)\right] = 0$.

via PCA, estimating $\eta_t$ in (17) yields the transformed residuals $\check{u}_t = Qu_t$. Then, the GIV is formed as:

$$z_t := S'\check{u}_t = S'Qy_t = \Gamma'y_t = \Gamma'u_t, \qquad \Gamma := Q'S. \tag{20}$$

Then, $z_t$ is a valid instrument, since it is composed of idiosyncratic shocks, as $z_t = \Gamma'u_t$. Since $p_t - \alpha y_{St} = \varepsilon_t$ and $\mathbb{E}[u_t\varepsilon_t] = 0$, we have $\mathbb{E}[(p_t - \alpha y_{St})z_t] = 0$, which is (10).[9]

We now detail the value of the matrix $Q$ – this is more advanced material and should be skipped at the first reading. Let us say that we have estimated $\eta_t$ in (18) by generalized least squares with a weight matrix $W$ of size $N \times N$ (one could use $W = I$, but typically it is optimal to take $W = c(V^u)^{-1}$ for $c$ an arbitrary constant).[10] To represent this, we define two useful matrices, $R^{\Lambda,W}$ and $Q^{\Lambda,W}$, with respective dimensions $r \times N$ and $N \times N$:

$$R^{\Lambda,W} := (\Lambda'W\Lambda)^{-1}\Lambda'W, \qquad Q^{\Lambda,W} := I - \Lambda R^{\Lambda,W}. \tag{21}$$

Roughly, $Q$ is the projection on the space orthogonal to $\Lambda$, and $R$ is the projection on $\Lambda$, with a scalar product that depends on $W$.[11]

Constructing the GIV as (20) generalizes our basic example (3) (when $W = I$) and the heteroskedastic shock case of Section 2.1.2, with $W = (V^u)^{-1}$. Indeed, in that example with uniform loadings we had $\Lambda = \iota$, $W = I$, $Q = I - \iota\tilde{E}'$, where $\tilde{E} = \frac{W\iota}{\iota'W\iota}$, hence $E = \frac{\iota}{N}$ when $W = I$ and $E = \tilde{E}$ defined in (16) when $W = (V^u)^{-1}$. So that $\check{u}_{it} = u_{it} - u_{\tilde{E}t}$, and the GIV was: $z_t = \check{u}_{St} = u_{St} - u_{\tilde{E}t}$. We therefore had $\Gamma = Q'S = S - \tilde{E}$.[12]

## 2.2 A simple demand and supply example with feedback loops

**A simple model** We next enrich the previous example, and consider a simple supply and demand example that features a "loop." Suppose that demand for some commodity (say, oil) is:

$$y_{it} = \phi^d p_t + \eta_t + u_{it}, \tag{24}$$

---

[9]This shows that the GIV is valid and possible as long as $\Gamma := Q'S \neq 0$. This is a very mild assumption, as discussed in Assumption 1. If $\Gamma$ were close to 0, that would be picked up by very large standard errors.

[10]This is, we estimated $\eta_t$ by $\min_{\eta_t^e} \|y_t - \Lambda\eta_t^e\|_W^2$ with $\|x\|_W^2 := x'Wx$. $W$ is assumed to be positive definite.

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

$$Q\Lambda = 0, \quad R\Lambda = I, \quad Q'W\Lambda = 0, \quad (I - Q)W^{-1}Q' = 0, \quad (I - Q')WQ = 0, \quad Q^2 = 0, \quad RW^{-1}Q' = 0. \tag{22}$$

[12]This implies that $\sigma_z^2 = \sigma_u^2 S'QS$ if the $u_{it}$ are homoskedastic. For instance, if $\lambda_i = (1, \check{x}_i)$ with $\check{x}_E = 0$, the variance of the GIV is

$$\sigma_z^2 = \sigma_u^2\left(h^2 - \frac{1}{N}\frac{\check{x}_S^2}{\check{x}_E^2}\right). \tag{23}$$

This illustrates how controlling for more factors reduces the standard deviation of the GIV, hence (as in Proposition 2, which features a standard deviation of the estimator $\sigma_\alpha = \frac{\sigma_\varepsilon}{\sigma_z}$), especially if $\check{x}_S^2$ 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 $\eta_t$ easier, and hence increase the precision of the GIV estimator (that is, increase $\sigma_z^2$ by shrinking the last term in (23), $\frac{1}{N}\frac{\check{x}_S^2}{\check{x}_E^2}$).

and supply is

$$s_t = \phi^s p_t + \varepsilon_t, \tag{25}$$

where $\eta_t, \varepsilon_t$ can be correlated. We can expect that the demand and supply elasticities (respectively $\phi^d$ and $\phi^s$) satisfy $\phi^d < 0 < \phi^s$. Again, to be more formal, $y_{it}$, $s_t$, and $p_t$ are understood as percent deviations from the average demand of country $i$, from supply, and from price, respectively.[13],[14]

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

$$p_t = \frac{u_{St} + \eta_t - \varepsilon_t}{\phi^s - \phi^d}. \tag{26}$$

There is a "loop" because the demand shocks $\eta_t$ and $u_{it}$ feed into the price $p_t$, which then in turns affects demand. The equilibrium quantity produced is

$$s_t = y_{St} = \frac{\phi^s u_{St} + \phi^s \eta_t - \phi^d \varepsilon_t}{\phi^s - \phi^d}. \tag{27}$$

The classic problem of estimating supply and demand equilibrium quantity $s_t$ and price $p_t$ is that we cannot regress: $s_t = \beta p_t + \varepsilon_t$, and hope to get $\beta = \phi^s$, as $\varepsilon_t$ and $p_t$ are correlated.

However, suppose that we form the GIV, as in (5):

$$z_t := y_{St} - y_{Et}. \tag{28}$$

Given that

$$y_{St} = \phi^d p_t + \eta_t + u_{St}, \qquad y_{Et} = \phi^d p_t + \eta_t + u_{Et},$$

we have:

$$z_t = u_{St} - u_{Et} =: u_{\Gamma t}. \tag{29}$$

As in the previous example, we assume that the shocks $u_{it}$ are idiosyncratic:

$$\mathbb{E}\left[u_{it}\eta_t\right] = \mathbb{E}\left[u_{it}\varepsilon_t\right] = 0 \text{ for all } i, t. \tag{30}$$

Then, we have again a valid instrument:

$$\mathbb{E}\left[z_t \varepsilon_t\right] = \mathbb{E}\left[z_t \eta_t\right] = 0 : \text{ Exogeneity,}$$

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

---

[13]We take the model of Section 2.1, and simply set $y_{it} = \phi^d p_t + \eta_t + u_{it}$, where $p_t = \frac{P_t - \bar{P}}{\bar{P}}$ is the proportional deviation from the average.

[14]For simplicity, we set all constant additive terms to 0. With non-zero such terms, one can talk about quantities and prices, rather than their deviations from the mean.

**Estimations of supply and demand elasticities by GIV**   The supply equation (25) implies $\mathbb{E}\left[(s_t - \phi^s p_t)\, z_t\right] = 0$, which gives the supply elasticity $\phi^s$ by

$$\phi^s = \frac{\mathbb{E}\left[s_t z_t\right]}{\mathbb{E}\left[p_t z_t\right]}. \tag{31}$$

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

Now, we want to estimate demand. For that, we make a stronger assumption: we assume that the shocks $u_{it}$ are i.i.d. across $i$'s and not just dates (we will relax this later). Then, this implies[15]

$$\mathbb{E}\left[u_{Et} u_{\Gamma t}\right] = 0. \tag{32}$$

So, given this, we have: $y_{Et} - \phi^d p_t = \eta_t + u_{Et}$, and $\mathbb{E}\left[\left(y_{Et} - \phi^d p_t\right) z_t\right] = 0$. This gives an estimate of the demand elasticity $\phi^d$,

$$\phi^d = \frac{\mathbb{E}\left[y_{Et} z_t\right]}{\mathbb{E}\left[p_t z_t\right]}, \tag{33}$$

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

**Estimation by OLS and interpreting it as a first- and second-stage IV estimator**   Let us recast our GIV in the language of applied microeconomics, and estimate the parameters by OLS (as we will often do in the general case). Much of this subsectionis very elementary, but for completeness and future reference we spell it out. Recall that the solutions are:

$$p_t = \frac{1}{\phi^s - \phi^d} u_{St} + \dot{\varepsilon}_t^p, \qquad s_t = y_{St} = \frac{\phi^s}{\phi^s - \phi^d} u_{St} + \dot{\varepsilon}_t^s,$$

where the $\dot{\varepsilon}_t^p, \dot{\varepsilon}_t^s$ are linear combinations of $\varepsilon_t, \eta_t$. So, if we run the OLS regression, with $z_t = u_{\Gamma t}$,

$$p_t = b^p z_t + \varepsilon_t^p, \tag{34}$$

we estimate $b^p = \frac{1}{\phi^s - \phi^d}$, which is the sensitivity of the price to the supply or demand shock. If we run the OLS regression

$$y_{St} = b^{y_S} z_t + \varepsilon_t^s, \tag{35}$$

we estimate $b^{y_S} = \frac{\phi^s}{\phi^s - \phi^d} = M$.

In the language of applied microeconomics, one can view the "first stage" as a regression of the

---

[15]Indeed, in the i.i.d. case we have $\mathbb{E}\left[u_{Et} u_{\Gamma t}\right] = \mathbb{E}\left[(\sum_i \frac{1}{N} u_{it})(\sum_i \Gamma_i u_{it})\right] = \frac{1}{N}\sum_i \Gamma_i \sigma_u^2 = 0$ as $\sum_i \Gamma_i = 0$. Equation (69) generalizes this to the non-i.i.d. case.

price on the GIV (34), which yields the instrumented price

$$p_t^e := b^p z_t. \tag{36}$$

The "second stage" regresses supply on the instrumented price:

$$s_t = \phi^s p_t^e + \varepsilon_t^s, \tag{37}$$

which gives $\phi^s$. Alternatively, one can run the "reduced form equation" (35), which estimates $M$. The supply elasticity is given by: $\phi^s = \frac{b^{y_S}}{b^p}$.

The demand elasticity is similar. In the second stage we regress equal-weighted demand on the instrumented change in the price, $p_t^e$:

$$y_{Et} = \phi^d p_t^e + \varepsilon_t^y, \tag{38}$$

which gives the demand elasticity $\phi^d$.[16] Alternatively, we can estimate the reduced form equation $y_{Et} = b^{y_E} z_t + \varepsilon_t^y$, which gives $b^{y_E} = \phi^d M$, and the demand elasticity is $\phi^d = \frac{b^{y_E}}{b^p}$.

In practice, we will add controls to those regressions, including estimates of $\eta_t$ recovered from PCA. This improves the precision of the estimator, by absorbing some of the variance of the noise.

From $M$ and $b^p$, we can recover the elasticities $\phi^s$ and $\phi^d$. This is exactly the same estimate as the IV estimator, derived earlier in (31) and (33).[17]

**Standard errors: When "weak instruments" are or are not a problem**  When estimating via OLS (e.g. $b^p$ and $M$), the standard errors are reliably estimated by the usual OLS method, even in small samples. When a ratio is implicitly performed (e.g. to estimate $\phi^d$, $\phi^s$), the two stage least square (2SLS) procedure as in (37) will also give correct standard errors when the instrument is strong enough. A traditional rule of thumb for the strength of the instrument (in the i.i.d., homoskedastic case) is that the $F$ statistics (which is the squared $t$-statistic on $b^p$) on the first stage (34) should be greater than 10, and this advice is being progressively enhanced in current IV research.[18]

## 2.3   Interpreting and diagnosing idiosyncratic shocks

**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.

---

[16]Here we used (32), which makes the OLS valid.

[17]Indeed, the OLS estimators are $M_T^e = \frac{\mathbb{E}_T[y_{St} z_t]}{\mathbb{E}_T[z_t^2]}$ and $b_T^e = \frac{\mathbb{E}_T[p_t z_t]}{\mathbb{E}_T[z_t^2]}$. We have $\phi_T^{s,e} = \frac{M_T^e}{b_T^e} = \frac{\mathbb{E}_T[y_{St} z_t]}{\mathbb{E}_T[p_t z_t]}$, which is the same as (31), as $s_t = y_{St}$ in equilibrium.

[18]See the literature on weak instruments, such as Stock and Yogo (2005) and Andrews et al. (2019).

In some cases it is quite clear – for example, a random productivity shock, 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, suppose that OPEC decided to cut down production, which in the language of our example is an aggregate $\eta_t$ shock. If Saudi Arabia cuts down production by more than anticipated, that is an idiosyncratic shock. Formally, if supply is $y_{it} = \phi^s p_t + \left(\lambda_i + \check{\lambda}_{it}\right)\eta_t + v_{it}$, with $\mathbb{E}_{t-1}\left[(1, \tilde{\eta}_t)\,\check{\lambda}_{it}\right] = 0$, then $u_{it} = \check{\lambda}_{it}\eta_t + v_{it}$ is a bona fide idiosyncratic shock. To take another example, suppose that we hear about a change in real estate prices in the economy, $\eta_t$, but that a bank $i$ was more exposed to it than anticipated: the market thought the bank's equity would move by $\lambda_i\eta_t$, but it moved by $r_{it} = \left(\lambda_i + \check{\lambda}_{it}\right)\eta_t$ for an expectational surprise $\check{\lambda}_{it}$ with $\mathbb{E}_{t-1}\left[(1, \tilde{\eta}_t)\,\check{\lambda}_{it}\right] = 0$. Then, the bank will have an idiosyncratic shock $u_{it} = \check{\lambda}_{it}\eta_t$ as part of its total return $r_{it}$.

Likewise suppose that the news is that a bank failed a stress test (while it was anticipated it would pass the test). This is an idiosyncratic shock. However, the bank could have failed the test because of some development in the macroeconomy $\eta_t$. Then, provided that the factor model allows for a rich enough structure in $\eta_t$, the latter will be controlled for.

The volatility of idiosyncratic shocks can depend on the common shocks. Suppose that $u_{it} = \sigma_t v_{it}$ where $\sigma_t$ and $\eta_t$ could be correlated (for instance, $\sigma_t$ could increase when $|\eta_t|$ is high), but $\mathbb{E}_{t-1}\left[\tilde{\eta}_t\sigma_t v_{it}\right] = 0$ (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 $\mathbb{E}_{t-1}\left[\tilde{\eta}_t u_{it}\right] = 0$.

**Thresholded and narrative GIVs**  In applications, it is possible to make further progress by assessing the drivers of the top shocks narratively. One procedure is to simply select the top $K$ shocks by $S_i|\check{u}_{it}|$ (where $\check{u}_{it}$ is the residual from factor analysis, e.g. $\check{u}_{it} = u_{it} - u_{Et}$, and we select across all actors $i$ and dates $t$), and check in the news what happened on that day (and check that the shocks are idiosyncratic indeed). We do that for some our applications. Formally, that means that we formulate a "thresholded" GIV,

$$z_t^\tau = \sum_i \tau\left(S_i\check{u}_{it}\right), \tag{39}$$

using the thresholding function $\tau(x) = x 1_{|x|\geq b}$, which only keeps granular shocks bigger than $b > 0$.[19] Then, the GIV procedure works using that "thresholded" GIV (see Section D.5). This thresholded GIV might also be useful to assess non-linear effects, for instance, in case of demand or supply curves.

After examining those largest shocks by looking at the news, some shocks might be eliminated as not idiosyncratic; we can call $I_t^{\mathcal{N}}$ the set of shocks that are "narratively certified" to be idiosyncratic

---

[19]We adjust $b$ to select a pre-specified expected number $K$ of shocks that survive the thresholding.

by this procedure, and form the alternative instrument

$$z_t^{\mathcal{N}} = \sum_{i \in I_t^{\mathcal{N}}} S_i \check{u}_{it}. \tag{40}$$

This is roughly what the "narrative" approach in the literature (e.g. Caldara et al. (2018)) 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 $\eta_t^e$ that are usable when running regressions, which increases the precision of estimators.

## 2.4 Over-identification test with multiple GIVs

Consider the model of demand with a single factor with heterogenous exposures

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

and assume we obtained an estimate of the factor, $\eta_t^e$. We abstract from estimation error in the factor in this section.

It is easy to get overidentifying tests with GIVs, because each $\check{u}_{it}^e$ is potentially a valid instrument. One way to do that is to construct two different GIV instruments, $z_{1t}$ and $z_{2t}$. These could include size-weighted averages of all $\check{u}_{it}^e$, a subset of the largest realizations of $S_i |\check{u}_{it}^e|$, or a subset of narratively-checked shocks as in Section 2.3. Alternatively, $z_{1t}$ (respectively $z_{2t}$) could be the size-weighted sum of the shocks to odd-numbered (respectively even-numbered) entities, or $z_{1t}$ might be based on idiosyncratic supply shocks and $z_{2t}$ on idiosyncratic demand shocks, as in Section D.4.[20]

Suppose we have constructed these two different GIV instruments, $z_{1t}$ and $z_{2t}$. One can then estimate separately the parameters of interest (e.g. $\phi^d$) based on $z_{1t}$ and $z_{2t}$, and see if they are economically different. We can also do a formal test.[21] In Gabaix and Koijen (2020), we provide an example of such a procedure. There, the goal is to estimate the elasticity of the aggregate equity demand curve using investor-level equity holdings data. After extracting the idiosyncratic shocks, we form two instruments. We rank investors by size in each period, and form an instrument based

---

[20]Yet another GIV procedure is to use characteristics $x_{it}$ measurable at time $t - 1$ (e.g., firm size, or GDP per capita, or a bank's credit risk), and form the $x$-weighted GIV: $z_t^x \coloneqq \sum_i S_i \check{u}_{it}^e x_{it}$. If the test fails, it's probably the case that $x_{it}$ is economically important and it should have been included as a factor loading in a larger factor model.

[21]We form the moment conditions as $\mathbb{E}[g_t(\theta)] = 0$, where $g_t(\theta) = (y_{Et} - \phi^d p_t - \lambda_E \eta_t^e)(z_{1t}, z_{2t}, \eta_t^e)'$ and $\theta = (\phi^d, \lambda_E)$. We can simply perform the Sargan-Hansen $J-$test for over-identifying moment conditions. The test statistic is given by $J = T g_t' W_T g_t \to^d \chi_1^2$, under the null, where $W_T = \left(\frac{1}{T} \sum_{t=1}^T g_t g_t'\right)^{-1}$. If we have $K$ instruments $z_{kt}$, $k = 1 \ldots K$, then the procedure is the same, with $g_t(\theta) = (y_{Et} - \phi^d p_t - \lambda_E \eta_t^e)(z_{1t}, \ldots, z_{Kt}, \eta_t^e)'$, and then $J \to^d \chi_{K-1}^2$.

on the odd and even ranks. We show that the point estimates using either instrument are not significantly different and that the two instruments are uncorrelated.[22]

## 2.5   Heterogeneous demand elasticities

We consider a generalization of the basic model to allow for heterogeneity in demand elasticities. We start from

$$y_{it} = \phi_i^d p_t + \lambda_i \eta_t + u_{it}, \tag{41}$$

while the supply curve is the same as in (25).

We consider the parametric version of the model (Section D.2 of the online appendix considers a non-parametric version, which is more complex). We suppose that the loadings are linear in the characteristics $X_{it}$ (a $k$-dimensional vector, with $k$ in practice a small number, and the first entry being a 1), so they can be expressed as:

$$\phi_i^d = X_{it}\dot{\phi}^d = \sum_{\ell=1}^{k} X_{i\ell t}\dot{\phi}_\ell^d$$

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 $\eta_t$ has dimension $r \times 1$. We want to evaluate $\phi^s$ and $\dot{\phi}^d$. Here is how we proceed.

For each date $t$ we run the cross-sectional regression of $y_{it}$ on $X_{it}$ (potentially the generalized least squares regression with weights $W$ to be discussed soon):[23]

$$y_{it} = X_{it}\dot{y}_t + \check{u}_{it} = \sum_{\ell=1}^{k} X_{i\ell t}\dot{y}_{\ell t} + \check{u}_{it}, \tag{42}$$

and we get regression slopes $\dot{y}_t = (\dot{y}_{\ell t})_{\ell=1\ldots k}$, then collect the residuals $\check{u}_{it}$. We next form the GIV $z_t := \sum_i S_i \check{u}_{it}$ as in (20). We then identify $\phi^s$ and $\dot{\phi}^d$ instrumenting $p_t$ by $z_t$, i.e. using the IV moments:[24]

$$\mathbb{E}\left[(s_t - \phi^s p_t) z_t\right] = 0, \tag{43}$$

$$\mathbb{E}\left[\left(\dot{y}_{\ell t} - \dot{\phi}_\ell^d p_t\right) z_t\right] = 0 \qquad \text{for } \ell = 1 \ldots k. \tag{44}$$

The identification of $\phi^s$ is valid for any weight matrix $W$ (including the identity matrix, as in

---

[22]If there are missing common factors, then the instruments are likely correlated. This procedure therefore provides an additional way to check for omitted common factors.

[23]That is, we use $\min_{\dot{y}_t} \sum_{i,j} (y_{it} - X_{it}\dot{y}_t) W_{ij} (y_{jt} - X_{jt}\dot{y}_t)$ to identify $\dot{y}_t$ at each date.

[24]Alternatively, we can use the first stage and second stage language of Section 2.2. In practice, to increase precision, we add controls to those regressions, including the estimates of $\eta_t$ recovered from PCA.

OLS) and that of $\dot{\phi}^d$ is valid if we have the "inverse variance" weights $W = c\,(V^u)^{-1}$ for some constant $c$.[25] Section C.4 gives the derivations, and discusses how to estimate $W$ if necessary.

## 2.6  Robustness to misspecification and threats to identification

The GIV procedure is robust to some forms of misspecification, and more fragile to others.

**Robustness of the GIV**  We may keep only the shocks to some actors (in a set $I_t$), i.e. set $z_t = \sum_{i \in I_t} S_i \breve{u}_{it}^e$ (with $\breve{u}_{it}^e = u_{it} - u_{Et}$), selecting for example the shocks to the top $K$ entities, the shocks for which we have data, or some subset of the entities based on size. Then again, everything goes through.[26] The estimator is still valid, just not the optimal GIV estimator.

The GIV also works if we have information only on some of the large actors (such as countries or firms): one uses the idiosyncratic shocks of those actors for which we have data.

Suppose that we misspecify the vector $S$ of sizes, for example by defining $z_t = \sum_i S_i^\circ \breve{u}_{it}^e$ using a wrong vector $S^\circ$. Then, the IV is still valid, but the OLS can be biased. In our basic example of Section 2.1, we still have $\mathbb{E}\left[(p_t - \alpha y_{St})\,z_t\right] = 0$, so that the IV procedure (11) still works. Likewise, in the more complex supply and demand case, the IV relations (31) and (33) still hold. But the OLS relations are slightly biased.[27]

If we assume homogeneous coefficients (e.g. on the elasticities of demand or supply), while in fact they are truly heterogeneous, then again (assuming that $\eta_t$ was well-estimated in the cross-section) the IV estimates are correct, and so are the OLS estimates, except that we obtain the equal-weighted averages of coefficients. For instance, the IV estimates yield $\phi^s$, $\phi_E^d$, and the OLS coefficients are those corresponding to the interpretation that the elasticity of demand is $\phi_E^d$ rather that $\phi_S^d$. Section D.6 provides the derivations.

If we misspecify the variance of the $u_{it}$ (but keeping them uncorrelated), things are essentially fine: 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 $N$.

**Threats to identification**  The threat to identification is that we might 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 $\eta_t^e$ in the regression we will not completely eliminate the $\lambda_\Gamma \eta_t - \lambda_\Gamma^e \eta_t^e$ error.[28] This danger is greater when $|\lambda_\Gamma|$ is greater, i.e. when loadings are correlated with size (indeed, we are safe if $\lambda_\Gamma = 0$). This

---

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

[26] For instance, we still have $u_{St} = z_t + \varepsilon_t^{us}$ with $z_t \perp \varepsilon_t^{us}$. Section D.5 gives for a formal analysis.

[27] 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 (in population) $b^{p,e} = b^p \psi$ 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.

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

is a small sample problem (with a large enough $T, N$ we measure $\eta_t, \lambda$ accurately). Fortunately, there are at least four approaches to handle this concern.

A first check, which we recommend, it to perform the over-identification test of Section 2.4, based on some meaningful alternative GIVs $z_t$, for example based on the shocks to some top actors, or their even-odd ranking (as we do in Gabaix and Koijen (2020)).

Second, one can perform a formal test for the number of factors (Bai and Ng (2002); Onatski (2009)), as we illustrate in Section 5. A missing factor may also be detected by testing the stability of estimates across GIVs: 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 recommended by formal tests for the number of factors.[29]

A third approach is to opt for the narrative GIV of Section 2.3, and as we do in our empirical applications below. If one checks the top, say, 15 events and they pass the narrative check, one can record that we could not reject the hypothesis of a misspecified factor model. In a more purist way, one could even restrict the GIV to those top events.

A fourth approach is to filter out "sporadic factors," i.e. a factor $\eta_t$ that affects a few actors special ways, but happens once in the sample, so is hard to detect with standard factor models. Then, one removes the dates with such sporadic factors, after identifying them via a narrative check, or via statistical methods as discussed in Section D.7. Section 6.6 provides an illustration of this.

# 3    General setup and multipliers

The previous section introduced the GIV in a simple context, with no loops or a single loop. We now propose a more general setup with potentially several factors, arbitrary loop structure, and rich heterogeneity. This section can be skipped when reading the paper for the first time and the reader can continue with Section 4.

## 3.1    Framework

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

$$y_{it} = \sum_f \lambda_{it}^f F_t^f + u_{it} + C_{it}^y m, \tag{45}$$

where each $F_t^f$ is a factor, $\lambda_{it}^f$ are factor loadings, $u_{it}$ is an idiosyncratic shocks, and $C_{it}^y$ is a vector of controls that may include lagged demands and other characteristics. We could also add constants,

---

[29]We do that in Section 5 and in Gabaix and Koijen (2020).

but we omit them for notational simplicity. Factor $f$ follows:

$$F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f. \tag{46}$$

It depends on an exogenous shock $\eta_t^f$, 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 (45)-(46) 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 $\alpha^f$ may be non-zero. As in the rest of the paper, we make the mild assumption that all our variables (e.g. $\eta_t^f, u_t$) 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:

$$\lambda_{it}^f = \dot{\lambda}_0^f + \dot{\lambda}_1^f x_{it}^f, \tag{47}$$

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:

$$\mathbb{E}\left[ u_{it} \left( \eta_t^f, C_t^y, C_t^f, x_t^f \right) \right] = 0, \tag{48}$$

but the $\eta_t^f$ may be correlated across $f$'s, and $\eta_t^f$ 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 (45) in vector form:

$$y_t = \Lambda_t F_t + u_t + C_t^y m, \qquad F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f, \tag{49}$$

with $\Lambda_t$ 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.[30]

---

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

## 3.2 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 $\alpha$ is the vector stacking the $\alpha^f$'s and $\varepsilon^y_t$ 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,

$$y_{St} = M_t \left( u_{St} + \varepsilon^y_t \right), \tag{50}$$

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

$$M_t = \frac{1}{1 - \Lambda_{St}\alpha} = \frac{1}{1 - \sum_f \Lambda^f_{St}\alpha^f}. \tag{51}$$

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:

$$F^f_t = M_t\alpha^f u_{St} + \varepsilon^f_t, \tag{52}$$

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 feedback loops in the economy, including those operating via the unobservable, endogenous factors. However, we can obviously not estimate $\alpha^f$ for those unobserved factors.

## 3.3 A formal identifiability result

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

**Assumption 1** (Condition for identification with GIV) *The vector $V^u S$ is not spanned by the factors loadings $\lambda^f$ (where $V^u$ is the covariance matrix of $u_t$).*

Assumption 1 ensures that the GIV is not identically 0 (as $z_t := S'Q^{(V^u)^{-1},\Lambda}_t u_t$, as in (20) and (21)). Economically, this assumption seems like a mild restriction. It is generically satisfied.[31,32]

For simplicity, we shall make here a strong further Assumption 2, which can be relaxed.

---

[31]One case that does prevent this assumption to hold is the case where the variance would be inversely proportional to size: then, GIV would fail, as then $V^u S = a\iota$ for some scalar $a$. 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)).

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

**Assumption 2** (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 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 $\Lambda_{St}$ 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 in Section C.5.

**Proposition 4** (Sufficient condition for identification with GIV) *Consider the factor model above, when $N$ is fixed but $T \to \infty$, and make Assumption 1 and 2. We assume the parametric case, where we know the actor (e.g., firm or country) characteristics $X_{it}$. Then, we can identify $\alpha^f$ 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.*

It may seem a bit surprising that the OLS standard errors are correct. The reason is that the GIV is directly obtained from an exact formula ($z_t := S' Q_t^{(V^u)^{-1}, \Lambda} y_t$ as in (20)), hence has no estimation error. The standard errors are made explicit in the proof of Proposition 4.

Proposition 4 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.

# 4 A user's guide

We summarize the above arguments in the form of a user's guide. The arguments in this section extend to multiple factors, as illustrated in Section 3.

## 4.1 User's guide: basics

The model can be summarized by

$$y_{it} = \phi^d p_t + m C_{it}^y + a_i + \lambda_{it} \eta_t + u_{it}, \tag{53}$$

$$p_t = \alpha y_{St} + m^p C_t^p + b + \eta_t^p, \tag{54}$$

where $C_{it}^y$ and $C_t^p$ are observable controls, which we assume to be scalar for notational convenience but the arguments directly extend to vectors of controls. We also assume that shocks are homoskedastic and discuss the extension to heteroskedastic shocks below. In outlining the basic

algorithm, we consider the possibility that we have additional information on factor loadings, $x_{it}$, so that $\lambda_{it} = \lambda_0 + \lambda_1 x_{it}$ for some factors, although this information is not required.

In describing the procedure, we refer to generic agent fixed effects as $a_i$ and generic time fixed effects as $b_t$. The basic algorithm is summarized by:

1. *Panel regression:* We estimate a panel regression with agent and time fixed effects,

$$y_{it} = a_i + b_t + mC_{it}^y + \check{y}_{it}.$$

   We refer to the estimated residuals as $\check{y}_{it}^e$.

2. *Factor estimation:* If information on factor loadings is available, we estimate a first set of factors using period-by-period cross-sectional regressions,

$$\check{y}_{it}^e = b_t + x_{it}\eta_t^x + e_{it},$$

   and we refer to the estimated factors as $\eta_t^{x,e}$. In addition, we estimate nonparametric factors using PCA on $\check{y}_{it}^e$ and refer to the estimated factors as $\eta_t^{PCA,e}$.[33] We stack the estimated factors together in $\eta_t^e = \left( \eta_t^{x,e}, \eta_t^{PCA,e} \right)$.

3. *Multiplier estimation using OLS:* We form $Z_t = y_{\Gamma t}$ and estimate $M = \frac{1}{1-\alpha\phi^d}$ and $\alpha M$ using OLS,

$$y_{St} = Z_t + \beta^y \eta_t^e + C_{St}^y \beta^{Cy} + a_S + M\varepsilon_t^y,$$

$$p_t = \alpha M Z_t + \beta^p \eta_t^e + C_t^p \beta^{Cp} + b + \varepsilon_t^p.$$

   The OLS standard errors on $M$ and $\alpha M$ are correct in case we only have parametric factors (see Proposition 4). In the case of nonparametric factors, the simulation results in Section 5 suggest that OLS standard errors remain reliable as well.

4. *Elasticity estimation using instrumental variables:* We estimate the elasticity $\alpha$ using instrumental variables, where we use $Z_t$ as an instrument for $y_{St}$ in (53). To estimate $\phi^d$, we consider the regression

$$y_{Et} = \phi^d p_t + mC_{Et}^y + \lambda_{Et}\eta_t^e + a_E + u_{Et}, \tag{55}$$

---

[33]In case of principal components, we estimate $(\check{\lambda}, \eta_t)$ using (with $\|a\| = \sqrt{a'a}$)

$$\min_{\check{\lambda}, \eta_t} \frac{1}{NT} \sum_t \left\| \check{y}_t^e - \check{\lambda}\eta_t \right\|^2,$$

subject to the normalization $\frac{1}{N}\check{\lambda}'\check{\lambda} = I$. The asymptotic theory for principal components analysis has been developed in Bai (2003) and relies on $N, T \to \infty$. Alternatively, one can rely on factor analysis in which $N$ is finite and $T \to \infty$, and the asymptotic theory is developed in Anderson and Amemiya (1988) under a wide class of distributional assumptions on the error terms and the factors. A variety of algorithms have been developed to extract latent factors in the presence of missing data.

and we use $Z_t$ as an instrument for $p_t$. In case of elasticity estimates, standard weak instrument concerns are relevant (Stock and Yogo (2005); Montiel Olea and Pflueger (2013); Andrews et al. (2019)).

After estimating the parameters of interest using GIV, we recommend exploring the specification tests in Section 2.6 to ensure that the procedure properly isolates idiosyncratic shocks. Depending on the data available, one can for instance perform a narrative check, an over-identifying restrictions test or explore the stability the estimates when adding additional factors.

## 4.2   User's guide: extensions

**Heteroskedasticity**   If there is heteroskedasticity, we compute

$$\tilde{E}_i = \frac{\bar{\sigma}_i^{-2}}{\sum_j \bar{\sigma}_j^{-2}},$$

where $\sigma_i^2 = \sigma_i^2\left(y_{it}\right)$ and $\bar{\sigma}_i = \max\left(\sigma_i, \kappa\times\text{median}\left(\sigma_j\right)\right)$, where $\kappa \in [0, 1]$. The adjustment using the max operator and the median ensures that very volatile entities do not distort the equal-weighted average, while not putting too large of a weight on entities with very low volatilities. When there is little dispersion, we do not adjust the inverse variance weights if $\kappa < 1$, say, $\kappa = 0.75$. In the presence of heteroskedasticity, we use $\tilde{E}_i$ in computing cross-sectional averages or in computing cross-sectional regressions, which improves efficiency.

**Robustness to large idiosyncratic shocks**   In case there are large idiosyncratic shocks, which is generally helpful in the context of GIV, this may influence the estimates of common factors, including the equal-weighted average. One approach is to adjust for heteroskedasticity, as discussed above. Alternatively, we can winsorize the data for the purposes of computing averages and factors. This extension is particularly relevant when the cross-section is small. If we denote the winsorized data by $y_{it}^w$, then in steps 1 and 2 of the user guide, we use $y_{it}^w$. In computing $Z_t$, we form $Z_t = y_{St} - y_{Et}^w$. In step 4, we use $y_{Et}^w$ in (55) as well.

**Autocorrelated shocks**   We have assumed that the idiosyncratic shocks $u_{it}$ are i.i.d. It is easy to form the GIV more generally: one replaces $u_{it}$ by $\tilde{u}_{it} := u_{it} - \mathbb{E}_{t-1}\left[u_{it}\right]$. For instance, suppose that the $u_{it}$ follow an autocorrelated process, $u_{it} = \rho u_{i,t-1} + \tilde{u}_{it}$, where $\tilde{u}_{it}$ is i.i.d. Then, we form the GIV based on the innovations $\tilde{u}_{it}$.

**Heterogeneous demands elasticities**   When demand elasticities are heterogeneous, we apply the procedure outlined in Section 2.5.

**At what level should one aggregate the data?** What's the impact of having aggregate data? Suppose that firm $i$ in industry $k$ has an action (e.g. demand for oil in our basic example) $y_{ikt} = \lambda_i \eta_t + u_{kt} + u_{ikt}$ where $u_{kt}$ is a industry-level common shock, and $u_{kit}$ is a firm-level shock. Then, aggregating at the industry $k$ level with weights $S_i^k$ (with $\sum_{i \in k} S_i^k = 1$), $y_{kt} = \sum_{i \in k} S_i^k y_{ikt} = \lambda^k \eta_t + u_{kt} + u_{k,S^k t}$ with $\lambda^k = \sum_{i \in k} S_i^k \lambda_i$ and $u_{k,S^k t} = \sum_{i \in k} S_i^k u_{ikt}$. So, the idiosyncratic industry level shock is $v_{kt} = u_{kt} + u_{k,S^k t}$: it is the sum of the primitive industry-level shock $u_{kt}$ and the sum of idiosyncratic firm-level shocks in industry $k$, $u_{k,S^k t}$. In general, it is of course better to have disaggregated data, in part because a higher number of units $N$ makes it easier to identify the common shock $\eta_t$, which leads to a higher-precision GIV (as we saw in our discussion of (23)). However, when the number of industries is large enough, there is little harm in having industry-level aggregation, as the marginal impact of having more industries is small.

# 5  Simulations

We use simulations to illustrate the precision of granularly identified parameters depending on the size of the sample (both $N$ and $T$), the degree of concentration, and the volatility of idiosyncratic shocks relative to aggregate shocks. Also, we show that PCA can be used to remove common factors and to estimate the idiosyncratic shocks.

## 5.1  Model

We start from the standard supply, $y_{it}^s$, and demand, $y_t^d$, model

$$y_{it}^s = \phi^s p_t + \lambda_i \eta_t + u_{it}, \qquad y_t^d = \phi^d p_t + \epsilon_t,$$

where $\phi^d < 0 < \phi^s$ and $\mathbb{E}[\eta_t u_{it}] = \mathbb{E}[\epsilon_t u_{it}] = \mathbb{E}[\eta_t \epsilon_t] = 0$. The model implies, with $M = -\frac{\phi^d}{\phi^s - \phi^d}$,

$$p_t = \frac{M}{\phi^d}\left(u_{St} + \lambda_S \eta_t - \epsilon_t\right), \qquad y_{St}^s = M u_{St} + M \lambda_S \eta_t + (1 - M)\epsilon_t.$$

## 5.2  Estimators and standard errors

To estimate $M$ and $\frac{M}{\phi^d}$, we can use standard OLS. To estimate $M$, we use

$$y_{St}^s = a + M y_{\Gamma t} + \theta' \eta_t^e + e_t, \tag{56}$$

and to estimate $\frac{M}{\phi^d}$, we use

$$p_t = a_p + \frac{M}{\phi^d} y_{\Gamma t} + \theta_p' \eta_t^e + e_t^p. \tag{57}$$

25

All standard OLS results apply if we observe the factors, $\eta_t$. However, we often do not directly observe all factors. We consider the cases in which we know the factor loadings, $\lambda_i^\eta$, and where the loadings are unobserved and estimated using PCA. To provide a point of reference, we also consider the case where we do not control for factors and impose that $\theta = \theta_p = 0$. In all cases, we report the OLS standard errors to assess to what extent the OLS standard errors need to be adjusted for the fact that we use $\eta_t^e$ instead of $\eta_t$. Note that we allow for a vector of estimated factors, $\eta_t^e$, as in practice we do not know the number of factors. To estimate the multiplier when using PCA, we estimate the number of common factors using the procedure in Bai and Ng (2002) by minimizing their $IC_{p2}(k)$ criterion, where $k \geq 1$. Also, we consider the possibility where we add factors extracted using principal components to factors extracted based on known loadings. The main conclusion is that adding additional factors does not distort the estimates in our simulations, while not properly controlling for factors can lead to biased estimates.

To estimate the demand and supply elasticities, we can recover them from the estimates of $M$ and $\frac{M}{\phi^d}$. However, as discussed before, this is equivalent to a 2SLS estimator using $y_{\Gamma t}$ as instrument for price, while controlling in this case for the factors. Hence, the first stage corresponds to

$$p_t = a_p + \xi y_{\Gamma t} + \theta_p' \eta_t^e + e_t,$$

and the second stage to estimate the demand elasticity is, with $\hat{p}_t = a_p^e + \xi^e y_{\Gamma t} + \theta_p' \eta_t^e$,

$$y_t^d = a_d + \phi^d \hat{p}_t + \theta_d' \eta_t^e + e_t^d,$$

and for the supply elasticity

$$y_{Et}^s = a_s + \phi^s \hat{p}_t + \theta_s' \eta_t^e + e_t^s.$$

The standard weak instrument tests can be used to assess whether $y_{\Gamma t}$ is a sufficiently strong instrument for price (Section 2.2). In this case, we report the 2SLS standard errors to assess whether their accuracy is impacted by the fact that we estimate the common factors.

## 5.3  Calibration

In calibrating the model, we target (i) concentration, as measured by the excess Herfindahl, $h = \sqrt{\sum_i S_i^2 - 1/N}$, and (ii) the ratio of the volatility of idiosyncratic shocks to the volatility of aggregate supply shocks.

We set the demand and supply elasticities to $\phi^d = -0.3$ and $\phi^s = 0.1$, respectively. The size weights are generated as $k_i = i^{-1/\zeta}$, $S_i = k_i / \sum_i k_i$, where $\zeta$ is chosen so that $h \in \{0.2, 0.3\}$.[34] For the shocks, we assume $\epsilon_t \sim N(0, 0.06^2)$ and $\eta_t \sim N(0, 1)$. To construct the loadings, we first draw $\bar{\lambda} \in \mathbb{R}^N$ from a uniform distribution, $\bar{\lambda} \sim U(0, 1)$. We scale $\bar{\lambda}$ so that the standard deviation of

---

[34]Here $\zeta$ is the power law exponent of the size distribution, see Gabaix (2009).

Table 1: Cases considered in simulations. We calibrate the supply-and-demand model in Section 5 under seven alternative parameterizations. The parameters are the following: $N$ is the cross-sectional sample size; $T$ is the number of simulated i.i.d. time periods; $h$ is the excess Herfindahl that we target in our simulation of the size weights (as described in Section 5.3); $\tau$ is the targeted ratio of the volatility of idiosyncratic shocks to the volatility of aggregate supply shocks; the multipliers $M = -\frac{\phi^d}{\phi^s - \phi^d}$ and $\frac{M}{\phi^d}$ are functions of the elasticities $\phi^d$ and $\phi^s$ of demand and supply with respect to price. The final column reports the share of the price volatility that is due to idiosyncratic shocks under each of the seven parameterizations.

| Case | $N$ | $T$ | $h$ | $\tau$ | $M$ | $\frac{M}{\phi^d}$ | % price vol. idiosyncratic |
|------|-----|-----|-----|--------|-----|--------------------|----------------------------|
| 1 | 25 | 360 | 0.2 | 3 | 0.75 | -2.5 | 12.6% |
| 2 | 25 | 360 | 0.2 | 4 | 0.75 | -2.5 | 20.4% |
| 3 | 25 | 360 | 0.3 | 3 | 0.75 | -2.5 | 19.1% |
| 4 | 25 | 360 | 0.3 | 4 | 0.75 | -2.5 | 29.5% |
| 5 | 25 | 120 | 0.2 | 4 | 0.75 | -2.5 | 20.4% |
| 6 | 50 | 120 | 0.2 | 4 | 0.75 | -2.5 | 16.1% |
| 7 | 50 | 360 | 0.2 | 4 | 0.75 | -2.5 | 16.1% |

the aggregate supply shock equals 3%, $V(\lambda_S \eta_t) = \lambda_S^2 = 0.03^2$, that is, $\lambda = \frac{0.03}{\lambda_S} \bar{\lambda}$. We consider two cases, namely where $Corr(\lambda, S) = 0$ and $Corr(\lambda, S) = -20\%$.[35] Lastly, we set $\sigma_u = \lambda_S \tau = 0.03\tau$ to target the ratio $\tau$ of idiosyncratic shock volatility to aggregate shock volatility. We vary $\tau \in \{3, 4\}$, $N \in \{25, 50\}$, and $T \in \{120, 360\}$.

The cases considered are summarized in Table 1. The final column reports the fraction of price volatility that is due to idiosyncratic shocks, which ranges approximately from 10% to 30%, in line with the recent literature on granularity that estimates the fraction of aggregate fluctuations that can be traced back to idiosyncratic shocks.

## 5.4 Simulation results

The simulation results when $Corr(\lambda, S) = 0$ are reported in Table 2. We consider four estimators. In the case of M1, we assume that the loadings are known in estimating the factors; this is an ideal case taken as a benchmark. In the case of M2, we use PCA to estimate the factors. In the case of M3, we control for the factors estimated using the known loadings and PCA. This case is helpful to see whether there is harm in the simulation from controlling for too many factors. In the case of M4, we use no factors and just use $y_{\Gamma t}^s$ without any factors. Note that we do not advocate M4 in practice: M4 simply shows that the GIV estimator can be biased if we do not properly control for

---

[35]Formally, we start from $\bar{\lambda} \sim U(0, 1)$. We then estimate the regression $\bar{\lambda} = a\iota_N + \bar{b}(S - \bar{S}) + e$ and find $b$ such that $\tilde{\lambda} = a\iota_N + bS + e$ satisfies $Corr(\tilde{\lambda}, S) = \rho$, the target correlation. This correlation equals $\rho = \frac{bV(S)}{\sigma(S)\sigma(\tilde{\lambda})} = \frac{bV(S)}{\sigma(S)\sqrt{b^2 V(S) + V(e)}}$, which provides the solution $b = Sign(\rho)\sqrt{\frac{\rho^2}{1 - \rho^2} \frac{V(e)}{V(S)}}$. We then scale $\tilde{\lambda}$ to obtain the right volatility of aggregate supply shocks, $\lambda = \frac{0.03}{\lambda_S} \tilde{\lambda}$.

Table 2: Simulation results when $Corr(\lambda, S) = 0$ based on 10,000 replications. The parameters used in the different cases are summarized in Table 1. In particular, the data are generated from a model in which $M = 0.75$, $\frac{M}{\phi^d} = -2.5$, $\phi^d = -0.3$, and $\phi^s = 0.1$. GIV estimators M1,..., M4 are described at the beginning of Section 5.4. For each estimator, we report the median, the mean, and percentiles 2.5% (P2.5) and 97.5% (P97.5) in the simulated distribution of estimates. "Coverage" is the fraction of estimates falling within the 95% confidence intervals constructed using OLS standard errors (columns 1 through 8) or the 2SLS standard errors (columns 9 through 16).

| Case | Statistic | $M$ M1 | M2 | M3 | M4 | $\frac{M}{\phi^d}$ M1 | M2 | M3 | M4 | $\phi^d$ M1 | M2 | M3 | M4 | $\phi^s$ M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.49 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.50 | -0.33 | -0.34 | -0.32 | -0.34 | 0.11 | 0.12 | 0.11 | 0.11 |
| 1 | P2.5 | 0.53 | 0.49 | 0.52 | 0.51 | -3.81 | -3.90 | -3.86 | -3.83 | -0.65 | -0.67 | -0.67 | -0.65 | 0.01 | 0.00 | 0.00 | 0.00 |
| | P97.5 | 0.98 | 1.01 | 0.99 | 0.99 | -1.20 | -1.11 | -1.15 | -1.17 | -0.17 | -0.17 | -0.17 | -0.17 | 0.27 | 0.32 | 0.28 | 0.30 |
| | Coverage | 0.95 | 0.93 | 0.95 | 0.95 | 0.95 | 0.94 | 0.95 | 0.95 | 0.93 | 0.93 | 0.93 | 0.93 | 0.96 | 0.94 | 0.95 | 0.96 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.50 | -2.51 | -2.51 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.51 | -2.51 | -2.51 | -0.31 | -0.31 | -0.32 | -0.31 | 0.10 | 0.11 | 0.10 | 0.10 |
| 2 | P2.5 | 0.60 | 0.56 | 0.59 | 0.58 | -3.48 | -3.54 | -3.50 | -3.52 | -0.52 | -0.52 | -0.53 | -0.52 | 0.04 | 0.02 | 0.04 | 0.03 |
| | P97.5 | 0.90 | 0.94 | 0.90 | 0.92 | -1.54 | -1.47 | -1.51 | -1.49 | -0.20 | -0.19 | -0.19 | -0.19 | 0.19 | 0.22 | 0.19 | 0.21 |
| | Coverage | 0.95 | 0.90 | 0.95 | 0.95 | 0.95 | 0.93 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 | 0.94 | 0.96 | 0.91 | 0.96 | 0.95 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.50 | -2.51 | -2.51 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.51 | -2.51 | -2.51 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.10 | 0.10 | 0.10 |
| 3 | P2.5 | 0.62 | 0.59 | 0.62 | 0.61 | -3.27 | -3.34 | -3.29 | -3.29 | -0.44 | -0.45 | -0.45 | -0.44 | 0.05 | 0.03 | 0.04 | 0.04 |
| | P97.5 | 0.88 | 0.92 | 0.88 | 0.89 | -1.74 | -1.68 | -1.72 | -1.72 | -0.22 | -0.22 | -0.21 | -0.22 | 0.17 | 0.21 | 0.18 | 0.18 |
| | Coverage | 0.95 | 0.90 | 0.95 | 0.95 | 0.94 | 0.93 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.95 | 0.95 | 0.90 | 0.95 | 0.95 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.51 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.50 | -2.51 | -2.51 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.10 | 0.10 | 0.10 |
| 4 | P2.5 | 0.65 | 0.63 | 0.65 | 0.64 | -3.15 | -3.18 | -3.17 | -3.18 | -0.42 | -0.42 | -0.42 | -0.42 | 0.06 | 0.05 | 0.06 | 0.05 |
| | P97.5 | 0.85 | 0.87 | 0.85 | 0.87 | -1.86 | -1.81 | -1.84 | -1.84 | -0.22 | -0.22 | -0.22 | -0.22 | 0.15 | 0.17 | 0.15 | 0.16 |
| | Coverage | 0.95 | 0.90 | 0.95 | 0.95 | 0.94 | 0.93 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.89 | 0.95 | 0.95 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.51 | -2.50 | -2.50 | -2.50 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.50 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| 5 | P2.5 | 0.66 | 0.64 | 0.66 | 0.66 | -3.01 | -3.05 | -3.02 | -3.02 | -0.38 | -0.38 | -0.39 | -0.38 | 0.06 | 0.05 | 0.06 | 0.06 |
| | P97.5 | 0.84 | 0.86 | 0.84 | 0.84 | -1.99 | -1.95 | -1.99 | -1.98 | -0.24 | -0.24 | -0.24 | -0.24 | 0.14 | 0.16 | 0.15 | 0.15 |
| | Coverage | 0.95 | 0.90 | 0.95 | 0.95 | 0.94 | 0.93 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.89 | 0.95 | 0.95 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.50 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.49 | -2.50 | -2.50 | -0.32 | -0.33 | -0.33 | -0.33 | 0.11 | 0.11 | 0.11 | 0.11 |
| 6 | P2.5 | 0.56 | 0.52 | 0.55 | 0.52 | -3.71 | -3.78 | -3.74 | -3.77 | -0.64 | -0.66 | -0.65 | -0.64 | 0.02 | 0.01 | 0.02 | 0.01 |
| | P97.5 | 0.94 | 0.98 | 0.95 | 0.97 | -1.24 | -1.16 | -1.24 | -1.20 | -0.17 | -0.17 | -0.17 | -0.17 | 0.22 | 0.26 | 0.22 | 0.26 |
| | Coverage | 0.95 | 0.93 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.93 | 0.93 | 0.93 | 0.93 | 0.96 | 0.95 | 0.96 | 0.96 |
| | Median | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.50 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.10 |
| | Mean | 0.75 | 0.75 | 0.75 | 0.75 | -2.50 | -2.50 | -2.50 | -2.50 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.10 | 0.10 | 0.10 |
| 7 | P2.5 | 0.64 | 0.62 | 0.64 | 0.62 | -3.21 | -3.27 | -3.23 | -3.24 | -0.44 | -0.44 | -0.45 | -0.44 | 0.06 | 0.05 | 0.06 | 0.05 |
| | P97.5 | 0.86 | 0.88 | 0.86 | 0.88 | -1.78 | -1.75 | -1.77 | -1.75 | -0.22 | -0.22 | -0.22 | -0.22 | 0.15 | 0.17 | 0.16 | 0.17 |
| | Coverage | 0.95 | 0.91 | 0.95 | 0.95 | 0.95 | 0.94 | 0.95 | 0.95 | 0.94 | 0.95 | 0.95 | 0.95 | 0.95 | 0.91 | 0.95 | 0.95 |

Table 3: Simulation results when $Corr\,(\lambda, S) = -20\%$ based on 10,000 replications. The parameters used in the different cases are summarized in Table 1. In particular, the data are generated from a model in which $M = 0.75$, $\frac{M}{\phi^d} = -2.5$, $\phi^d = -0.3$, and $\phi^s = 0.1$. GIV estimators M1,..., M4 are described at the beginning of Section 5.4. For each estimator, we report the median, the mean, and percentiles 2.5% (P2.5) and 97.5% (P97.5) in the simulated distribution of estimates. "Coverage" is the fraction of estimates falling within the 95% confidence intervals constructed using OLS standard errors (columns 1 through 8) or the 2SLS standard errors (columns 9 through 16).

| Case | Statistic | M ||| | $\frac{M}{\phi^d}$ |||| $\phi^d$ |||| $\phi^s$ ||||
| | | M1 | M2 | M3 | M4 | M1 | M2 | M3 | M4 | M1 | M2 | M3 | M4 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Median | 0.75 | 0.69 | 0.75 | 0.56 | -2.49 | -2.29 | -2.50 | -1.85 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.13 | 0.10 | 0.24 |
| | Mean | 0.75 | 0.69 | 0.75 | 0.56 | -2.50 | -2.30 | -2.50 | -1.85 | -0.33 | -0.34 | -0.33 | -0.30 | 0.11 | 0.16 | 0.11 | 0.21 |
| 1 | P2.5 | 0.52 | 0.41 | 0.51 | 0.29 | -3.84 | -3.72 | -3.87 | -3.23 | -0.66 | -0.74 | -0.67 | -0.97 | 0.01 | 0.01 | 0.00 | 0.07 |
| | P97.5 | 0.98 | 0.97 | 0.99 | 0.82 | -1.17 | -0.88 | -1.13 | -0.50 | -0.17 | -0.16 | -0.17 | -0.14 | 0.28 | 0.47 | 0.29 | 0.96 |
| | Coverage | 0.95 | 0.88 | 0.95 | 0.67 | 0.95 | 0.93 | 0.95 | 0.84 | 0.93 | 0.93 | 0.93 | 0.92 | 0.96 | 0.97 | 0.95 | 1.00 |
| | Median | 0.75 | 0.74 | 0.75 | 0.42 | -2.51 | -2.46 | -2.51 | -1.38 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.11 | 0.10 | 0.42 |
| | Mean | 0.75 | 0.74 | 0.75 | 0.41 | -2.51 | -2.47 | -2.51 | -1.39 | -0.31 | -0.32 | -0.32 | 0.00 | 0.10 | 0.11 | 0.10 | -0.12 |
| 2 | P2.5 | 0.60 | 0.56 | 0.59 | 0.19 | -3.51 | -3.50 | -3.54 | -2.48 | -0.52 | -0.53 | -0.54 | -1.05 | 0.04 | 0.03 | 0.04 | 0.17 |
| | P97.5 | 0.90 | 0.92 | 0.91 | 0.63 | -1.52 | -1.43 | -1.49 | -0.30 | -0.19 | -0.19 | -0.19 | -0.12 | 0.19 | 0.23 | 0.19 | 1.86 |
| | Coverage | 0.95 | 0.91 | 0.95 | 0.08 | 0.94 | 0.94 | 0.95 | 0.41 | 0.93 | 0.93 | 0.94 | 0.92 | 0.96 | 0.94 | 0.96 | 0.72 |
| | Median | 0.75 | 0.73 | 0.75 | 0.56 | -2.51 | -2.43 | -2.51 | -1.86 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.11 | 0.10 | 0.24 |
| | Mean | 0.75 | 0.73 | 0.75 | 0.56 | -2.51 | -2.43 | -2.51 | -1.86 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.12 | 0.10 | 0.26 |
| 3 | P2.5 | 0.62 | 0.56 | 0.61 | 0.39 | -3.29 | -3.26 | -3.31 | -2.68 | -0.44 | -0.45 | -0.45 | -0.52 | 0.04 | 0.04 | 0.04 | 0.12 |
| | P97.5 | 0.88 | 0.89 | 0.89 | 0.72 | -1.73 | -1.59 | -1.69 | -1.02 | -0.22 | -0.21 | -0.21 | -0.19 | 0.18 | 0.23 | 0.18 | 0.51 |
| | Coverage | 0.95 | 0.88 | 0.95 | 0.29 | 0.94 | 0.93 | 0.94 | 0.62 | 0.94 | 0.94 | 0.94 | 0.94 | 0.95 | 0.92 | 0.95 | 0.55 |
| | Median | 0.75 | 0.74 | 0.75 | 0.51 | -2.50 | -2.48 | -2.51 | -1.69 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.10 | 0.10 | 0.29 |
| | Mean | 0.75 | 0.74 | 0.75 | 0.51 | -2.51 | -2.48 | -2.51 | -1.69 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.11 | 0.10 | 0.32 |
| 4 | P2.5 | 0.65 | 0.62 | 0.64 | 0.34 | -3.17 | -3.16 | -3.19 | -2.44 | -0.42 | -0.42 | -0.43 | -0.51 | 0.06 | 0.05 | 0.06 | 0.16 |
| | P97.5 | 0.85 | 0.86 | 0.86 | 0.66 | -1.85 | -1.79 | -1.82 | -0.93 | -0.22 | -0.22 | -0.22 | -0.20 | 0.15 | 0.18 | 0.16 | 0.64 |
| | Coverage | 0.95 | 0.91 | 0.95 | 0.07 | 0.94 | 0.94 | 0.95 | 0.36 | 0.94 | 0.95 | 0.95 | 0.94 | 0.95 | 0.92 | 0.95 | 0.11 |
| | Median | 0.75 | 0.74 | 0.75 | 0.61 | -2.50 | -2.45 | -2.50 | -2.03 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.11 | 0.10 | 0.19 |
| | Mean | 0.75 | 0.74 | 0.75 | 0.61 | -2.51 | -2.46 | -2.51 | -2.03 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.11 | 0.10 | 0.20 |
| 5 | P2.5 | 0.66 | 0.62 | 0.66 | 0.49 | -3.03 | -3.02 | -3.04 | -2.59 | -0.39 | -0.39 | -0.39 | -0.41 | 0.06 | 0.06 | 0.06 | 0.12 |
| | P97.5 | 0.84 | 0.85 | 0.84 | 0.72 | -1.98 | -1.89 | -1.96 | -1.46 | -0.24 | -0.24 | -0.24 | -0.23 | 0.15 | 0.18 | 0.15 | 0.32 |
| | Coverage | 0.95 | 0.88 | 0.95 | 0.25 | 0.95 | 0.93 | 0.95 | 0.57 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.89 | 0.95 | 0.29 |
| | Median | 0.75 | 0.71 | 0.75 | 0.46 | -2.49 | -2.36 | -2.50 | -1.53 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.12 | 0.10 | 0.35 |
| | Mean | 0.75 | 0.71 | 0.75 | 0.46 | -2.50 | -2.35 | -2.50 | -1.53 | -0.33 | -0.34 | -0.33 | -0.25 | 0.11 | 0.14 | 0.11 | 0.27 |
| 6 | P2.5 | 0.55 | 0.47 | 0.55 | 0.21 | -3.73 | -3.64 | -3.76 | -2.84 | -0.65 | -0.71 | -0.67 | -1.27 | 0.02 | 0.03 | 0.02 | 0.12 |
| | P97.5 | 0.95 | 0.94 | 0.95 | 0.71 | -1.22 | -1.00 | -1.21 | -0.21 | -0.17 | -0.16 | -0.17 | -0.10 | 0.22 | 0.34 | 0.23 | 1.84 |
| | Coverage | 0.95 | 0.90 | 0.95 | 0.35 | 0.95 | 0.94 | 0.95 | 0.68 | 0.93 | 0.93 | 0.93 | 0.91 | 0.96 | 0.97 | 0.96 | 0.97 |
| | Median | 0.75 | 0.74 | 0.75 | 0.46 | -2.50 | -2.45 | -2.50 | -1.54 | -0.30 | -0.30 | -0.30 | -0.30 | 0.10 | 0.11 | 0.10 | 0.35 |
| | Mean | 0.75 | 0.74 | 0.75 | 0.46 | -2.50 | -2.45 | -2.50 | -1.53 | -0.31 | -0.31 | -0.31 | -0.31 | 0.10 | 0.11 | 0.10 | 0.35 |
| 7 | P2.5 | 0.64 | 0.61 | 0.64 | 0.30 | -3.23 | -3.21 | -3.24 | -2.31 | -0.45 | -0.45 | -0.45 | -0.61 | 0.05 | 0.05 | 0.05 | 0.19 |
| | P97.5 | 0.86 | 0.87 | 0.87 | 0.62 | -1.77 | -1.70 | -1.76 | -0.74 | -0.22 | -0.21 | -0.21 | -0.18 | 0.16 | 0.18 | 0.16 | 0.82 |
| | Coverage | 0.95 | 0.91 | 0.95 | 0.03 | 0.95 | 0.94 | 0.95 | 0.29 | 0.95 | 0.94 | 0.95 | 0.93 | 0.96 | 0.93 | 0.95 | 0.16 |

aggregate factors. The first four columns correspond to the estimates of $M$, the next four columns to estimates of $\frac{M}{\phi^d}$, the next four columns to estimates of $\phi^d$, and the last four columns to estimates of $\phi^s$.

For each of the estimators, we report the median, the mean, and the 2.5% and 97.5% percentiles. For each estimator, we also report its 95% coverage rate using OLS standard errors (columns 1 to 8) or the 2SLS standard errors (columns 9 to 16).

As is clear from all cases, the estimators are mean- and median-unbiased in the simulation. Moreover, confidence intervals tighten when concentration increases (case 3 relative to case 1 and case 4 relative to case 2) and when the volatility of idiosyncratic shocks increases (case 2 relative to case 1 and case 4 relative to case 3). Naturally, the confidence interval tightens when we increase $N$ and $T$. The coverage is generally accurate and OLS standard errors only slightly overstate the precision in the case of M2 in estimating $M$; the 2SLS standard errors are somewhat small in small samples in estimating $\phi^s$.

It is tempting to conclude that using $y^s_{\Gamma t}$ as instrument, even without estimating the factors, results in accurate and unbiased estimates of the parameters of interest. However, this is only the case when $Corr\left(\lambda, S\right) = 0$. To illustrate this, we consider a negative correlation between size and exposures, $Corr\left(\lambda, S\right) = -20\%$.

The results are presented in Table 3. We now find a large bias in the case of M4, both in terms of the mean and median. The coverage estimates are also heavily distorted. Intuitively, $y^s_{\Gamma t}$ does not filter out aggregate shocks, and the exogeneity restriction is violated. This is why factor estimates are required when loadings may be correlated with size. Even in the case where we have no information about factor loadings (in the case of M2, which relies only on PCA), accounting for common factors removes most of the bias and leads to much improved coverage estimates. When we know the factor loadings (in case of M1), there is no bias and the coverage estimates are accurate. In addition, combining the PCA estimate and the estimate using the known loadings results in almost the same accuracy as M1. This simulation illustrates the importance of accounting for factors in using GIV when loadings correlate with size.

# 6    Estimating sovereign yield spillovers

We study spillovers in sovereign yield markets in the Euro area as a first application of GIVs. We focus on the transmission and amplification of idiosyncratic shocks during the European sovereign debt crisis.

## 6.1 Data

We use daily data on 10-year zero coupon yields from Bloomberg.[36] The countries and Bloomberg tickers that we use are listed in Table F.7 in Appendix F.4.[37] We use data on general government gross debt for each country from Eurostat.[38] The sample is from July 2009 to May 2018.[39]

## 6.2 An empirical model of sovereign yield spillovers

Section F provides a fleshed-out 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 and vice versa.[40] We index countries by $i$. We define the yield spread, $y_{it}$, as the yield in country $i$ relative to Germany's yield. The model implies that relative changes in yield spreads, $r_{it} := \frac{\Delta y_{it}}{y_{i,t-1}}$, satisfy the following empirical model

$$r_{it} = \gamma r_{St} + \lambda_i' \eta_t + u_{it}, \tag{58}$$

where the size weights are

$$S_{i,t-1} = \frac{B_{i,t-1} y_{i,t-1}}{\sum_j B_{j,t-1} y_{j,t-1}}, \tag{59}$$

with $B_{it}$ the outstanding government debt of country $i$.

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.[41] Second, the spillover impact is such that $\frac{\Delta y_{it}}{y_{i,t-1}}$, rather than $\Delta y_{it}$, depends linearly on $\gamma r_{St}$. This means that a country with almost no default risk should have almost no sensitivity of its yield $\Delta y_{it}$, 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.[42]

The interpretation of $\gamma$ and $M = \frac{1}{1-\gamma}$ is that if a country suffers 1 billion euros in losses on its debt because of some idiosyncratic bad news, then the aggregate debt of all European governments

---

falls by $M = \frac{1}{1-\gamma}$ billions.[43]

It is essential to control for common factors $\eta_t$. It is well understood, see for instance Forbes and Rigobon (2002), that omitted factors and endogeneity impact measures of spillovers and contagion.[44]

## 6.3 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 get close to zero, $y_{it} \simeq 0$.

1. We compute the rolling standard deviation of relative changes in yield spreads using the trading days of the last two months, $\sigma_t(r_{it})$. We then define[45]

$$\sigma_{it} = \max\left(\sigma_t(r_{it}), m_t\right), \tag{60}$$

where $m_t = median\left(\sigma_t(r_{jt})\right)$, that is, the cross-sectional median at time $t$. We define the $\tilde{E}$-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 (60) to avoid that the $\tilde{E}$-weights 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.[46]

2. We compute $\check{r}_{it} = r_{it} - r_{\tilde{E}t}$ and adjust $\check{r}_{it}$ for heteroskedasticity, $n_{it} = \frac{\check{r}_{it}}{\sigma_{it}}$. We use PCA based on $n_{it}$, $n_{it} = \check{\lambda}'_i \eta_t + \check{u}_{it}$, to estimate the factors, $\eta_t^e$.

3. We estimate the multiplier $M = \frac{1}{1-\gamma}$ via the regression

$$r_{St} = k + M r_{\tilde{\Gamma}t} + \lambda'_S \eta_t^e + e_t. \tag{61}$$

To identify the largest shocks and to verify narratively that the shocks are truly idiosyncratic, we run the weighted panel regression

$$r_{it} - r_{\tilde{E}t} = c + \lambda' \eta_t^e + u_{it},$$

---

[43]Indeed, suppose that country $i$ 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

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

[44]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 $\gamma$ in equation (58)) is very different from theirs.

[45]We compute the rolling standard deviation including the current observation. Given the objective to most accurately estimate the time fixed effect and the factor realizations, we want to account for the same-period realizations, in particular given the fat-tailed distribution of yield changes.

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

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 2009 to May 2018. The spreads are based on 10-year zero-coupon bonds and are constructed using data from Bloomberg. The right panel displays the size weights based on the definition in (59) for the same countries and the same sample period.
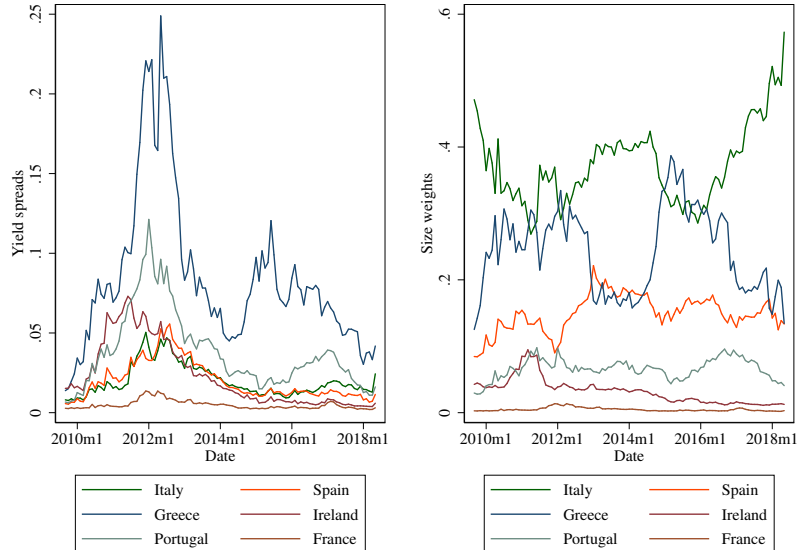


using the size weights. In this panel regression, $u_{St}^e$ is identical to the residual of the regression $r_{\tilde{\Gamma}t} = c + \lambda'_{\hat{\Gamma}}\eta_t^e + u_{St}^e$. We discuss the largest $|u_{St}^e|$ in detail in Section 6.6.

## 6.4 Empirical results

We plot the dynamics of spreads, $y_{it}$, in the left panel, and size weights, using the definition in (59), in the right panel of Figure 1 for France, Greece, Ireland, Italy, Portugal, and Spain. The sample is from September 2009 to May 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 Section 6.6.

Table 4 reports the estimates of the multiplier, $M$. The first column regresses $r_{St}$ on $Z_t = r_{\tilde{\Gamma}t}$. The second to the fourth column add principal components. The multiplier estimate drops after adding the first principal component from 1.63 to 1.46, but then stabilizes and adding more principal components does not impact the estimate of the multiplier in an economically meaningful way. In

Table 4: Multiplier estimates of sovereign yield spillovers. The table reports the estimates of the multiplier in (61). The first to the fourth column include zero to three principal components as controls. In the final column, we re-estimate the model excluding Greece. In this table, $Z_t = r_{\tilde{\Gamma}t}$ and $PC_{it}$ corresponds to the $i$th principal component that we extract based on $\frac{\tilde{r}_{it}}{\sigma_{it}}$. The size-weighted average of relative yield spread changes, $r_{St}$, uses the size weights as defined in (59). The model is estimated using daily data from July 2009 until May 2018.

|  | $r_{St}$ | $r_{St}$ | $r_{St}$ | $r_{St}$ | $r_{St}$ (excluding Greece) |
|---|---|---|---|---|---|
| $Z_t$ | 1.632 | 1.456 | 1.433 | 1.488 | 1.426 |
|  | (73.90) | (44.09) | (45.42) | (46.48) | (29.68) |
| $PC_{1t}$ |  | 0.00215 | 0.00230 | 0.00192 | 0.00215 |
|  |  | (7.13) | (8.00) | (6.67) | (5.47) |
| $PC_{2t}$ |  |  | -0.00332 | -0.00330 | -0.00160 |
|  |  |  | (-14.85) | (-14.95) | (-6.35) |
| $PC_{3t}$ |  |  |  | 0.00193 | 0.00249 |
|  |  |  |  | (7.53) | (6.50) |
| $N$ | 2264 | 2264 | 2264 | 2264 | 2264 |
| $R^2$ | 0.707 | 0.714 | 0.739 | 0.745 | 0.752 |

$t$ statistics in parentheses

the final column, we omit Greece, which plays an important role during this period. However, using the shocks from other countries does not impact the estimates in an economically meaningful way.

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 $\eta_t^e$. 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 4. The R-squared of this regression is 24%, implying that a quarter of the variation in aggregate yield spread changes is due to idiosyncratic shocks.

The idiosyncratic shocks to relative changes in yield spreads are fat-tailed, as can be seen from the left panel of Figure 2, which plots the time series of $u_{St}^e$. The right panel of the same figure plots $u_{St}^e$ (horizontal axis) against $r_{St}$ (vertical axis). If there are no spillovers, $\gamma = 0$, then the multiplier is one and the points fall along the 45-degree line (the red dashed line). The fact that the estimated slope is steeper, as indicated by the blue solid line, implies that there are significant yield spillovers.

Figure 2: Idiosyncratic shocks over time and aggregate yield changes. The figure shows the time-series dynamics of $u_{St}^e$ in the left panel. We construct $u_{St}^e$ as the residual of a regression of $r_{\tilde{\Gamma}t}$ on $\eta_t^e$. The right panel shows a scatter plot of $u_{St}^e$ (horizontal axis) against $r_{St}$ (vertical axis). The size-weighted average of relative yield spread changes, $r_{St}$, uses the size weights as defined in (59). The series are constructed using daily data from July 2009 until May 2018.
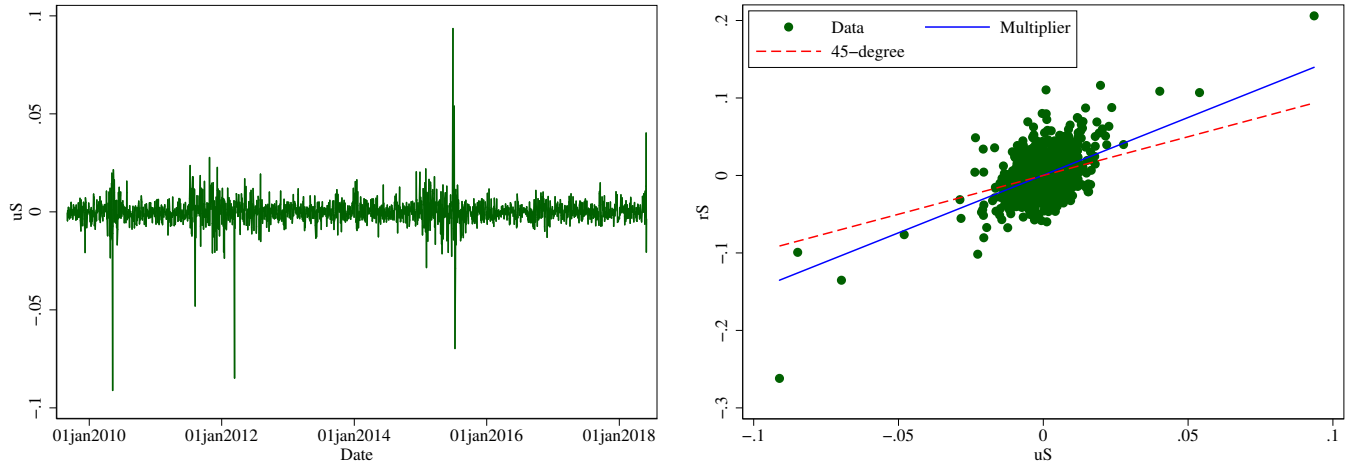


## 6.5  Interpretation of the coefficients

We find a multiplier $M = \frac{1}{1-\gamma} \simeq 1.5$ and hence a spillover parameter $\gamma \simeq \frac{1}{3}$.[47] The interpretation is as follows: 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. 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.[48]

## 6.6  Narrative GIVs

To further inspect the variation that the GIVs are exploiting to estimate the multiplier, we narratively check the largest shocks in Table 5. In particular, we order the dates based on the size of $|u_{St}^e|$. To illustrate the relevance of the largest shocks, we re-estimate the model that includes three principal components, that is, the fourth column of Table 4, using only the days with the largest $k$ shocks. In Figure 3, we show the multiplier estimate, alongside the 95%-confidence interval, where

---

[47]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 $r_{St} = M u_{St}$.

[48]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 double ($u_i = 1$ for $i =$ Italy); that is, the Italian yield spread goes from 2% to 4%. That makes the other yields 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.

Table 5: Summary of the largest idiosyncratic shocks and narratives. The table reports the properties of yield spread changes, $\Delta y_{it}$, on the 10 days with the largest realization of $|u^e_{St}|$ in Panel A. In Panel B, we report the corresponding relative yield changes, $r_{it} = \frac{\Delta y_{it}}{0.01+y_{i,t-1}}$. In Panel C, we provide the narratives associated with these events. The series are constructed using daily data from July 2009 until May 2018.

Panel A: Unscaled idiosyncratic shocks

| Date | Austria | Belgium | Finland | France | Greece | Ireland | Italy | Netherlands | Portugal | Slovenia | Spain |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10-May-10 | -0.08% | -0.24% | -0.07% | -0.08% | -3.87% | -1.18% | -0.43% | -0.06% | -1.62% | -0.30% | -0.62% |
| 8-Aug-11 | -0.03% | -0.09% | 0.00% | 0.08% | 0.01% | 0.05% | -0.66% | 0.00% | -0.10% | -0.16% | -0.73% |
| 26-Oct-11 | -0.05% | -0.03% | -0.03% | -0.09% | 3.11% | 0.01% | -0.02% | -0.03% | -0.15% | 0.01% | -0.04% |
| 12-Mar-12 | 0.04% | 0.04% | 0.01% | 0.05% | -7.32% | 0.00% | 0.11% | 0.01% | 0.06% | -0.04% | 0.09% |
| 3-Feb-15 | -0.01% | -0.01% | -0.01% | 0.00% | -1.21% | -0.02% | -0.07% | -0.01% | -0.12% | -0.06% | -0.05% |
| 29-Jun-15 | 0.05% | 0.07% | 0.03% | 0.07% | 3.21% | 0.13% | 0.36% | 0.04% | 0.48% | 0.11% | 0.36% |
| 6-Jul-15 | 0.02% | 0.04% | 0.01% | 0.03% | 2.35% | 0.06% | 0.16% | 0.01% | 0.27% | 0.03% | 0.18% |
| 10-Jul-15 | -0.04% | -0.05% | -0.03% | -0.05% | -3.28% | -0.10% | -0.21% | -0.04% | -0.24% | -0.23% | -0.22% |
| 13-Jul-15 | -0.01% | 0.01% | 0.00% | 0.01% | -1.11% | 0.00% | 0.02% | 0.00% | -0.03% | -0.01% | 0.02% |
| 29-May-18 | 0.07% | 0.04% | 0.03% | 0.04% | 0.36% | 0.08% | 0.49% | 0.03% | 0.20% | 0.08% | 0.17% |

Panel B: Scaled idiosyncratic shocks

| Date | Austria | Belgium | Finland | France | Greece | Ireland | Italy | Netherlands | Portugal | Slovenia | Spain |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10-May-10 | -5.2% | -14.2% | -5.2% | -5.9% | -41.6% | -29.5% | -17.5% | -4.8% | -38.0% | -12.8% | -24.1% |
| 8-Aug-11 | -1.9% | -3.1% | 0.2% | 4.4% | 0.1% | 0.6% | -14.4% | -0.3% | -1.2% | -4.0% | -15.9% |
| 26-Oct-11 | -2.6% | -1.0% | -2.1% | -4.3% | 19.4% | 0.2% | -0.5% | -2.3% | -1.5% | 0.3% | -0.8% |
| 12-Mar-12 | 1.9% | 1.6% | 0.8% | 2.5% | -31.7% | 0.0% | 2.8% | 0.9% | 0.5% | -0.9% | 2.2% |
| 3-Feb-15 | -1.2% | -0.7% | -0.7% | -0.1% | -11.5% | -1.2% | -3.0% | -0.7% | -3.7% | -2.9% | -2.0% |
| 29-Jun-15 | 3.8% | 5.0% | 2.8% | 5.1% | 33.4% | 6.9% | 16.6% | 3.1% | 17.4% | 4.7% | 16.6% |
| 6-Jul-15 | 1.8% | 2.5% | 0.7% | 2.0% | 18.9% | 3.3% | 6.8% | 1.2% | 8.6% | 1.1% | 7.7% |
| 10-Jul-15 | -3.0% | -3.5% | -2.2% | -3.7% | -22.1% | -5.0% | -8.8% | -3.0% | -7.6% | -9.1% | -9.0% |
| 13-Jul-15 | -0.5% | 1.0% | -0.2% | 1.1% | -9.6% | 0.1% | 1.0% | 0.3% | -0.9% | -0.5% | 1.1% |
| 29-May-18 | 5.2% | 2.6% | 2.3% | 2.9% | 7.2% | 5.0% | 14.6% | 2.3% | 7.3% | 4.7% | 7.7% |

Panel C: Narrative analysis

| Date | Event |
|---|---|
| 10-May-10 | Stock markets leap across Europe as EUR750bn eurozone rescue package is agreed |
| 8-Aug-11 | ECB decides to start buying Italian and Spanish bonds as part of the Securities Markets Program |
| 26-Oct-11 | EU leaders announced an agreement, including deal with private sector investors to take a 50% loss on Greek bonds |
| 12-Mar-12 | Greece Bailout Package Signed Off by EU Leaders |
| 3-Feb-15 | Greek government said to retreat from a demand for a debt writedown. |
| 29-Jun-15 | Greece closes banks |
| 6-Jul-15 | Greece bailout referendum on July 5th where voters reject austerity package |
| 10-Jul-15 | The Greek government submitted its highly anticipated plan for the country's economic overhaul to bailout authorities |
| 13-Jul-15 | Greek PM Alexis Tsipras conceded to a further swathe of austerity measures and economic reforms |
| 29-May-18 | Italian political turmoil (snap election plus new budget plan) cause largest 1-day decline in Italian bonds in 25 years |

Figure 3: Multiplier estimates using an expanding window. The figure reports the estimates of the multiplier in (61) using three principal components as controls. We estimate the model using an expanding sample where the data are ordered by $|u^e_{St}|$, that is, the magnitude of the idiosyncratic shocks. The number of dates included is depicted on the horizontal axis, starting with 15 observations. The solid blue line corresponds to the point estimate and the dashed red lines to the 95%-confidence interval. The model is estimated using daily data from July 2009 until May 2018.
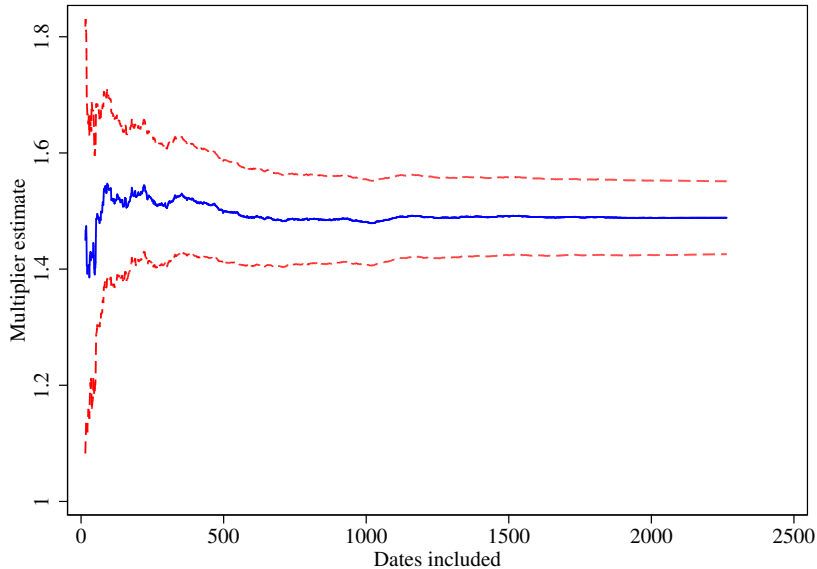


we indicate the number of dates included on the horizontal axis (starting at 15 observations). The estimate is stable for different samples but obviously standard errors tighten as the sample expands.

Panel A in Table 5 reports the yield changes on the 10 days with the largest realization of $|u^e_{St}|$. In Panel B, we scale the yield changes by $0.01 + y_{i,t-1}$. In Panel C, we provide the narratives. If we inspect some of the largest shocks in Table 5, then is is clear that most of them are truly idiosyncratic shocks. Examples include the decision by Greece to close all banks or the outcome of the referendum. There are two shocks, however, on May 10, 2010 and August 8, 2011 that involve actions by the ECB and hence are more likely aggregate shocks as opposed to idiosyncratic shocks. Removing these dates does not impact our estimates, but illustrates the empirical relevance of sporadic factors during times of crisis (see Section 2.6). Most of the shocks are associated with Greece, although the final date corresponds to Italy.[49]

---

[49]Also following the end of our sample, many of the idiosyncratic shocks in recent months happened in Italy.

# 7 Multipliers and elasticities in crude oil markets

## 7.1 Data

Our data construction follows the existing literature (Kilian (2009), Caldara et al. (2018), Baumeister and Hamilton (2019), henceforth BH19). The data on crude oil production are from the January 2019 Monthly Energy Review published by the U.S. Energy Information Administration (EIA). We observe the monthly oil production for 20 countries from January 1985 until December 2015.[50] The selection of countries includes OPEC and non-OPEC members. In addition to the production data of these 20 countries, we also observe the production by (current) OPEC and non-OPEC members as well as the world's production. We use the total non-OPEC production to construct a fictitious country that produces the residual non-OPEC supply.[51]

We focus on estimating short-run (monthly) multipliers as well as demand and supply elasticities, consistent with the literature. To construct innovations, we use a state vector $X_t$ that includes lagged (i) world supply growth, (ii) monthly oil price changes, (iii) changes in inventories, and (iv) growth in industrial production. We use the data of BH19 for the latter three series.

## 7.2 Model

We model the supply growth of country $i$ in period $t$ as

$$\Delta y_{it} = a_i + \phi^s \Delta p_t + \gamma'_y X_{t-1} + \lambda_i \eta_t + u_{it},$$

and model changes in aggregate oil demand (both in use and inventories) as

$$\Delta d_t = c_d + \phi^d \Delta p_t + \gamma'_d X_{t-1} + \epsilon_t.$$

Market clearing, $\Delta y_{St} = \Delta d_t$, implies

$$\Delta p_t = c_p + \frac{M}{\phi^d} u_{St} + \gamma^{p\prime} \mathcal{C}_t, \qquad \Delta y_{St} = c_{y_S} + M u_{St} + \gamma^{y\prime} \mathcal{C}_t,$$

where $M = -\frac{\phi^d}{\phi^s - \phi^d} \in [0,1]$ is the multiplier, and $(\gamma^p, \gamma^y)$ are loadings on $\mathcal{C}_t = (\eta_t, \varepsilon_t, X_{t-1})$, and whose precise values do not matter here.

Our goal is to estimate the short-run supply and demand elasticities, $\phi^s$ and $\phi^d$ (with presumably $\phi^d < 0 < \phi^s$). The equations for aggregate supply and price changes are part of the VAR models that are commonly used in the recent literature on oil prices and their impact on economic growth.

---

[50]We follow Caldara et al. (2018) and remove Gabon from the sample due to concerns about data quality. In addition, we scale the supply of the USSR using the ratio of supply of the USSR to the supply of Russia to obtain a continuous series and to avoid a sudden jump in the non-OPEC supply.

[51]This fictitious country includes the countries that were not part of OPEC during our sample.

## 7.3 GIV estimation

The supply changes in a couple of months are extreme for some countries during supply disruptions, and we therefore winsorize the growth rates at 2.5% and 97.5% across all countries and periods to estimate $\Delta y_{Et} = \Delta y_{Et}^w$, where $\Delta y_{it}^w$ denotes winsorized supply growth.[52] We then estimate the model using the following steps, which follows the procedure in Section 4:

1. Run a panel regression with country and time fixed effects,[53]

$$\Delta y_{it}^w = a_i + b_t + \Delta \breve{y}_{it}^w.$$

2. Use $\Delta \breve{y}_{it}^w$ to estimate a parametric factor based on OPEC membership.[54] Let $I_{it}^{OPEC} = 1$ if a country is an OPEC member in month $t$ and $I_{it}^{OPEC} = 0$ otherwise. We then estimate

$$\Delta \breve{y}_{it}^w = b_t + I_{it}^{OPEC} \eta_t^{OPEC} + e_{it},$$

and we denote the estimated factor by $\eta_t^{OPEC,e}$. We use $\Delta \breve{y}_{it}^w$ also to separately estimate a principal component, and we denote the estimated factor by $\eta_t^{PCA,e}$. We stack the estimated factors together in $\eta_t^e = \left( \eta_t^{OPEC,e}, \eta_t^{PCA,e} \right).$

3. Estimate $\frac{M}{\phi^d}$ and $M = -\frac{\phi^d}{\phi^s - \phi^d}$ using respectively, with $Z_t = y_{St} - y_{Et}^w$,

$$\Delta p_t = a_p + \frac{M}{\phi^d} Z_t + \beta_p^{\eta\prime} \eta_t^e + \beta_p^{X\prime} X_{t-1} + e_t^p, \tag{62}$$

$$\Delta y_{St} = a_S + M Z_t + \beta_S^{\eta\prime} \eta_t^e + \beta_S^{X\prime} X_{t-1} + e_t^y. \tag{63}$$

4. Estimate the demand and supply elasticities using the following regressions, where we instrument $\Delta p_t$ using $Z_t$,

$$\Delta y_{St} = a_S + \phi^d \Delta p_t + \beta_S^{\eta\prime} \eta_t^e + \beta_S^{X\prime} X_{t-1} + e_t^d, \tag{64}$$

$$\Delta y_{Et} = a_E + \phi^s \Delta p_t + \beta_E^{\eta\prime} \eta_t^e + \beta_E^{X\prime} X_{t-1} + e_t^s. \tag{65}$$

## 7.4 Empirical results

We report the estimation results of the multipliers $M = 0.91$ and $\frac{M}{\phi^d} = -1.97$ in Table 6 alongside both elasticities. The estimate of $\frac{M}{\phi^d}$ implies that a 1% decline in the global oil supply leads to a 1.97% increase in oil prices. The estimate of $M$ implies that a 1% shock to supply leads to an overall

---

[52]To ensure growth rates are always defined, we set supply to one in case it drops to zero, which happens in seven country-months.

[53]Note that the time fixed effects absorb the controls, $X_{t-1}$, in this case.

[54]OPEC membership is largely stable during our sample except for Angola and Ecuador. We account for changes in membership in estimating the factor.

Table 6: Multiplier estimates in the oil market. The first column reports the estimate of $M$ as the coefficient on $Z_t$ in (62). The second column reports the estimate of $\frac{M}{\phi^d}$ as the coefficient on $Z_t$ in (63). The third column reports the Two Stage Least Squares (2SLS) estimate of the demand elasticity $\phi^d$, see (64), and the fourth column the 2SLS estimate of the supply elasticity $\phi^s$, see (65). In both cases, we use $Z_t$ as an instrument for $\Delta p_t$. We suppress the coefficients on the controls, $X_{t-1}$, that include lagged (i) world supply growth, (ii) monthly oil price changes, (iii) changes in inventories, and (iv) growth in industrial production. The $t$-statistics are based on OLS and 2SLS standard errors in parentheses. We also report the t-statistics using standard errors that are robust to heteroskedasticity and autocorrelation in brackets. We select the optimal lag length as in Newey and West (1994). The sample is from January 1985 to December 2015.

|  | $y_{St}$ | $\Delta p_t$ | $y_{St}$ | $y_{Et}$ |
|---|---|---|---|---|
| $Z_t$ | 0.913 | -1.974 | | |
| | (16.21) | (-3.60) | | |
| | [14.11] | [-2.34] | | |
| $\eta_t^{PCA}$ | -0.118 | 0.173 | -0.0379 | -0.126 |
| | (-15.29) | (2.30) | (-0.97) | (-13.67) |
| | [-11.71] | [2.51] | [-0.76] | [-9.78] |
| $\eta_t^{OPEC}$ | 0.409 | -0.197 | 0.318 | 0.418 |
| | (15.07) | (-0.75) | (2.55) | (14.35) |
| | [8.92] | [-0.72] | [2.25] | [7.89] |
| $\Delta p_t$ | | | -0.463 | 0.0438 |
| | | | (-3.54) | (1.43) |
| | | | [-2.52] | [0.97] |
| $N$ | 370 | 370 | 370 | 370 |
| $R^2$ | 0.578 | 0.262 | | |

reduction in oil supply of 0.91% as other countries step in to increase their supply in response to higher prices. We estimate a demand elasticity of $\phi^d = -0.46$ and a supply elasticity of $\phi^s = 0.044$. Changes in demand also include changes in inventories, which respond more elastically to changes in prices (Kilian and Murphy (2014)). In parentheses, we report the t-statistics using OLS and 2SLS standard errors. We also report the t-statistics with more conservative standard errors in brackets that are robust to heteroskedasticity and autocorrelation. We select the optimal lag length as in Newey and West (1994). Recall that while the multiplier estimates rely on standard OLS inference, the elasticity estimates use instrumental variables. This is important, as the instrument for price is quite weak, and we are therefore cautious when interpreting the elasticity estimates and emphasize the multiplier estimates instead.

To put these estimates in perspective, we compare them to recent estimates in the literature and find them to be in broad agreement. Indeed, Baumeister and Hamilton (2019) use sign restrictions in combination with a Bayesian estimator to find supply and demand elasticities of 0.15 and $-0.35$, respectively, with 68% credibility intervals of $(0.09, 0.22)$ for the supply elasticity and $(-0.51, -0.24)$ for the demand elasticity. The point estimates imply multipliers of $M = 0.7$ and $\frac{M}{\phi^d} = -2$. Kilian and Murphy (2014) also combine sign restrictions and a Bayesian estimator with short-run supply elasticities bounded at 0.025, 0.05, and 0.1, and corresponding demand elasticities ranging from -0.44 to -0.47. For the 5% bound, the multiplier estimates equal $M = -0.9$ and $\frac{M}{\phi^d} = -2$. Caldara et al. (2018) use a narrative approach and estimate a supply elasticity of 0.08 (with a standard error of 0.037) and a demand elasticity of -0.08 (with a standard error of 0.08), yielding multipliers of $M = 0.5$ and $\frac{M}{\phi^d} = -6.3$.

To further compare the GIV to instruments used in the literature, we juxtapose it to that employed by Caldara et al. (2018). We construct our instrument as the residual from a regression of $y_{\Gamma t}$ on $X_{t-1}$ and the two factors, $\eta_t^e$, and refer to it as $u_{\Gamma t}$. If we regress it on the instrument of Caldara et al. (2018), which is non-zero only during 14 months in this sample, we get a slope coefficient of 0.92, with a t-statistic of 13.5 and an R-squared of 94%. Moreover, if we restrict ourselves to more extreme episodes by only using data when $u_{\Gamma t}$, in absolute value, exceeds a threshold, then the estimates of $\frac{M}{\phi^d}$ becomes more negative.[55] A potential interpretation is that demand becomes more inelastic in case of more extreme shocks, for instance, as inventories are running low. This reconciles our estimates with those of Caldara et al. (2018).

In summary, the GIV estimator results in estimates that are in the range of estimates documented in the recent literature, thereby providing some external validation of GIVs as an approach to estimating demand and supply elasticities. At the same time, the GIV procedure arguably requires less domain-specific ingenuity than the previous studies we mentioned.

Several interesting extensions are worth exploring.[56] First, we can estimate demand elasticities

---

[55]If we restrict $|u_{\Gamma t}|$ exceeding 0.5%, 1%, and 1.5%, the first-stage slope decreases from -1.97 to -2.19, -3.46, and -4.76, respectively.

[56]We are grateful to Lutz Kilian for suggesting these potential extensions.

for oil demand in production and oil demand in use (Kilian and Murphy (2014)). Second, Saudi Arabia plays a key role within OPEC and steps in quickly in times of supply disruptions, for instance, during the First Gulf War. One can use GIV to estimate a different supply curve for Saudi-Arabia than for the other countries, and in particular allow Saudi Arabia to respond to the production by other OPEC countries. Third, detailed country-level data on (net) imports and oil consumption can be used to construct a second instrument that can be used to both sharpen the estimates and to test for overidentifying restrictions. Fourth, one could use GIVs to explore non-linear demand curves to understand the multipliers during large supply disruptions (see Section D.8).[57]

# 8 Discussion and extensions of the framework

## 8.1 Extensions

There are many ways to increase the number of setups in which the GIV idea can be applied. In our experience, one can extend the GIV idea to many settings and generalizations.

**Multidimensional GIV** One can handle multidimensional "actions:" for instance, a firm could have a shock that affects both productivity and labor demand. A country could have a shock that affects both productivity and oil demand. Formally, the actions $y_{it}$ and shocks $u_{it}$ are now multidimensional. The GIV idea goes through, and this is developed in Section D.1. We have seen that with one GIV, we can estimate $1 + d_F$ parameters ($M$, $M\alpha^f$), where $d_F$ is the number of endogenous, observed factors. With $q$-dimensional actions, we have $q$ GIVs, and we can estimate $q^2 + qd_F$ parameters, which correspond to $M$ and $\alpha^f$.[58] So, potentially many parameters can be recovered with multidimensional "actions" by firms or countries.

**GIV with a more complex matrix of influences** The GIV can also be extended to non-homogeneous influences in the context of loops. Suppose a model $y_{it} = \gamma \sum_j G_{ij} y_{jt} + \lambda_i \eta_t + u_{it}$, i.e., in vector form

$$y_t = \gamma G y_t + \Lambda \eta_t + u_t, \tag{66}$$

with a given "influence" matrix $G$ (in our baseline model, $G = \iota S'$). We'd like to identify $\gamma$, the strength of linkages.

A simple generalization of our GIV is to define a "size" vector $S := G'E$. Then, left-multiplying (66) by $E'$, we get

$$y_{Et} = \gamma y_{St} + \Lambda_E \eta_t + u_{St}.$$

The key moment is still $\mathbb{E}\left[(y_{Et} - \gamma y_{St}) z_t\right] = 0$, where the GIV is again $z_t = y_{St} - y_{Et}$ in the simple

---

[57]One would interact with $z_t$ by, for example, the log previous-period price to see the elasticity at different prices.

[58]As $u_{St}$ and $y_{St}$ are $q$-dimensional, $M = \frac{dy_{st}}{du_{St}}$ $q \times q$ dimensional, and each of the $\frac{dF_t^f}{du_S} = M\alpha^f$ is also $q$-dimensional.

case where $\Lambda = \iota$ and $G\iota = \iota$; see Section D.14 for the general case. Hence, GIV generalizes to "spatial" models with common shocks (most spatial models do not have latent common shocks).[59]

**Bayesian GIV**   Another extension is a Bayesian interpretation of the GIVs. This way, we can interpret GIVs in a Bayesian framework – see Section E. In particular, this opens the possibility of marrying GIV estimation with priors on other parameters. In the simplest cases with Gaussian shocks, the maximum likelihood estimate is our GIV – confirming its optimality properties. At the same time, the basic GIV doesn't actually use normality assumptions.

**Vector autoregressions with GIV**   One can do vector autoregressions with GIVs: if $Y_t = AY_t + X_t$, 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).[60] One can also do Jordà (2005) style local projections, regressing $\mathbb{E}_t[Y_{t+h}] = \beta_t X_t$, and instrumenting some of the regressors $X_t$ by a GIV. Both methods give a way to do impulse responses with GIVs.

## 8.2   Comparison with Bartik instruments and other procedures

**Comparison with Bartik instruments**   The GIV estimator shares some similarities with Bartik instruments,[61] but also major differences To put it simply, Bartik instruments allow to estimate the cross-sectional (or micro) sensitivities to shocks, but not aggregate sensitivities; whereas GIVs are mostly designed to estimate aggregate (or macro) elasticities. Hence, they are complementary.

To see this, let us use the notation established earlier. Shift-share estimators aim to estimate the coefficients $\lambda_1^f$ in the structural equation $y_{it} = \sum_f (\lambda_0^f + \lambda_1^f x_{it}) F_t^f + \eta_t^y + u_{it}$ using $x_{it} g_t^f$ as an instrument for $x_{it} F_t^f$. In this notation, the "shares" are $x_{it}$ and the "shifters" are $g_t^f$ (for instance $g_t^f$ could be the China shock, and be correlated with $\eta_t^y$). Shift-share estimators have been the study of much recent econometric work including Goldsmith-Pinkham et al. (2020); Adao et al. (2019); Borusyak et al. (2020). Borusyak et al. (2020) lay out sufficient identifying conditions for the shift-share estimator to estimate the structural parameter. Shift-share estimators estimate $\lambda_1^f$ but are unable to estimate the mean effect $\lambda_0^f$.[62] In contrast, the GIV estimator provides a general strategy to estimate $\lambda_0^f$ if the shock $F_t^f$ is affected by some large idiosyncratic shocks.

**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)

---

[59]See Brownlees and Mesters (2020) for a potential way to extend this approach when $G$ is unknown.

[60]See Plagborg-Moller and Wolf (2020) for a recent development in this area.

[61]Known as "shift-share estimators," they were first introduced in Bartik (1991) and popularized in Blanchard and Katz (1992).

[62]If we have just one date, one can identify $\lambda_1^f F_t^f$ for that date $t$ by a cross-sectional regression. Hence, a Bartik approach is viable with just one cross-section, whereas the GIV needs a whole panel (fixed $N$, large $T$).

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:[63]

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

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, $\eta_t$, 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 5).

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 feedback loops: shocks to banks affect aggregate investment, but aggregate investment does not circle back around to affect individual bank behavior (so, they assume $\alpha^f = 0$ in our notations). 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.

**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.

---

[63]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.

**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. This is sketched in (66) and Section D.15 develops this in some detail.[64]

**Other literatures**   In a tangentially related recent paper, Sarto (2018) uses factor analysis to extract values of $\eta^f$ (much as we do when we "recover" a factor $\eta^f$). Take the basic example in our paper. Then, Sarto does not identify $\alpha$: even if $\eta$ (the aggregate shock to demand) were perfectly identified, that would not allow to estimate $p$. In the supply and demand example, Sarto would identify the demand elasticity $\phi^d$, but not the supply elasticity $\phi^s$.

*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 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 feedback loops from banks to sovereign yields (and vice versa). 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 we will see in the empirical examples below. 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.[65]

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

---

[64]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.

[65]We discuss the robustness of GIVs to various forms of misspecification in Section 2.6.

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.

## 8.3   When aggregate shocks are made of idiosyncratic shocks

We now discuss how GIVs extend to economies where aggregate shocks $\eta_t$ are themselves made of idiosyncratic shocks $u_{it}$, as in Carvalho and Gabaix (2013). Take the basic supply and demand model of Section 2.1. In the case without loops, 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. 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 the factor model:

$$y_{it}^k = \phi^k p_t + \lambda_i^k \eta_t^k + u_{it}^k, \tag{67}$$

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 "fracking shock" that affects both supply and demand, it will be captured by both $u_{it}^s$ and $u_{it}^d$ for $i = \text{USA}$. This is a concrete case in which supply and demand shocks are correlated: this happens via the correlations in country-level shocks. Suppose that this correlation captures the common shocks, so that $\eta_t^s \perp u_{\Gamma^d t}^d$ (where $\Gamma^d$ are the residual granular weights given by the demand-side relative size): then, we can identify the elasticity of supply, via $u_{\Gamma^d t}^d$. Likewise, if $\eta_t^s \perp u_{\Gamma^s t}^s$ then the GIV $u_{\Gamma^s t}^s$ allows to estimate the supply elasticity $\phi^s$. Section D.4 details this, and gives more variants.

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 D.13. 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 D.14 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.

# 9   Conclusion

We developed granular instrumental variables (GIVs): we remark that idiosyncratic shocks offer a rich source of instruments, and we lay out econometric procedures to optimally extract them from aggregate shocks. Many econometric extensions might be useful, for instance with stochastic volatility and various dimensions of autocorrelations. We leave those extensions to future research.

We provided two empirical applications. Many more applications seem within reach, and the introduction listed some. We hope that GIVs will aid identifications 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.

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

**Anderson, T. W. and Yasuo Amemiya**, "The Asymptotic Normal Distribution of Estimators in Factor Analysis under General Conditions," *The Annals of Statistics*, 1988, *16* (2), 759–771.

**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.

**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**, "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.

**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.

**Baumeister, Christiane and James D Hamilton**, "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," *American Economic Review*, 2019, *109* (5), 1873–1910.

**Blanchard, Olivier Jean and Lawrence F. Katz**, "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, *23* (1), 1–76.

**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," *NBER Working Paper No. 24997*, 2020.

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

**Caballero, Ricardo J and Alp Simsek**, "A model of fickle capital flows and retrenchment," *Journal of Political Economy*, 2020, *128* (6), 2288–2328.

**Caldara, Dario, Michele Cavallo, and Matteo Iacoviello**, "Oil price elasticities and oil price fluctuations," *Journal of Monetary Economics*, 2018.

**Camanho, Nelson, Harald Hau, and Hélène Rey**, "Global portfolio rebalancing and exchange rates," Technical Report, National Bureau of Economic Research 2020.

**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 Miguel and Basile Grassi**, "Large Firm Dynamics and the Business Cycle," *American Economic Review*, 2019.

**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," *Review of Financial Studies*, 2021.

**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.

**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**, "Power Laws in Economics and Finance," *Annual Review of Economics*, May 2009, *1* (1), 255–294.

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

**Gabaix, Xavier and Matteo Maggiori**, "International liquidity and exchange rate dynamics," *The Quarterly Journal of Economics*, 2015, *130* (3), 1369–1420.

**Gabaix, Xavier and Ralph SJ Koijen**, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," *Available at SSRN 3686935*, 2020.

**Galaasen, Sigurd, Rustam Jamilov, Ragnar Juelsrud, Hélène Rey et al.**, "Granular credit risk," Technical Report, NBER Working Paper No. 27994 2020.

**Gaubert, Cecile and Oleg Itskhoki**, "Granular comparative advantage," *Forthcoming at the Journal of Political Economy*, 2020.

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

**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.

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

**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.

**Kilian, Lutz**, "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 2009, *99* (3), 1053–69.

**Kilian, Lutz and Daniel P Murphy**, "The role of inventories and speculative trading in the global market for crude oil," *Journal of Applied Econometrics*, 2014, *29* (3), 454–478.

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

**Koijen, Ralph SJ and Motohiro Yogo**, "An equilibrium model of institutional demand and asset prices," *Journal of Political Economy*, 2019.

**Kramarz, Francis, Julien Martin, and Isabelle Mejean**, "Volatility in the small and in the large: The lack of diversification in international trade," *Journal of International Economics*, 2020, *122*, 103276.

**Kundu, Shohini and Nishant Vats**, "Banking Networks and Economic Growth: From Idiosyncratic Shocks to Aggregate Fluctuations," Technical Report, Working paper Chicago Booth 2020.

**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.

**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.

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

**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.

**Newey, Whitney K and Kenneth D West**, "Automatic lag selection in covariance matrix estimation," *The Review of Economic Studies*, 1994, *61* (4), 631–653.

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

**Pasten, Ernesto, Raphael Schoenle, and Michael Weber**, "Price Rigidity and the Origins of Aggregate Fluctuations," *NBER Working Paper No. 23750*, 2020.

**Plagborg-Moller, Mikkel and Christian K Wolf**, "Instrumental Variable Identification of Dynamic Variance Decompositions," Technical Report 2020.

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

**Rigobon, Roberto**, "Identification through heteroskedasticity," *Review of Economics and Statistics*, 2003, *85* (4), 777–792.

**Sarto, Andres Pablo**, "Recovering Macro Elasticities from Regional Data," *Working Paper*, 2018.

**Schubert, Gregor, Anna Stansbury, and Bledi Taska**, "Monopsony and Outside Options," *Working Paper*, 2020.

**Stock, James H and Mark W Watson**, "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," in "Handbook of macroeconomics," 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.

**Stock, James H and Motohiro Yogo**, "Testing for weak instruments in linear IV regression," in Donald WK Andrews and James H Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 2005.

**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.

<p style="text-align:center">Online Appendix for</p>

<p style="text-align:center">"Granular Instrumental Variables"</p>

<p style="text-align:center">Xavier Gabaix and Ralph S.J. Koijen</p>

<p style="text-align:center">December 5, 2020</p>

## Table of contents

# C  Proofs omitted in the paper

## C.1  Variance facts

We will repeatedly use the fact that if $(u_i)_{i=1\ldots N}$ is a series of uncorrelated random variables with mean 0 and common variance $\sigma_u^2$, then

$$\mathbb{E}\left[u_\Gamma u_E\right] = 0, \qquad \mathbb{E}\left[u_\Gamma^2\right] = \mathbb{E}\left[u_S u_\Gamma\right] = h^2 \sigma_u^2. \tag{68}$$

Hence, the standard deviation of the granular residual $u_{\Gamma t}$ is proportional to the Herfindahl. In the general heteroskedastic case, the quasi-equal weight vector is $\tilde{E} = \frac{(V^u)^{-1}\iota}{\iota(V^u)^{-1}\iota}$. Then, for any $\Gamma$ such that $\iota'\Gamma = 0$, we have:[66]

$$\mathbb{E}\left[u_\Gamma u_{\tilde{E}}\right] = 0. \tag{69}$$

## C.2  Proof of Proposition 2

The proof is quite elementary, and uses well-known ingredients. We have

$$\alpha_T^e - \alpha = \frac{\mathbb{E}_T\left[(\alpha y_{St} + \varepsilon_t)\, u_{\Gamma t}\right]}{\mathbb{E}_T\left[y_{St} u_{\Gamma t}\right]} - \alpha = \frac{\mathbb{E}_T\left[\varepsilon_t u_{\Gamma t}\right]}{\mathbb{E}_T\left[y_{St} u_{\Gamma t}\right]} = \frac{A_T}{D_T}.$$

---

[66]Here is the proof. We have $\tilde{E} = k\left(V^u\right)^{-1}\iota$ for $k = \frac{1}{\iota(V^u)^{-1}\iota}$. So

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

as $\iota'\Gamma = 0$.

Next, the law of large number gives:

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

with

$$D = \mathbb{E}\left[y_{St}u_{\Gamma t}\right] = \mathbb{E}\left[(\eta_t + u_{St})u_{\Gamma t}\right] = \mathbb{E}\left[u_{St}u_{\Gamma t}\right] = \mathbb{E}\left[(u_{\Gamma t} + u_{Et})u_{\Gamma t}\right] = \mathbb{E}\left[u_{\Gamma t}^2\right] = \sigma_{u_\Gamma}^2.$$

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

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

where (assuming that $\varepsilon_t$ is homoskedastic conditional on the $u_{it}$'s):

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

so that

$$\frac{\sigma_A}{D} = \frac{\sigma_\varepsilon\sigma_{u_\Gamma}}{\sigma_{u_\Gamma}^2} = \frac{\sigma_\varepsilon}{\sigma_{u_\Gamma}} =: \sigma_\alpha.$$

Hence,

$$\sqrt{T}\left(\alpha_T^e - \alpha\right) \to^d N\left(0, \sigma_\alpha^2\right).$$

Finally, if $u_{it}'$s are i.i.d. across $i$'s, then $\sigma_{u_\Gamma} = h\sigma_u$, see (68).

## C.3  Proof of Proposition 3

We have

$$\alpha_T^e - \alpha = \frac{\mathbb{E}_T\left[(\alpha y_{St} + \varepsilon_t)z_t\right]}{\mathbb{E}_T\left[y_{St}z_t\right]} - \alpha = \frac{\mathbb{E}_T\left[\varepsilon_t z_t\right]}{\mathbb{E}_T\left[y_{St}z_t\right]},$$

so the same proof as for Proposition 2 yields the asymptotic error

$$\sigma_\alpha(\Gamma) = \frac{\sigma_\varepsilon\sigma_z}{|\mathbb{E}\left[y_{St}z_t\right]|} = \frac{\sigma_\varepsilon\sigma_z}{|\mathbb{E}\left[u_{St}z_t\right]|} = \frac{\sigma_\varepsilon}{\sigma_{y_S}\left|corr\left(u_{St},z_t\right)\right|}.$$

So, the best estimator $z_t = u_{\Gamma t}$ maximizes the squared correlation $C(\Gamma) := corr\left(u_{St}, u_{\Gamma t}\right)^2$:

$$\max_\Gamma C(\Gamma) \text{ subject to } \iota'\Gamma = 0.$$

We next solve this problem.

Call $V = V^u$ the variance covariance matrix of the $u_i$. We have:

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

54

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

$$\mathcal{L} = \frac{1}{2}\Gamma'V\Gamma - b\Gamma'\iota - c\Gamma'VS \tag{70}$$

with some Lagrange multipliers $b$, $c$. The first order condition in $\Gamma'$ is: $0 = V\Gamma - b\iota - cVS$, i.e.

$$\Gamma = cS + bV^{-1}\iota.$$

Now, using $\iota'\Gamma = 0$ gives $0 = c + b\iota'V^{-1}\iota$, i.e., with $\tilde{E} := \frac{V^{-1}\iota}{\iota'V^{-1}\iota}$,

$$\Gamma = c\left(S - \tilde{E}\right).$$

The factor $c$ doesn't affect the results, (as $\Gamma$ and $c\Gamma$ give the same estimator $\alpha_T^e$), so we may choose $c = 1$.

## C.4  Justification of the procedure in Section 2.5 for heterogeneous demand elasticities

In vector notation, $y_t = \phi^d p_t + \lambda\eta_t + u_{it}$. We have

$$y_{it} = \phi_i^d p_t + \lambda_i\eta_t + u_{it} = X_{it}\left(\dot{\phi}^d p_t + \dot{\lambda}\eta_t\right) + u_{it},$$

i.e.

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

The more abstract way to think about the procedure in the main text is the following. We form the matrices $R$ and $Q$ from (21) (using $\Lambda = X$) for any $W$ at first, e.g. the identity matrix. We define: $\breve{u}_t := Qy_t$, so that $\breve{u}_t = Qu_t$ and the GIV is $z_t := S'\breve{u}_t$. As a column of $X$ is made of ones, we also have $z_t = \Gamma'\breve{u}_t$. Indeed, the cross-sectional regression of $y_{it}$ on $X_{it}$ identifies $\dot{y}_t = \dot{\phi}^d p_t + \dot{\lambda}\eta_t + Ru_t$, and the residuals $\breve{u}_{it}$. We have $\dot{y}_t = Ry_t$ and $\breve{u}_t = Qu_t$.

Then, we identify $\phi^s$ using

$$\mathbb{E}\left[(s_t - \phi^s p_t)z_t\right] = 0, \tag{71}$$

which is valid for any $W$. Next, if we have $W = c(V^u)^{-1}$ for some constant $c$, we can also estimate $\dot{\phi}^d$ using the moment (which is $r$-dimensional):[67]

$$\mathbb{E}\left[\left(\dot{y}_t - \dot{\phi}^d p_t\right)z_t\right] = 0. \tag{72}$$

---

[67]To see this, we form $Ry_t = \dot{\phi}^d p_t + \dot{\lambda}\eta_t + Ru_t$, and we have $\mathbb{E}[Ru_t\breve{u}_t'] = \mathbb{E}[Ru_t u_t'Q'] = RV^u Q' = 0$, because (22) indicates that $RV^u Q' = 0$ if $W = K(V^u)^{-1}$.

To see this, we form $Ry_t = \dot{\phi}^d p_t + \dot{\lambda}\eta_t + Ru_t$, and we have $\mathbb{E}\left[Ru_t \check{u}_t'\right] = \mathbb{E}\left[Ru_t u_t' Q'\right] = RV^u Q' = 0$, because (22) indicates that $RV^u Q' = 0$ if $W = c\left(V^u\right)^{-1}$ for some scalar $c$.[68]

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 $V^{\check{u}} = QV^u Q'$):[69]

$$\mathbb{E}\left[\check{u}_{ii}^2\right] = \left(QV^u Q'\right)_{ii}.$$

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 (72).

## C.5 Proof of Proposition 4

We first prove the proposition in the case where there are no controls $C_{it}m$ and non-zero mean values (so that $m$ and the mean values do not have to be estimated). Once this is done, we treat the case with controls and non-zero mean values.

**A canonical decomposition**   The model can be written as:

$$y_t = \lambda\eta_t + \iota\left(\left(M - 1\right)u_{St} + \varepsilon_t\right) + u_t, \qquad F_t^f = \alpha^f M u_{St} + \varepsilon_t^f \tag{73}$$

We next derive a "canonical representation" that, at the cost of some overhead and concepts and notations, will make the identifiability very crisp.

We take the constant $\lambda$ case to alleviate notation, but we could have a time-varying $\lambda_t$. We use the $Q$ and $R$ matrices of (21), reproduced here:

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

where $W$ will be specified soon and that we keep implicit here to alleviate notations.

We decompose $\lambda = \check{\lambda} + \iota\lambda_0$, where $\check{\lambda} = Q^\iota \lambda$ is orthogonal to $\iota$ —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 (22)).

This gives the following decomposition:

**Proposition 5** (Canonical representation) *When $W = c\left(V^u\right)^{-1}$ for some constant $c > 0$, we define $\Gamma := \left(Q^\lambda\right)' S$, and we have the following "canonical representation"*

$$y_t = \check{\lambda}\dot{\eta}_t + \iota y_{Et} + \check{u}_t, \qquad y_{Et} = \left(M - 1\right)u_{\Gamma t} + \dot{\varepsilon}_t, \qquad F_t^f = \alpha^f M u_{\Gamma t} + \dot{\varepsilon}_t^f \tag{74}$$

---

[68]If we take $W$ as the identity rather than the "ideal" $W$ above, the error in (72) is typically quite moderate: for instance, with $X = \iota$, it is only of order $\frac{1}{N}$.

[69]See Section D.10. 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 $\mathbb{E}\left[\check{u}_t \check{u}_t'\right] = QV^u Q'$, and not just its diagonal terms.

with $\check{u}_t = Q^\lambda u_t$ and $u_{\Gamma t} = S' \check{u}_t$. Also, $\check{u}_t$ and $u_{\Gamma t}$ are uncorrelated with $\dot{\eta}_t, \dot{\varepsilon}_t, \dot{\varepsilon}_t^f$, but the $\dot{\eta}_t, \dot{\varepsilon}_t, \dot{\varepsilon}_t^f$ can be correlated between themselves; and $\check{\lambda}, \iota$ and $\check{u}_t$ are all orthogonal (using the scalar product $\langle A, B \rangle = A'WB$ with $W = (V^u)^{-1}$). One advantage of this representation is that we have:

$$R^{\check\lambda} y_t = \dot{\eta}_t, \qquad y_{Et} = R^\iota y_t = (M - 1) u_{\Gamma t} + \dot{\varepsilon}_t \qquad Q^\lambda y_t = \check{u}_t \qquad (75)$$

with $R$ and $Q$ are the projection matrices defined in (21). Hence, we can recover $\dot{\eta}_t$ and $\check{u}_t$ by simple projections. Finally we have the following ancillary relations, with $\varepsilon_t^u := R^\lambda u_t$ uncorrelated with $\check{u}_t, u_{\Gamma t}, \dot{\eta}_t, \dot{\varepsilon}_t, \varepsilon_t^f$:

$$\dot{\eta}_t = \eta_t + \varepsilon_t^u, \qquad \dot{\varepsilon}_t = \varepsilon_t + (M - 1) \lambda_S \varepsilon_t^u + \lambda_0 \dot{\eta}_t, \qquad \dot{\varepsilon}_t^f = \varepsilon_t^f + \alpha^f M \lambda_S \varepsilon_t^u$$

**Proof of Proposition 5**   We decompose $u_t = Q^\lambda u_t + \left(I - Q^\lambda\right) u_t$, so:

$$u_t = \check{u}_t + \lambda \varepsilon_t^u, \qquad \varepsilon_t^u := R^\lambda u_t \qquad (76)$$

Taking $W = c (V^u)^{-1}$ has the important consequence that $\check{u}_t$ (an $N$ dimensional vector) and $\varepsilon_t^u$ (a $r$ dimensional vector) are uncorrelated. Indeed (dropping the superscript $\lambda$ in $R$ and $Q$ for simplicity)

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

using (22). 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 (73):

$$F_t^f = \alpha^f M u_{St} + \varepsilon_t^f = \alpha^f M \left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t^f = \alpha^f M u_{\Gamma t} + \dot{\varepsilon}_t^f$$

with $\dot{\varepsilon}_t^f = \varepsilon_t^f + \alpha^f M \lambda_S \varepsilon_t^u$.

In the same way, (73) gives

$$
\begin{aligned}
y_t &= \lambda \eta_t + \iota \left((M - 1) u_{St} + \varepsilon_t\right) + u_t \\
&= \lambda \eta_t + \iota \left[(M - 1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t\right] + \check{u}_t + \lambda \varepsilon_t^u \\
&= \lambda \dot{\eta}_t + \iota \left[(M - 1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t\right] + \check{u}_t \text{ with } \dot{\eta}_t = \eta_t + \varepsilon_t^u \\
&= \check{\lambda}_t \dot{\eta}_t + \iota \left[(M - 1)\left(u_{\Gamma t} + \lambda_S \varepsilon_t^u\right) + \varepsilon_t + \lambda_0 \dot{\eta}_t\right] + \check{u}_t \text{ as } \lambda = \check{\lambda} + \iota \lambda_0 \\
&= \check{\lambda}_t \dot{\eta}_t + \iota y_{Et} + \check{u}_t
\end{aligned}
$$

by defining

$$y_{Et} = (M - 1) u_{\Gamma t} + \dot{\varepsilon}_t, \qquad \dot{\varepsilon}_t = \varepsilon_t + (M - 1) \lambda_S \varepsilon_t^u + \lambda_0 \dot{\eta}_t$$

□

Note that $\eta_t$ cannot be recovered without error by $R^{\check{\lambda}} y_t$, but $\dot{\eta}_t$ can be. This is why the analysis is easier in that space.

This leads to the following procedure.

**An identification procedure**

1. We set $Q^\lambda$ as in (21), we construct the GIV $z_t := S' Q^\lambda y_t$, which implies that $z_t = S' Q u_t = u_{\Gamma t}$. We recover exactly $\dot{\eta}_t^e := R^{\check{\lambda}} y_t = \dot{\eta}_t$.

2. To estimate $\alpha^f$, we use the moment $\mathbb{E}\left[\left(F_t^f - \alpha^f y_{St}\right) z_t\right] = 0$, or, for more precision we can also use $\mathbb{E}\left[\left(F_t^f - \alpha^f y_{St} - \beta^f \dot{\eta}_t\right)(z_t, \dot{\eta}_t)\right] = 0$. We can also do OLS, e.g. run $F_t^f = \alpha^f M z_t + \beta^f \dot{\eta}_t + e_t$ (calling $e_t$ a random variable uncorrelated with $z_t$) and obtain an estimate of $\alpha^f M$.

3. We can also estimate $M$ similarly, e.g. we can run the OLS $y_{Et} = (M - 1) z_t + e_t$, or even $y_{St} = M z_t + e_t$. We can also add controls for $\eta_t^e$. Indeed, as $R^\iota y_t = (M - 1) z_t + e_t$, we can regress: $y_{Et} = (M - 1) z_t + \beta \dot{\eta}_t + e$, with $\dot{\eta}_t$ as control.

Proposition 5 implies that we can write

$$y_{Et} = (M - 1) z_t + \dot{\varepsilon}_t, \qquad y_{St} = M z_t + \dot{\varepsilon}_t, \qquad F_t^f = \alpha^f M z_t + \dot{\varepsilon}_t^f \tag{77}$$

with $z_t$ uncorrelated with $\dot{\varepsilon}_t$ and $\dot{\varepsilon}_t^f$ uncorrelated. Hence the assumptions of OLS are satisfied: the OLS standard errors are thus correct. The same holds if we control for $\dot{\eta}_t$.

Those standard errors are easy to calculate, repeating the arguments of the proof of Proposition 2. For instance, (77) implies that the OLS estimator $\tilde{\alpha}_T^{f,e}$ of $\tilde{\alpha}^f := \alpha^f M$ satisfies $\sqrt{T}\left(\tilde{\alpha}_T^{f,e} - \tilde{\alpha}^f\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\tilde{\alpha}^f}^2\right)$ with

$$\sigma_{\tilde{\alpha}^f} = \frac{\sigma_{\dot{\varepsilon}^f}}{\sigma_z}, \qquad \sigma_z = \mathbb{E}\left[u_{\Gamma t}^2\right]^{1/2}. \tag{78}$$

**Case with controls and non-zero mean values** We treat the case where there are controls $C_{it}^y$. A very similar argument applies when there are in addition controls $C^F m^F$ in the factors, and additive constants (so that terms do not need to have a zero mean) in all equations.

The model is then:

$$y_t = \lambda \eta_t + \iota\left((M - 1)\left(u_{St} + C_{St}^y m\right) + \varepsilon_t\right) + C_t m + u_t, \qquad F_t^f = \alpha^f M\left(u_{St} + C_{St}^y m\right) + \varepsilon_t^f \tag{79}$$

We call $\check{y}_t = R^\iota y_t$, i.e. $\check{y}_{it} = y_{it} - y_{\tilde{E}t}$ is the cross-sectionally demeaned value. Given a candidate value $m$, we construct $\dot{\eta}_t(m) := R^{\check{\lambda}}\left(y_t - C_t^y m\right)$ and the associated $z_t(m) := S' Q^\lambda\left(y_t - C_t^y m\right)$, as

58

we do in Section 4.1. Define $\theta$ to be $(m, M, \alpha)$ and $\theta^e$ to be the GMM estimator of $\theta$ associated with the following moments:

$$\mathbb{E}\left[\left(\check{y}_t - \check{x}_t \dot{\eta}_t(m) - m\check{C}_t^y\right) W \check{C}_t^y\right] = \mathbb{E}\left[g_1(m)\right] = 0, \tag{80}$$

$$\mathbb{E}\left[\left(y_{St} - Mz_t(m) - \beta^y \dot{\eta}_t(m)\right)\left(z_t(m), \dot{\eta}_t(m)\right)\right] = \mathbb{E}\left[g_2(\theta)\right] = 0 \tag{81}$$

$$\mathbb{E}\left[\left(p_t - \alpha Mz_t(m) - \beta^f \dot{\eta}_t(m)\right)\left(z_t(m), \dot{\eta}_t(m)\right)\right] = \mathbb{E}\left[g_3(\theta)\right] = 0. \tag{82}$$

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 $\frac{\partial \dot{\eta}_t(m)}{\partial m}$ and $\frac{\partial z_t(m)}{\partial m}$ are proportional $C_t^y$). Then, by Theorem 6.1 of Newey and McFadden (1994), the standard errors on $M, \alpha$ 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 $M$ and $\alpha^f M$ 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 $M$ and $\alpha^f M$ returned by OLS in this procedure are valid) holds even when $m$ needs to be estimated.[70]

# D   Complements

## D.1   Multi-dimensional actions

Suppose now that the 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 firms 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 have:

$$y_{S^a t} = \sum_f \lambda_{S^a, f}^a F^f + u_{S^a t},$$

$$F_t^f = \eta_t^f + \sum_a \alpha_a^f y_{S^a, t}^a,$$

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

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

where

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

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

---

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

$q \times q$, and

$$u_{St} = (u^a_{S^a t})_{a=1\ldots q}.$$

This implies

$$y_{St} = M\left(u_{St} + \varepsilon_t\right), \tag{83}$$

there 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: $u_\Gamma = (u^a_{\Gamma^a})_{a=1\ldots q}$. We want, with $E^a = S^a - \Gamma^a$,

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

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

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

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 have a 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:

$$Q^{ab} = \sum_i E^a_i \Gamma^b_i v^{ab}_i, \qquad v^{ab}_i := \mathbb{E}\left[u^a_{it}u^b_{it}\right]. \tag{84}$$

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

$$v^{ab}_i = \sigma^2_i v^{ab}. \tag{85}$$

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,

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

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

Given (83), we have

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

so

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

hence our estimator is

$$M = \mathbb{E}\left[y_{St}z_t'\right]\mathbb{E}\left[z_t z_t'\right]^{-1}. \tag{86}$$

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

**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

$$u_{\Gamma t,-i} := u_{\Gamma t} - S_i^u u_{it}, \tag{87}$$

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 $\mu_i$, by regression

$$y_{it} = \mu_i M u_{\Gamma t,-i} + \phi^i \mathcal{C}_t + \varepsilon_{it}^y, \tag{88}$$

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

## D.2 Heterogeneous demand elasticities: Non-parametric extension

**Non-parametric version for $\phi^d$**  We present a variant of the procedure in Section 2.5, but now with non-parametric heterogeneous demand elasticities $\phi_i^d$. The model is

$$y_t = \phi^d p_t + \lambda \eta_t + u_t, \tag{89}$$

We still assume parametric loading of unobserved factors $\eta$.

### D.2.1 Recommended procedure

We premultiply (89) 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 $\psi^c$ of $\psi$ we form the associated GIV: $z_t(\psi^c) := y_{\Gamma t} - \psi^c p_t$.

If we have the correct $z_t$, the following moments hold[71,72], with $b^p = \frac{1}{\phi^s - \phi_S^d}$ the coefficient of the first stage regression (34), $p_t = b^p z_t + \varepsilon_t^p$,

$$\mathbb{E}\left[\left(y_t - \phi^d p_t\right) z_t\right] = V^u \Gamma, \qquad \mathbb{E}\left[(p_t - b^p z_t) z_t\right] = 0 \tag{92}$$

$$\mathbb{E}\left[\left(\check{y}_{it} - \check{\phi}_i^d p_t\right)^2\right] = V_{ii}^{\check{u}}, \qquad \mathbb{E}\left[z_t^2\right] = \Gamma' V^u \Gamma \tag{93}$$

which potentially allow to estimate, respectively, $\phi^d$, $b^p$ (hence $\phi^s$) the $V_{ii}^u$ and $\phi_\Gamma$. Indeed, of we know $z_t$, we know $\phi^d$ and $b^p$.[73]

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 (93). So, we solve for $\psi^c$ (a *c*andidate answer for $\psi$) so that the empirical variance of the GIV is equal to its theoretical variance:

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

i.e. $\mathbb{E}\left[p_t^2\right](\psi^c)^2 - 2\mathbb{E}\left[y_{\Gamma t} p_t\right]\psi^c + \mathbb{E}\left[y_{\Gamma t}^2\right] - \sigma_{u_\Gamma}^2 = 0$. This is a quadratic equation in $\psi^c$, which yields two roots:[74] a good (i.e. correct) root, $\psi^G = \psi$, and a bad root, $\psi^B = \psi + 2\frac{\mathbb{E}[z_t^* p_t]}{\mathbb{E}[p_t^2]}$. Fortunately, 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:

$$b^{p,B} = -b^{p,G}, \tag{94}$$

Hence, if we have a prior on the sign of of the first stage coefficient $b^p$ (e.g. we know that $b^p > 0$

---

[71] 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 $V^{\check{u}} = QV^u Q'$.

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

$$\mathbb{E}\left[\left(y_{\tilde{E}t} - \phi_{\tilde{E}} p_t\right) z_t\right] = 0 \tag{90}$$

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

$$\mathbb{E}\left[\left(\check{y}_{it} - \check{\phi}_i p_t\right) z_t\right] = (QV^u \Gamma)_i \tag{91}$$

[73] We recommend starting from the parametric estimates of Section 2.5, which gives potentially good starting values for $\phi^d, z_t$ and $V^u$.

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

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

in a demand and supply model), we can choose the correct root as the one yielding a positive $b^p$ in the first stage.

### D.2.2 Justification

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

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

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

$$\mathbb{E}\left[z_t^B p_t\right] = -\mathbb{E}\left[z_t p_t\right], \qquad \mathbb{E}\left[\left(z_t^B\right)^2\right] = \mathbb{E}\left[z_t^2\right] \tag{95}$$

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:

$$b^{p,B} = \frac{\mathbb{E}\left[p_t z_t^B\right]}{\mathbb{E}\left[\left(z_t^B\right)^2\right]} = \frac{-\mathbb{E}\left[p_t z_t\right]}{\mathbb{E}\left[z_t^2\right]} = -b^{p,eG} \tag{96}$$

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 an nonparametric model with heterogeneous elasticities, and that fortunately (as in our recommended 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 $\varepsilon_t$ 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 $\eta_t$), $p_t = \alpha M u_{St} + \ddot{\varepsilon}_t$, 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 $\phi$ and $\alpha M$.

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[75]

$$Y_t = \Lambda \eta_t + \Phi \ddot{\varepsilon}_t + \alpha M \Phi u_{St} + U_t \tag{97}$$

---

[75]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.

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

$$Y_t = \Lambda \eta_t + \Psi \varepsilon_t + \Psi u_{St} = \Lambda \eta_t + \Psi \varepsilon_t + (I + \Psi S') U_t \qquad (98)$$

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

$$
\begin{align}
V^Y &= \sigma_\eta^2 \Lambda \Lambda' + \sigma_\varepsilon^2 \Psi \Psi' + \sigma_{\eta\varepsilon}(\Lambda \Psi' + \Psi \Lambda') + (I + \Psi S') V^U (I + S \Psi') \qquad (99) \\
&= \sigma_\eta^2 \Lambda \Lambda' + \gamma \Psi \Psi' + \Psi b' + b \Psi' + V^U \qquad (100) \\
b &= \sigma_{\eta\varepsilon} \Lambda + V^U S \qquad (101) \\
\gamma &= \sigma_\varepsilon^2 + S' V^U S \qquad (102)
\end{align}
$$

The idea for the multiplicity of roots in $\Psi$ is that we have a second degree equation in $\Psi$, so that 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

$$(\Psi - a)(\Psi - a)' = C := \frac{1}{\gamma}\left(V^Y - \sigma_\eta^2 \Lambda \Lambda'\right) + aa' \qquad (103)$$

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

$$\Psi^B = 2a - \Psi \qquad (104)$$

Now, because $\Lambda_p$ 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 101) and thus $a_p = 0$. So the component of the bad root on the price is $\Psi_p^B = 2a_p - \Psi_p = -\Psi_p$:

$$\Psi_p^B = -\Psi_p \qquad (105)$$

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 $\alpha M$ (e.g., it is positive in a supply and demand context), we can pick the good root by inspecting the sign of $\Psi_p$.

### D.2.3 Another procedure for the nonparametric estimation of heterogeneous demand elasticities

Recall the model (89). We replace $\phi^d$ by $\phi$ for simplicity:

$$y_t = \phi p_t + \lambda \eta_t + u_t, \tag{106}$$

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

1. Guess a candidate for $\phi$, called $\phi^c$, 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

$$Q^\phi y_t = \left(Q^\phi \lambda\right) \eta_t + Q^\phi u_t \tag{107}$$

2. We can apply the "singular factor analysis" procedure of Section D.11 to $Q^\phi y_t$ (so, in the notation of that section, $G = Q^\phi$). This returns: $\check{\lambda} := Q^\phi \lambda$, $\eta^e_t$, $V^u$, $\check{u}_t = Q^{\check{\lambda}, \phi} u_t$. We form $z_t := S' \check{u}_t$, and $\Gamma := \left(Q^{\check{\lambda}, \phi}\right)' S$.

3. We estimate $b^p$ and $\phi$ by (92). To gain some precision, instead of $\mathbb{E}\left[(y_t - \phi p_t) z_t\right] = V^u \Gamma$, we premultiply it by $Q^{\check{\lambda}, V_u^{-1}}$, i.e. we solve for $\phi$ in the moment:

$$\mathbb{E}\left[Q^{\check{\lambda}, V_u^{-1}} (y_t - \phi p_t) z_t\right] = Q^{\check{\lambda}, V_u^{-1}} V^u \Gamma \tag{108}$$

   This is equivalent to controlling for the $\eta_t$ in the regression.

4. Given this new estimates of $\phi$ 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 2.5), but keep the loadings $\lambda$ non-parametric. Then, in steps 1-2 we replace $\phi$ by $X$, and in the last step we replace $\phi$ by $X\dot{\phi}$ and estimate $\dot{\phi}$.

## D.3 Complements to the general procedure

**The procedure can be simplified in some cases.** *When we have a long time-series.* Recall that

$$y_{St} = \sum_f \lambda^f_{St} F^f_t + u_{St}. \tag{109}$$

Hence, if all factors with $\lambda^f_{St}$ possibly non-zero are observables and exogenous, we can measure the $\lambda^f_{St}$ 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.[76]

*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, (45) 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 ($y_{[k],t} = \sum_i \omega_{[k]i,t} y_{it}$, $\alpha^f_{[k]} = \sum_i \omega_{[k]i} \alpha^f_i$). 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 $var(u_{it}) \simeq 0$).

## D.4 When we have disaggregated data for both the demand and the supply side

When we have disaggregated data for both the demand and the supply side, we can refine the "exclusion restriction". So far we assumed that $\mathbb{E}[u_{it}\varepsilon_t] = 0$, i.e. no covariance between idiosyncratic demand and supply shock. If that's not the case, we can also decompose each supply with a factor model:

$$y^k_{it} = \phi^k p_t + \lambda^k_i \eta^k_t + u^k_{it}, \tag{110}$$

for type $k = s, d$ for supply and demand. We allow $\mathbb{E}[u^s_{it} u^d_{it}]$ to be nonzero: for instance, if the US has a "fracking shock" that affects both supply and demand, it will be captured by both $u^s_{it}$ and $u^d_{it}$ for $i = \text{USA}$.

The price $p_t$ adjusts so that supply equals demand, $y^s_{S^s t} = y^d_{S^d t}$ (where $S^d_i$ (resp. $S^s_i$) is the average fraction of demand (resp. supply) accounted by country $i$), i.e.

$$p_t = \frac{u^d_{S^d} - u^s_{S^s} + \lambda^d_{S^d} \eta^d_t - \lambda^s_{S^s} \eta^s_t}{\phi^s_{S^s} - \phi^d_{S^d}} \tag{111}$$

Then, we have two GIVs, based on supply and demand respectively:

$$z^k_t := \Gamma^{k\prime} y^k_t = u^k_{\Gamma^k t}, \tag{112}$$

for $k = s, d$ (with $\Gamma^k = S^k - E^k$ in the basic case $\lambda^k = \iota$, and $\Gamma^k = Q^{\lambda^k} S^k$ in the general case). We can also form the difference:

$$z^{d-s}_t := z^d_t - z^s_t. \tag{113}$$

Now, assume $\mathbb{E}[u^k_{it} \eta^{k\prime}_t] = 0$ for $k, k'$ in $\{s, d\}$. Then we have $\mathbb{E}[(y^k_{Et} - \phi^k p_t) z_t] = 0$ for $z_t$ equal

---

[76]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.

to either $z_t^s$ or $z_t^d$, or some combination of them. The optimal instrument is $z_t = z_t^d - z_t^s$, as this is the most correlated with the price (111) (this generalizes the reasoning of Proposition 3). We can also use an overidentification test like in Section 2.4, based on the those two GIVs based on supply and demand.

If we assume only that $\mathbb{E}\left[z_t^\ell \eta_t^k\right] = 0$ for some $(k, \ell)$, we can identify $\phi^k$ via $\mathbb{E}\left[\left(y_{Et}^k - \phi^k p_t\right) z_t^\ell\right] = 0$.

## D.5   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

$$\tau\left(x\right) = x 1_{|x| \geq b}$$

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 (11) still works. Likewise, in the more complex supply and demand case, the IV relations (31) and (33) still hold.

Furthermore, the OLS estimates still hold. The key is that we can write:

$$u_{\Gamma t} = z_t + z_t^<,$$

where $\tau^<\left(x\right) = x 1_{|x| < b}$, and $z_t^< = \sum_i \tau^<\left(S_i \check{u}_{it}\right)$, 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.

## D.6   When the researcher assumes too much homogeneity

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

The model (24)-(25) 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

$$p_t = \frac{u_{St} + \lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi^d_S}. \tag{114}$$

In this thought experiment, the econometrician assumes identical elasticities of demand across countries, $\phi^d_i = \phi^d$. He runs a panel model for $y_{it} - y_{Et}$, and we assume that it's large enough that he can extract $\eta_t$, successfully.[77] The GIV (we use the notation $Z_t$ rather than $z_t$ to denote the GIV before controls by $\eta_t$) is then

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

so

$$Z_t = \frac{1}{\psi} u_{\Gamma t} + \lambda^Z \tilde{\eta}_t, \qquad \frac{1}{\psi} = \frac{\phi^s - \phi^d_E}{\phi^s - \phi^d_S}, \tag{115}$$

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

Hence, when we run the first stage

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

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 $\phi^s$. Likewise, the IV on the elasticity of demand will give $\phi^d_E$.

In the polar opposite case where $\eta_t$ cannot be estimated or controlled for, then the simple procedure becomes biased, however, as (115) shows. To fix it, one can estimate the model with non-parametric coefficients (Section D.2).

## D.7 Sporadic factors

A potential issue is that of a "sporadic factor", i.e. a factor $\eta_t$ that affects a few actors special ways, but is not recurrent. An example would be a one-off policy announcement by the European

---

[77]One of the factors, formally, will be $p_t$. We assume that it is not included in the vector of factors $\eta_t$.

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 $\mathcal{S}_t = b_{(2)t}$.[78] Over the entire sample, we might remove the days with anomalously high sporadicity statistics, e.g. in the top 5% by that metric.

## D.8  Nonlinear GIV

We imagine a nonlinear GIV. Suppose that instead of the simple $s_t = \phi^s p_t + \varepsilon_t$ (equation (25)) we have a more complex

$$s_t = \Phi\left(p_t, \phi^s\right) + \varepsilon_t \tag{116}$$

for $\Phi$ a nonlinear function. We can use the moment:

$$\mathbb{E}\left[\left(s_t - \Phi\left(p_t, \phi^s\right)\right) z_t\right] = 0 \tag{117}$$

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

## D.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.[79]

### D.9.1  Logit demand

The utility that household $h$ derives from product $i$, for $i = 0, ..., N$, is given by[80]

$$U_{hit} = \delta_{it} + e_{hit},$$
$$\delta_{it} = -\gamma p_{it} + \beta' x_{it} + \alpha_i + \xi_{it},$$

---

[78]We could also sum over the most active $K$ entities, excluding the most active one.

[79]We thank Robin Lee, Alex MacKay, and Ariel Pakes for very helpful feedback on this section.

[80]We use the log price, $p_{it}$, instead of the price, $P_{it}$, in the formulation of $\delta_{it}$ to simplify some of the expressions, but the basic logic extends to the case where $\delta_{it}$ depends on $P_{it}$.

where $e_{hit}$ follows a Type-1 extreme-value distribution, $p_{it}$ denotes the log price, $x_{it}$ observable characteristics, and $\mathbb{E}[\xi_{it}] = 0$. 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 $s_{it} = \mathbb{P}(U_{hit} > \max_{j \neq i} U_{hjt})$, 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

$$\max_{P_{it}} Q_{it}(P_{it} - C_{it}),$$

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 $\beta$, it is commonly assumed that $\mathbb{E}[x_{it}\xi_{it}] = 0$ and we maintain this assumption. However, as prices respond to demand shocks, $\xi_{it}$, we cannot assume $\mathbb{E}[p_{it}\xi_{it}] = 0$. 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 therein).[81] The resulting moment is $\mathbb{E}[z_{it}^{BLP}\xi_{it}] = 0$.[82] 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).

---

[81]For other recent advances to construct instruments, see Sweeting (2013) and MacKay and Miller (2019).

[82]If a firm offers multiple products, the average of characteristics of other products produced by the same firm can be used as well.

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,

$$\xi_{it} = \eta_t + u_{it}, \tag{118}$$

which can be extended to allow for heterogeneous exposures, i.e. replacing $\eta_t$ by $\lambda_i \eta_t = \sum_k \lambda_i^k \eta_t^k$. Also, we assume for simplicity that $\eta_t$ 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:

$$z_{it} = \sum_{j:j \neq i} \bar{s}_{j,t-1} u_{jt}, \tag{119}$$

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

$$\mathbb{E}\left[z_{it}\xi_{it}\right] = 0, \tag{120}$$

which identifies $\gamma$. Remember that we use $\mathbb{E}\left[x_{it}\xi_{it}\right] = 0$ to identify $\beta$.

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 $s_{it}$ falls, 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 (119) is as follows. Recall that in this simple model the demand elasticity is

$$\epsilon_{it} = \gamma(1 - s_{it}),$$

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 $\epsilon_{it}$, is

$$\sum_{j:j \neq i} \frac{\partial \log s_{it}}{\partial u_{jt}} u_{jt} = -\sum_{j:j \neq i} s_{jt} u_{jt}. \tag{121}$$

Hence, shocks to companies with larger market shares have a larger impact.

### D.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 $\Theta$. 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 $\delta_{it}$ from the market shares using a contraction mapping (see Nevo (2000) for an introduction). With $\delta_{it}$ in hand, we form moment conditions as before to estimate $(\theta, \Theta)$.

To construct a GIV instrument in this model, one can also use (119) 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 suggesting forming

$$z_{it} := \sum_{j: j\neq i} s_{j,t-1}^i u_{jt}, \tag{122}$$

where $s_{j,t}^i$ is

$$s_{j,t}^i := -\frac{\partial \log s_{it}}{\partial u_{jt}}. \tag{123}$$

Indeed, in the homogeneous elasticity case, $s_{j,t}^i = s_{jt}$. 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 $s_{j,t}^i$ will be high, and $u_{jt}$ receives a high weight in the firm-$i$ specific GIV $z_{it}$.

## D.10   When the variance-covariance matrix of the $u$'s is estimated

In some cases, we'll want to identify the matrix $V^u$. We discuss this in Section 4.2. We add some thoughts on extensions here.

A simple sufficient condition is the following.

**Assumption 3** (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 3(a) could be relaxed in number of ways.[83] 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.

Assumption 3(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.[84,85] We now show how to do that in the most basic (and useful) case.

### D.10.1   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 $\sigma_i^2 = var\,(u_{it})$. We only measure $\check{u}_{it} = u_{it} - u_{Et}$. Here's a bit of algebra to recover $\sigma_i^2$.

We define $\sigma_E^2 := \frac{1}{N}\sum_i \sigma_i^2$. We have:

$$var\,(u_E) = \frac{\sigma_E^2}{N}, \qquad var\,(\check{u}_{it}) = \sigma_i^2\left(1 - \frac{2}{N}\right) + \frac{\sigma_E^2}{N} \tag{124}$$

---

[83]However, relaxations of Assumption 3 will still need to ensure some restrictions on the space of variance-covariances allowed.

[84]If we estimate $V^u$ via $V^{\check{u}}$, we need: $\frac{1}{2}(N-r)(N-r+1) \geq N$ as the projection $Q$ on a space of dimensions $N - r$ leaves only $\frac{1}{2}(N-r)(N-r+1)$ 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 $V^y$.

[85]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 not extraneous observable factors. Matrix $V^y$ gives $\frac{1}{2}N(N+1)$ parameters. We want to estimate: the $n_{V^u}$ (equal to $N$ if we assume a diagonal matrix $V^u$, equal to 1 is we assume homoskedasticity), parameter $M$, and matrix $V^\eta$ (which has $\frac{1}{2}r(r+1)$ degrees of freedom). If the factor model is parametric with $X_{it}$ a $r-$dimensional vector, then we need: $\frac{1}{2}N(N+1) \geq 1 + \frac{1}{2}r(r+1) + n_{V^u}$.

which implies:

$$\sigma_{\hat{u},E}^2 := \frac{1}{N}\sum_i var\left(\check{u}_{it}\right) = \sigma_E^2\left(1 - \frac{1}{N}\right) \tag{125}$$

So, we can recover

$$\sigma_i^2 = \frac{var\left(\check{u}_{it}\right) - \frac{\sigma_{\hat{u},E}^2}{N-1}}{1 - \frac{2}{N}} \tag{126}$$

If the estimation has a small error, the optimal thing to do is to define $E$ as $E^0$ in (518), and then $\Gamma = S - E$.

### D.10.2 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 (21)). Suppose that we assume 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 we have with the matrix $H_{ij} := Q_{ij}^2$, we have:[86]

$$\check{D} = HD \tag{127}$$

Hence, when $H$ is invertible (which is typically true), we can recover the variance of $\sigma_{u_i}^2 = D_i$ by $D = H^{-1}\check{D}$.[87] One disadvantage of this procedure is that it does not guaranty that $D_i$ is positive (indeed, in (126), the right-hand side is not necessarily positive).

### D.10.3 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 (21), 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 $m \times n$.

---

[86]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 $V^{\check{u}} = QV^uQ'$,

$$\check{D}_i = e_i'V^{\check{u}}e_i = e_i'QV^uQ'e_i = \sum_j e_i'Qe_je_j'Q'e_iD_j = \sum_j Q_{ij}^2 D_j = \sum_j H_{ij}D_j.$$

[87]Numerically, with in finite samples, we can get negative $\sigma_{u_i}^2$ (see (126)). So, one can imagine variants that guaranty positivity, e.g. adding a winsorization step, $D_i := \max\left(D_i, \xi \text{median}\left(D_j\right)\right)$ for a low $\xi$ such as $\xi = 0.5$. Another procedure is to do

$$\min_D \left\|\check{D} - HD\right\|^2 \text{ subject to } \min D_i \geq \xi\text{median}(D_j)$$

or another constraints, 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 those constraints.

We do a singular value decomposition of $Q$. We call its rank $k$. We can write:

$$Q = UDV^*$$ (128)

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

$$\bar{u}_t := B\breve{u}_t$$ (129)

which is a $k$ dimensional vector that gathers the useful $k$ degrees of freedom in $\breve{u}_t$. While $V^{\breve{u}}$ 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, $V^{\bar{u}} = MV^u M'$. 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 $V^u = Diag\left(\sigma^2_{u_1}, \ldots, \sigma^2_{u_n}\right)$. We assume that

$$\frac{1}{2}k(k+1) \geq n$$

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 $2\pi$, and use the notation $|W^{\bar{u}}|$ for the determinant of $W^{\bar{u}}$),

$$\begin{aligned} \ell &:= \frac{2L}{T} = -\mathbb{E}_T\left[\mathrm{tr}\left(\bar{u}'_t W^{\bar{u}} \bar{u}_t\right)\right] + \ln |W^{\bar{u}}| = -\mathrm{tr}\left(\mathbb{E}_T\left[\bar{u}_t \bar{u}'_t\right] W^{\bar{u}}\right) + \ln |W^{\bar{u}}| \\ \ell &= -\mathrm{tr}\left(V^{\bar{u},e} W^{\bar{u}}\right) + \ln |W^{\bar{u}}| \end{aligned}$$ (130)

We recover $V^u$ by maximizing numerically $\ell$ over the $\sigma^2_{u_i} > 0$.

## D.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 $\eta_t$ dimension $r$, but we only observe:

$$Y_t = GY_t^*$$

for a known matrix $G$ such that $G^2 = G$ (e.g. $G$ could be a $Q$ matrix as in (21)). 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 $\eta_t$.

**Projecting $Y_t$ in into a lower-dimensional $y_t$** We do an eigendecomposition of $G$. Calling $K$ the rank of $G$, we can write:

$$G = A^{-1}DA \tag{131}$$

where $D = \begin{pmatrix} I_K & 0 \\ 0 & 0 \end{pmatrix}$, and $A^{-1}$ is the matrix whose columns are the corresponding eigenvectors of $G$.[88],[89] We introduce $\pi := \begin{pmatrix} I_K & 0 \end{pmatrix}$ and set $B := \pi A$, both with dimension $K \times N$. . We have[90]

$$D = \pi\pi', \qquad \pi D = \pi, \qquad B = \pi A, \qquad BG = B, \qquad A^{-1}\pi'B = G \tag{132}$$

We call

$$y_t := BY_t \tag{133}$$

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:

$$y_t = \lambda\eta_t + v_t \tag{134}$$

**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 $\lambda$, unless $V^v$ is proportional to the identity matrix.[91] Ideally, we'd like then to estimate the PCA on $Ly_t = L\lambda\eta_t + Lv_t$, for $L = k(V^v)^{-1/2}$ (with $k$ an

---

[88]We order the columns of $A^{-1}$ with first the eigenvectors with eigenvalue 1, then those with eigenvalue 0.

[89]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 21. Then, calling $H = W^{-1/2}$,then $\tilde{G} = H^{-1}GH$ is a symmetric matrix, so that Matlab recognizes that the eigenvectors should be all 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 $\tilde{G}$ with the sam eigenvalue $k$. This way, we recover real eigenvectors of $G$.

[90]Indeed, $BG = (\pi A)\left(A^{-1}DA\right) = \pi DA = \pi A = B$, and

$$A^{-1}\pi'B = A^{-1}\pi'\pi AG = A^{-1}DAG = GG = G.$$

[91]Indeed, if $V^v$ is proportional to the identity matrix, then $V^y\lambda$ is proportional to $\lambda$, so that the column vectors of $\lambda$ are eigenvectors of $V^y$ (and PCA, which extracts the eigenvectors, will successfully recover $\lambda$). But this is not the case if $V^v$ is not proportional to the identity matrix.

constant) because then the covariance of residuals $Lv_t$ will be proportional to the identity, and the PCA will correctly recover $L\lambda$. But we do not know $V^v$: we need to estimate it.

This motivate the following algorithm. For an invertible matrix $V$ of dimensions $K \times K$, call[92]

$$J(V) := V^{-1/2}\sqrt{\frac{|\text{tr}\,(V)|}{K}} \tag{135}$$

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:

$$\text{tr}\,(J(V)VJ(V)) = \text{tr}\,V \tag{136}$$

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 $y_{nt} = L_n\lambda\dot\eta_t + \dot v_{nt}$ and get the residuals, $\dot v_{nt}$.

2. We define $\ddot u_t := A^{-1}\pi' L_n^{-1}\dot v_{nt}$. We note that $\dot v_{nt} = Q^{L_n\lambda, I_K}L_n v_t$ and $v_t = Bu_t$, so that $\ddot u_{nt} = q_n u_t$ with matrix $q_n := A^{-1}\pi' L_n^{-1}Q^{L_n\lambda, I_K}L_n B$. As the rank of $\lambda$ is $r$, matrix $q_n$ as dimension $N \times N$ but only rank $K - r$. From $V^{\ddot u_n}$ we obtain $V_n^u$, using the procedure in Section D.10.2, observing that $\ddot u_{nt} = q_n u_t$.

   (a) We can use the procedure in Section D.10.2, 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 D.10.3 which does guaranty that $\sigma_{u_i}^2$ is positive.[93]

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 ($\left\|\frac{L_n V_n^v L_n}{\text{tr}(L_n V_n^v L_n)/K} - I\right\|$ should be small).

The procedure returns $V^{u,e} = V_n^u$. It also returns the factors $\dot\eta_t$ (up to a rotation, as usual). We also obtain an estimator of $\Lambda$:

$$\Lambda^e = A^{-1}\pi' L_n^{-1}(L_n\lambda) \tag{137}$$

where $L_n\lambda$ is estimated from the PCA in Step 1.[94]

---

[92]Given a positive definite matrix $V$, and $\alpha$ a scalar, $V^\alpha$ is defined as follows. Do an eigendecomposition $V = P\Delta P^{-1}$, where the columns of $P$ are the eigenvectors of $V$, and $\Delta = Diag(\Delta_i)$ is the diagonal matrix with $V$'s eigenvalues. Then, we define $V^\alpha = PDiag(\Delta_i^\alpha)P^{-1}$. We apply this to $\alpha = -\frac{1}{2}$.

[93]We can alternatively use the same procedure to $\dot v_{nt} = Q^{L_n\lambda, I_K}L_n Bu_t$ as the observed vector, rather than going through the higher-dimensional $\ddot u_t$ as the measured vector.

[94]Indeed, then at convergence $\Lambda^e = A^{-1}\pi'\lambda = A^{-1}\pi'B\Lambda = G\Lambda = GG\Lambda^* = G\Lambda^* = \Lambda$, by (132).

## D.12 Full recovery when different factors have different "size" weights

In the basic model, we can identify $\alpha^f$, $M = \frac{1}{1 - \sum_f \lambda^f \alpha^f}$, but not $\lambda^f$.

We give some conditions under which we can actually also identify the $\lambda^f$ (in addition to $\alpha^f$ 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).

Here we take the basic set up as in Section 3.1, 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 $\eta$ all have an impact of 1:

$$y_{it} = u_{it} + \sum_f \lambda^f F_t^f + \eta_t^y, \tag{138}$$

$$F_t^f = \alpha^f y_{S^f, t} + \eta_t^f. \tag{139}$$

This implies

$$y_t = u_t + \iota \sum_f \lambda^f F_t^f + \iota \eta_t^y = u_t + \iota \sum_f \lambda^f \left( \eta_t^f + \alpha^f S^{f'} y_{it} \right) + \iota \eta_t^y.$$

With "$\varepsilon^k$" denoting some combination of the various $\eta$'s, and as usual $M = \frac{1}{1 - \sum_f \alpha^f \lambda^f}$,

$$y_t = \left( I - \iota \sum_f \lambda^f \alpha^f S^{f'} \right)^{-1} \left( u_t + \iota \varepsilon_t^1 \right)$$

$$= \left( I + M \iota \sum_f \lambda^f \alpha^f S^{f'} \right) \left( u_t + \iota \varepsilon_t^1 \right)$$

$$y_t = u_t + M \iota \sum_f \lambda^f \alpha^f u_{S^f, t} + \iota \varepsilon_t^y, \tag{140}$$

i.e., since that $F_t^f = \eta_t^f + \alpha^f y_{S^f, t}$ this gives:

$$F_t^f = \alpha^f \left( u_{S^f, t} + M \sum_g \lambda^g \alpha^g u_{S^g, t} \right) + \varepsilon_t^f. \tag{141}$$

Hence, suppose that we extracted the $\check{u}_{it} = u_{it} - u_{Et}$ (following our usual procedure). Then, we form

$$z_{\Gamma^f t} := S^{f'} \check{u}_t = u_{S^f t} - u_{Et}. \tag{142}$$

Then, regressing $F_t^f$ on the various $z_{\Gamma^g t}$

$$F_t^f = \sum_g b_g^f z_{\Gamma^g t} + \varepsilon_t^{f,1} \tag{143}$$

78

(for $\varepsilon^{f1}$ some residual noise) yields a regression coefficient:

$$b_g^f = \alpha^f \left(1_{f=g} + M\lambda^g \alpha^g\right). \tag{144}$$

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 $f$), hence $M$. Hence, we obtained $\alpha^f M$, $M$ and $\lambda^g \alpha^g$ — hence all quantities: $\alpha^f, \lambda^f, M$.

## D.13 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 $\eta_t$ 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 $\phi$ is the Frisch elasticity of labor supply. So, as $C_t = \Lambda_t L_t$, labor supply is $\hat{L}_t = \phi\left(\hat{\Lambda}_t - \eta_t^L\right),$[95] and GDP is $\hat{Y}_t = \hat{L}_t + \hat{\Lambda}_t$, i.e.

$$\hat{Y}_t = m\hat{\Lambda}_t - \phi\eta_t^L, \qquad m = 1 + \phi \tag{145}$$

We seek to find the "GDP multiplier" $m = 1 + \phi$, so that a TFP of 1 percent translates into a GDP increase of $m$ percent.[96]

This is potentially a complicated problem, as for instance, in the Long and Plosser (1983) case with input-output matrix $A$, output changes are $\hat{Y}_t = (I - A)^{-1}\hat{\Lambda} + \hat{L}$, 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:

$$\hat{\Lambda}_{it} = \lambda_i \eta_t^\Lambda + u_{it}. \tag{146}$$

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).

---

[95]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.

[96]If more than one factor change, $m$ has the broader interpretation of a multiplier between TFP and 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 productivities $\hat{\tilde{\Lambda}}_{it}$ growth (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

$$\hat{\tilde{\Lambda}}_{it} = \tilde{\lambda}_i \tilde{\eta}_t^\Lambda + \tilde{u}_{it}. \tag{147}$$

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.

## D.14 When the influence matrix is not proportional to size

### D.14.1 Position of the problem

Suppose a model

$$y_{it} = \gamma \sum_j G_{ij} y_{jt} + \lambda_i \eta_t + u_{it}, \tag{148}$$

i.e.

$$y_t = \gamma G y_t + \Lambda \eta_t + u_t, \tag{149}$$

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 $\gamma$.

### D.14.2 A simple approach

We study the model(149), where the factor loading $\Lambda$ (an $N \times r$ matrix) is not necessarily equal to $\iota$ (but we keep imposing that the $\Lambda$ spans $\iota$, i.e. there is a $q$ such that $\iota = \Lambda q$). As before, $\eta_t$ is a

low-dimensional vector of factors.

First, we suppose that we have a first estimate of $\gamma$, which we call $\gamma^e$. We will later iterate on it. Then, we form:

$$\tilde{y}_t(\gamma^e) := (I - \gamma^e G)\, y_t. \tag{150}$$

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 $\Lambda$ and $W = (V^u)^{-1}$. We introduce $Q$ as in (21) so that $Q\Lambda = 0$ and set

$$\check{u}^e := Q\tilde{y}_t(\gamma^e),$$

so that at $\gamma^e = \gamma$, $\check{u}_t^e = Qu_t$. We observe that $Gy_t = G(I - \gamma G)^{-1} u_t + B\eta_t$ for some $B$. This suggests the following procedure.

We define the GIV $z_t$ as a vector (with dimension $N$):

$$z_t := G(1 - \gamma^e G)^{-1} \check{u}_t^e = G(1 - \gamma^e G)^{-1} Q(I - \gamma^e G)\, y_t. \tag{151}$$

Indeed that $z_t$ will imitate the movements of the idiosyncratic shocks on $y_t$.

Our key moment is:[97]

$$\mathbb{E}\left[(y_t - \gamma Gy_t)' W^u z_t\right] = \delta, \tag{153}$$

where $\delta$ is a discrepancy term

$$\delta := \operatorname{tr}\left(QG(1 - \gamma G)^{-1}\right). \tag{154}$$

This yields an estimate of $\gamma$.[98]

The discrepancy term $\delta$ 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, (153) generalizes out 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 $\delta = 0$.

---

[97]Here is the proof. At the right estimator $\gamma = \gamma^e$,

$$z_t = G(1 - \gamma G)^{-1} Q(I - \gamma G)\, y_t = G(1 - \gamma G)^{-1} Qu_t = Hu_t, \tag{152}$$

for $H = G(1 - \gamma G)^{-1} Q$. We also have $y_t - \gamma Gy_t = \Lambda \eta_t + u_t$. This implies that

$$
\begin{aligned}
\mathbb{E}\left[(y_t - \gamma Gy_t)' W^u z_t\right] &= \mathbb{E}\left[(\Lambda \eta_t + u_t)' W^u Hu_t\right] = \mathbb{E}\left[\operatorname{tr}(u_t' W^u Hu_t)\right] = \mathbb{E}\left[\operatorname{tr}(W^u Hu_t u_t')\right], \\
&= \operatorname{tr}(W^u HV^u) = \operatorname{tr}(HV^u W^u) = \operatorname{tr}(H).
\end{aligned}
$$

[98]We have a fixed point: an initial $\gamma^e$ gives an estimate of $\gamma$; that's then the new estimate $\gamma^e$, and we re-iterate the process, until convergence.

## D.15 Identification of social interactions and the reflection problem

Superficially, there seems to be a contradiction between Section 8.2'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:

$$\boldsymbol{y}_t = \beta \boldsymbol{G} \boldsymbol{y}_t + \gamma \boldsymbol{x}_t + \delta \boldsymbol{G} \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \tag{155}$$

with $\mathbb{E}\left[\boldsymbol{\varepsilon}_t | \boldsymbol{x}_t\right] = 0$. 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 $\iota$ 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 8.2. 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:

$$\mathbb{E}\left[\boldsymbol{y}_t | \boldsymbol{x}_t\right] = \beta \boldsymbol{G} \mathbb{E}\left[\boldsymbol{y}_t | \boldsymbol{x}_t\right] + \gamma \boldsymbol{x}_t + \delta \boldsymbol{G} \boldsymbol{x}_t, \tag{156}$$

where all the noise has been averaged out.

Indeed, we do impose some structure, namely:

$$\varepsilon_{it} = \eta_t + u_{it}, \qquad u_{it} \text{ i.i.d., orthogonal to } \eta_t, \tag{157}$$

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[99]), which seems to leads to the impossibility of identification in Bramoullé et al. (2009). This is formalized here.

**Proposition 6** (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* (157). *Suppose also the existence of two n-dimensional vectors $S$ and $\Gamma$ satisfying*

$$G'S = S, \qquad G'\Gamma = 0, \qquad \iota'S \neq 0, \qquad \Gamma'S \neq 0. \tag{158}$$

*Then GIV is possible in that setup, i.e. with the GIV $z_t = \Gamma'y_t$, we can identify the coefficients $(\beta, \gamma, \delta)$.*

---

[99] It can be reduced to that case by rescaling $H = b_0 + b_1 G$ with the right coefficient, with $H^2 = H$.

*In our basic setup, we had $S_i$ the relative sizes, and $G = \iota S'$, $\Gamma = S - \frac{\iota}{N}$. Hence (158) is an abstract generalization of our concrete conditions.*

Hence, in many situations of interest we can be quite confident that condition (158) is satisfied.[100]

There could be another way to analyze social influence with a matrix of influence, like in Section D.14.

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 6**   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

$$y_{Et} = E'y_t, \qquad y_{St} := S'y_t,$$

which are our generalized "equal weighted" and "value weighted" averages – for more abstract setting. Then, premultiplying (155) by $\Gamma'$ gives:

$$z_t := \Gamma'y_t = \delta x_{\Gamma t} + u_{\Gamma t}.$$

Hence, estimating this by OLS we can obtain $\gamma$, and $var\,(u_{\Gamma t})$, so that we obtain also $\sigma_u^2$. Next,

$$y_E = \beta y_S + \gamma x_E + \delta x_S + \eta + u_E,$$

so that

$$\mathbb{E}\left[(y_{Et} - \beta y_{St} - \gamma x_{Et} - \delta x_{St})'\,(z, x_{St})\right] = (\mathbb{E}u_{Et}u_{\Gamma t}, 0)\,. \tag{159}$$

The right-hand side is known, as $\mathbb{E}u_{Et}u_{\Gamma t} = E'\Gamma\sigma_u^2$, which is known. So, we have two unknowns $\beta$, $y$ and two equations: we can solve the system. The condition $\Gamma'S = 0$ ensures that $\mathbb{E}\,[y_{St}z_t] \neq 0$. $\square$

## D.16   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 the estimate elasticity of demand in partially segmented markets. The first problem uses the second one.

---

[100]As $G^2 = G$, one can always find vectors $\Gamma, S$ 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.

As a motivation, imagine that industry $i$ has the CES production function[101]

$$Q_{it} = B_{it} \left( K_{it}^{\frac{\phi_i - 1}{\phi_i}} + A_{it}^{\frac{1}{\phi_i}} L_{it}^{\frac{\phi_i - 1}{\phi_i}} \right)^{\frac{\phi_i}{\phi_i - 1}} \tag{160}$$

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:

$$\frac{L_{it}}{K_{it}} = A_{it} \left( \frac{W_{it}}{R_t} \right)^{-\phi_i} \tag{161}$$

We'd like to estimate the elasticity of substitution $\phi_i$ between capital and labor. This is the wage elasticity of demand. GIV 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 (161) as:

$$y_{it}^d = \phi_i^d p_{it} + C_{it} + \lambda_i^d \eta_t + u_{it}^d \tag{162}$$

where $C_{it}$ is a control, and as usual vector $\eta_t$ 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 it $C_{it}$, but this is not important.

Now, log labor supply is modeled as:

$$y_{it}^s = \phi_i^s p_{it} - \psi_i p_{St} + \lambda_i^s \eta_t + u_{it}^s \tag{163}$$

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 $i$:

$$p_{it} = \frac{\psi_i p_{St} + u_{it}^d - u_{it}^s + \left( \lambda_i^d - \lambda_i^s \right) \eta_t}{\phi_i^s - \phi_i^d} \tag{164}$$

i.e.

$$p_{it} = \gamma_i p_{St} + v_{it} + \lambda_i^p \eta_{it} \tag{165}$$

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 (165) is a standard GIV. In the general case, we can estimate $\gamma_i$ as in Section (D.2).[102]

So, we obtain $\gamma_i$ and $v_{it}^e$ (the proxy for $v_{it}$) in (165). We also form $z_{it} = z_t - \Gamma_i^u v_{it}$ as in (453). This is the GIV formed of the idiosyncratic shock of all industries but industry $i$. We will use the

---

[101]We thank Julieta Caunedo for prompting us to think about this identification problem.

[102]This procedure is much simplified if the $\gamma_i$ and $\lambda_i^p$ 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 $b = \frac{1}{1 - \gamma}$.

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 (162), and instrument for $p_{it}$ using the $z_{it}$

$$p_{it} = b_i z_{it} + \varepsilon_{it}^p \tag{166}$$

we estimate $b_i$, and define $p_{it}^e = b_i z_{it}$ the price in industry $i$ instrumented by the changes in other industries. We use the estimated $\eta_t^e$ as controls, and run

$$y_{it}^d = \phi_i^d p_{it}^e + C_{it} + \lambda_i^d \eta_t^e + u_{it}^d \tag{167}$$

which yields a consistent estimate $\phi_i^d$ of the labor demand.

**Extension**   We can extend (163) to

$$y_{it}^s = \phi_i^s p_{it} - \psi_i \sum_j G_{ij} p_{jt} + \lambda_i^s \eta_t + u_{it}^s \tag{168}$$

where the "influence" matrix $G_{ij}$ capture 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 8.1 and D.14 to handle this case.

# E   A Bayesian perspective on GIVs

We will see that, under conditions of Gaussianity, our estimators are basically the MLE. As variables may not be Gaussian, we keep the general exposition (showing identification) free of distributional assumptions. If we assume that variables $(u_{it}, \eta_t)$ are Gaussian, then a Bayesian analysis can be performed. We detail it here.

## E.1   The general model: Bayesian version

Here we treat the general model of Section 3.1, in the case where the $\lambda_{it}$ are the same, and equal to 1, and all factors are observed (except the $\eta_t^y$, as in $y_{it} = \sum_f \lambda^f F_t^f + u_{it} + \eta_t^y$). The general case with heterogeneous loadings is done later, in Section E.5, and uses much the same ideas.

The data $D$ is $D = \left( y_t, F_t^f \right)_{f=1\dots d_F, t=1\dots T}$, made of i.i.d. draws from a fixed distribution. To simplify the notations, we'll just denote by $f$ the collection of all variables corresponding to factors (without explicitly mentioning that $f = 1 \dots d_F$).

The solution of the system features:

$$
\begin{aligned}
y_{St} - M y_{\Gamma t} &= b^y \varepsilon_t, \\
F_t^f - \alpha^f M y_{\Gamma t} &= b^f \varepsilon_t,
\end{aligned}
$$

for some vector $b^y$, $b^f$, and $\varepsilon_t := \left( u_{Et} + \eta_t^y, \eta_t^f \right)$.

Hence, we form: $\theta = \left( M, \alpha^f M \right)$; $\omega^V$ a parametrization of the relevant variance matrices – so the variance matrix $V^u$ depends on $\omega^V$ (for instance, if $V^u$ is diagonal, $\omega^V$ is the vector of its diagonal terms); $\omega = \left( \theta, \omega^V \right)$ the set of parameters to be estimated; $E\left( \omega \right) = \frac{V^u\left(\omega^V\right)^{-1}\iota}{\iota' V^u\left(\omega^V\right)^{-1}\iota}$ the corresponding quasi-equal weights vector, and form the key quantities:

$$Y_t\left( \omega \right) := \left( y_{St} - M y_{\Gamma(\omega),t}, F_t^f - \alpha^f M y_{\Gamma(\omega),t} \right). \tag{169}$$

We also keep track of

$$\check{y}_{it}\left( \omega \right) = y_{it} - y_{E(\omega),t} \tag{170}$$

and stack those two vectors together as $X_t\left( \omega \right)$, which contains all our information:

$$X_t\left( \omega \right) = \left( Y_t\left( \omega \right), \check{y}_t\left( \omega \right) \right). \tag{171}$$

The key "trick to tractability" is to transform the data into that $X_t$.

There is an invertible matrix $A\left( \theta \right)$ such that $D_t = A\left( \theta \right) X_t$. Hence, there is no loss of information in using $X_t$ as "conveniently processed" data, rather than the "unprocessed" data $D_t$. Hence, instead of $\ln \mathbb{P}\left( D_t | \omega \right)$, we'll consider

$$\ln \mathbb{P}\left( X_t | \omega \right) = \ln \mathbb{P}\left( D_t | \omega \right) + \ln \left| A\left( \theta \right) \right|. \tag{172}$$

The Jacobian $\left| A\left( \theta \right) \right| := \det A$ is independent of all parameters $\omega$.[103] Hence it can be discarded as a constant in the calculations.

The key simplifying observation is that (under the correct model), $\mathbb{E}\left[ \check{y}_t y_{Et} \right] = 0$, so that $Y_t\left( \omega \right)$ and $\check{y}_t\left( \omega \right)$ have zero covariance. Hence, the log likelihood decouples, and we have

$$-2\ln \mathbb{P}\left( D_t | \omega \right) = Y_t'\left( \omega \right) V^Y\left( \omega^V \right)^{-1} Y_t\left( \omega \right) + \check{y}_t\left( \omega \right)' \left( V^{\check{y}}\left( \omega \right) \right)^{-1} \check{y}_t\left( \omega \right) + \ln \left| V^Y\left( \omega^V \right) \right| + \ln \left| V^{\check{y}} \right|. \tag{173}$$

As $\check{y}_t$ lives in a space of dimension $N - 1$ (as $E'\check{y}_t = 0$), the value of $V^{\check{y}}$ is understood as being of the corresponding dimensions, $(N - 1) \times (N - 1)$.

Now, imagine that $\omega^V$ has already been estimated, and do only the optimization w.r.t. $\theta$. That gives:

$$\min_\theta \mathbb{E}_T \left[ Y_t'\left( \theta, \omega^V \right) V^Y\left( \omega^V \right)^{-1} Y_t\left( \theta, \omega^V \right) \right]. \tag{174}$$

The first order conditions are:

$$\mathbb{E}_T \left[ \left( y_{\Gamma t}, 0 \right) \left( V^Y \right)^{-1} Y_t \right] = 0, \qquad \mathbb{E}_T \left[ \left( 0, y_{\Gamma t} \right) \left( V^Y \right)^{-1} Y_t \right] = 0,$$

---

[103]First, go from $X_t$ to $\tilde{D}_t = (F_t, y_{St}, \check{y}_t)$, which is upper triangular with 1 on the diagonal, so has determinant 1; second, go from $\tilde{D}$ to $D$, which is independent of the $\omega$.

i.e. (given that $0 = \mathbb{E}_T \left[ y_{\Gamma t} \left( V^Y \right)^{-1} Y_t \right] = \left( V^Y \right)^{-1} \mathbb{E}_T \left[ y_{\Gamma t} Y_t \right]$) we have $\mathbb{E}_T \left[ y_{\Gamma t} Y_t \right] = 0$, yielding

$$\mathbb{E}_T \left[ y_{\Gamma t} \left( y_{St} - M y_{\Gamma t}, F_t^f - \alpha^f M y_{\Gamma t} \right) \right] = 0. \tag{175}$$

Those are precisely the first order conditions of the OLS estimation:

$$\min_M \mathbb{E}_T \left[ \left( y_{St} - M y_{\Gamma t} \right)^2 \right], \qquad \min_{\alpha^f M} \mathbb{E}_T \left[ \left( F_t^f - \alpha^f M y_{\Gamma t} \right)^2 \right]. \tag{176}$$

Hence, our GIV is also the MLE estimator of $M, M\alpha^f$, when we have Gaussian distributions.

We can also go beyond MLE, and calculate full Bayesian posteriors. Then, the GIV gives an easy way to do finite-sample Bayesian updating. Assuming again for simplicity that we know the variance matrices, we have

$$\ln \mathbb{P} \left( \theta | D \right) = \ln \mathbb{P} \left( \theta \right) - \frac{1}{2} \sum_t Y_t \left( \theta \right) \left( V^Y \right)^{-1} Y_t \left( \theta \right) + K \left( D \right), \tag{177}$$

where $K \left( D \right)$ ensures that the probability sums to 1.

The rest of this section examines instantiations and variants of the general idea we just saw.

## E.2 The supply and demand model of Section 2.2

This model corresponds exactly to the general case, with a factor $F_t^f = p_t$, $p_t = \alpha^f y_{St} + \eta_t^f$ with $\alpha^f = \frac{1}{\phi^s}$ and $\eta^f = -\frac{\varepsilon}{\phi^s}$. Then, everything goes through.

## E.3 The basic example with self-loop of Section 8.2

We give a Bayesian treatment of this model of Section 8.2:

$$y_{it} = \gamma y_{St} + \eta_t + u_{it}.$$

We are given $D_t = y_t$. We wish to estimate $M = \frac{1}{1-\gamma}$. The vector of parameters of interest is $\theta = M$:

$$Y_t \left( \omega \right) = y_{St} - M y_{\Gamma(\omega^V), t}.$$

As in the general procedure, we set:

$$\check{y}_{it} \left( \omega \right) = y_{it} - y_{E(\omega^V), t} \tag{178}$$

and

$$X_t \left( \omega \right) = \left( Y_t \left( \omega \right), \check{y}_t \left( \omega \right) \right). \tag{179}$$

87

In the true model, we have $Y_t = M (u_{Et} + \eta_t)$, so[104]

$$-2 \ln \mathbb{P} (D_t | \omega) = \frac{Y_t^2 (\omega)}{\sigma_Y^2} + \check{y}_t (\omega)' \left(V^{\check{y}}\right)^1 \check{y}_t (\omega) + \ln \sigma_Y^2 + \ln \left|V^{\check{y}}\right|.$$

Suppose first that we know the variance terms. Then, the MLE is simply to do

$$\max_M \mathbb{E}_T \left[Y_t (\omega)^2\right],$$

which is the identification condition we used, and it corresponds to running the OLS

$$\min_M \mathbb{E}_T \left(y_{St} - M y_{\Gamma(\omega^V)t}\right)^2.$$

Next, for the estimation of the variance terms, we optimize on $\sigma_Y^2$, $V^{\check{y}}$. Asymptotically, that gives the true values.

## E.4 The basic example without loop of Section 2.1.

We now detail the Bayesian version of our example in Section 2.1:

$$y_{it} = \eta_t + u_{it}, \qquad p_t = \alpha y_{St} + \varepsilon_t.$$

We'd like to estimate $\alpha$ especially (or, in a Bayesian context, update our prior on $\alpha$). This example is actually a bit non-generic, as it endows the economist with a knowledge that $\lambda^f = 0$, which creates some subtle changes: it features the "recovered" factor $y_{Et}$, used as a regressor. We can also write $p_t = \alpha y_{\Gamma t} + \beta y_{Et} + \varepsilon_t^{\perp}$ for some $\beta$.

The data $D$ is a set of $D = (y_t, p_t)_{t=1...T}$, assumed to be i.i.d. draws from a fixed distribution.

We call $\theta = (\alpha, \beta)$ the set of "key" model parameters, and $\omega^V$, the parametrization of the variance-covariance matrix $V^{(u+\eta_t, \varepsilon)}$ (e.g. if we assume that $u$ is diagonal), the auxiliary parameter, and $\omega = (\theta, \omega^V)$ the full set of parameters. The correct value is $\omega^*$.

Given $y_t, p_t$, we form

$$Y_t (\theta) = p_t - (\alpha y_{\Gamma t} + \beta y_{Et})$$

and $X_t (\theta) = (Y_t (\theta), y_t)$. At the correct parameter $\omega^*$,

$$Y_t (\theta^*) = \varepsilon_t^{\perp},$$

which is defined in the analysis is the "enriched OLS estimator" that comes from running the regression: $p_t = \alpha y_{\Gamma t} + \beta y_{Et} + \varepsilon_t^{\perp}$. Hence, at the correct value, $Y_t$ and $y_t$ are uncorrelated. Call $V^X (\omega)$ the variance-covariance matrix of $X_t$.

---

[104]The inverse of $V^{\check{y}}$ is taken in the space of matrices operating on vectors orthogonal to $E$.

We can start the Bayesian analysis:

$$\mathbb{P}\left(\omega|D\right) \propto \mathbb{P}\left(D|\omega\right)\mathbb{P}\left(\omega\right)$$

and

$$\ln\mathbb{P}\left(D|\omega\right) = \sum_t \ln\mathbb{P}\left(D_t|\omega\right)$$

with

$$-2\ln\mathbb{P}\left(D_t|\omega\right) = \frac{Y_t\left(\theta\right)^2}{\sigma_{\varepsilon\perp}^2} + y_t'V^y\left(W\right)y_t + \ln\sigma_{\varepsilon\perp}^2 + \ln\left|V^y\left(W\right)\right|, \tag{180}$$

where here $|A|$ is the determinant of a matrix $A$.

Hence, the MLE estimator maximizes $\sum_t \ln\mathbb{P}\left(D_t|\omega\right)$ over $\omega = \left(\theta, W\right)$. The problem for $\theta$ separates as:

$$\min_{\alpha,\beta} \sum_t Y_t\left(\theta\right)^2,$$

i.e.

$$\min_{\alpha,\beta} \sum_t \left(p_t - \left(\alpha y_{\Gamma t} + \beta y_{Et}\right)\right)^2,$$

which is the "enriched OLS estimator" $p_t = \alpha y_{\Gamma t} + \beta y_{Et} + \varepsilon_t^\perp$. This shows that, with Gaussian distributions, the MLE is just our enriched GIV-OLS estimator.

Maximizing over the other parameters $W$ will allow to recover the variance matrix (including that of $\varepsilon_t, \eta_t$).

If we have a small sample, we can just update rather than do MLE. The above shows that the "simplifying trick" is to form that statistic $Y_t\left(\theta\right)$, which allows for an interpretable updating of the parameters. For simplicity, suppose that we know the value of $V^y\left(W\right)$, and $\sigma_{\varepsilon\perp}^2$.[105] However, we have a prior on $\theta = \left(\alpha, \beta\right)$, perhaps Gaussian. Then, our posterior after observing the data $D$ is:

$$\ln\mathbb{P}\left(\theta|D\right) = \ln\mathbb{P}\left(\theta\right) - \sum_t \frac{Y_t\left(\theta\right)^2}{2\sigma_{\varepsilon\perp}^2} + K\left(D\right),$$

where $K\left(D\right)$ ensures that the probability sums to 1.

---

[105]Otherwise, we can update our knowledge of those, which is standard though tedious to lay out.

## E.5 Heterogeneous loadings

### E.5.1 Bayesian model with heterogeneous loadings

Here we extend the basic model of this section to heterogeneous, nonparametric loadings. The model is

$$y_{it} = \sum_f \lambda^f F_t^f + \eta_t^y + \sum_{k=1}^K \lambda_i^k \eta_t^k + u_{it}, \tag{181}$$

where now there are $K$ unobserved factors, with unknown factors $\eta_t^k$, and non-uniform loadings $\lambda_i^k$. As in the more basic Section E.1, we assume the existence of a factor $\eta_t^y$ with uniform loadings (which can be taken to be uncorrelated with $\eta_t$), and factors $F_t^f$ are endogenous and observed. More compactly, we can write the model as:

$$y_t = \lambda^F F_t + \iota \eta_t^y + \lambda \eta_t + u_t. \tag{182}$$

We will now see how this case can be reduced to the one of Section E.1. We define the theoretical object:

$$\tilde{y}_{it} := y_{it} - \sum_{k=1}^K \lambda_i^k \eta_t^k. \tag{183}$$

Then, the results of Section E.1 apply to $\tilde{y}_t$, conditional on $(\lambda, \eta_t)$. Equation (173) becomes:

$$-2\ln \mathbb{P}\left(D_t | \omega, \lambda, \eta_t\right) = \tilde{Y}_t'\left(\omega\right) V^{\tilde{Y}}\left(W\right)^{-1} \tilde{Y}_t\left(\omega\right) + \check{\tilde{y}}_t\left(\omega\right)'\left(V^{\check{\tilde{y}}}\left(\omega\right)\right)^{-1} \check{\tilde{y}}_t\left(\omega\right) + \ln\left|V^{\tilde{Y}}\left(W\right)\right| + \ln\left|V^{\check{\tilde{y}}}\right| \tag{184}$$

with

$$\tilde{Y}_t\left(\omega\right) := \left(\tilde{y}_{St} - M\tilde{y}_{\Gamma(W),t}, F_t^f - \alpha^f M\tilde{y}_{\Gamma(W),t}\right), \tag{185}$$

$$\check{\tilde{y}}_{it}\left(\omega\right) = \tilde{y}_{it} - \tilde{y}_{E(W),t}. \tag{186}$$

So, given $\lambda, \eta_t$, the procedure is as in Section E.1, applied to the tilde variables. In turn, suppose that we have some priors on $\lambda$ (we'll take them to be diffuse) and on $\eta_t$ (we'll normalize them to be independent standard normals). Then, the full likelihood is:

$$\mathbb{P}\left(D_t | \omega\right) = \mathbb{P}\left(D_t | \omega, \lambda, \eta_t\right) \mathbb{P}\left(\lambda, \eta_t\right), \qquad \mathbb{P}\left(\lambda, \eta_t\right) = (2\pi)^{-K/2} e^{-\frac{1}{2}\eta_t \eta_t'}. \tag{187}$$

So, we can estimate $\lambda$ and $\eta_t$ by Bayesian methods, e.g the E-M method summarized in Section E.5.2.

**What the MLE gives** It is worth pausing to see what the MLE does. Consider the MLE estimator of $M$ (keeping the $\alpha^f M$ constant). It is as the analysis of (175) but in $\tilde{y}_t$ space, i.e. it is

equivalent to running the OLS regression:

$$\tilde{y}_{St} = M\tilde{y}_{\Gamma t} + e_t, \tag{188}$$

i.e.,

$$y_{St} - \lambda_S \eta_t = Mz_t + e_t, \tag{189}$$

where

$$z_t := z_{\Gamma t} + z_{\eta t}, \qquad z_{\Gamma t} := u_{\Gamma t}, \qquad z_{\eta t} := \lambda_{\Gamma t} \eta_t. \tag{190}$$

This means that the MLE uses two primitive sources of shocks for identification (i) $z_{\Gamma t} := u_{\Gamma t}$, which is the "pure" GIV, and (ii) $z_{\eta t} := \lambda_{\Gamma t} \eta_t$, which traces the ripple effects of the aggregate shocks $\eta_t$ on the aggregate action $y_{St}$, after controlling for the "direct" effects (this is why $y_{St} - \lambda_S \eta_t$ is on the left-hand side of (189)). Those are two economically very different styles of identification. For economic clarity we find it useful to single out solely the "pure" GIV identification (i.e. regress only on $z_{\Gamma t}$ rather than on $z_{\Gamma t} + z_{\eta t}$).

### E.5.2 Maximum likelihood estimation with heterogeneous loadings

We consider the model with heterogeneous loadings

$$y_t = \gamma y_{St}\iota + \lambda\eta_t + u_t,$$

where $u_t \sim N(0, V_u)$ and $\eta_t \sim N(0, 1)$. Define $\delta_t(\gamma) = y_t - \gamma y_{St}\iota = \lambda\eta_t + u_t$ and note that the log likelihood contribution of $y_t$ is

$$\mathbb{P}(y_t) = \mathbb{P}(\delta_t) + \ln(1 - \gamma).$$

The likelihood of $\delta_t$ can be computed efficiently using the expectation-maximization (EM) algorithm. The steps are as follows, and we refer to Ghahramani and Hinton (1997) for details, where subscripts $(n)$ refer to the $n$-th iteration of the algorithm.

- Expectation step

  - $\mathbb{E}_{(n)}[\eta_t \mid \delta_t] = \beta_{(n)}\delta_t$, where $\beta_{(n)} = \lambda'\Sigma^{-1}$ and $\Sigma_{(n)} = \lambda_{(n)}\lambda'_{(n)} + V_{u(n)}$.
  - $\mathbb{E}_{(n)}[\eta_t^2 \mid \delta_t] = 1 - \beta_{(n)}\lambda_{(n)} + \left(\mathbb{E}_{(n)}[\eta_t \mid \delta_t]\right)^2.$

- Maximization step

  - $\lambda_{(n+1)} = \left(\sum_t \mathbb{E}_{(n)}[\eta_t^2 \mid \delta_t]\right)^{-1} \sum_t \delta_t \mathbb{E}_{(n)}[\eta_t \mid \delta_t].$
  - $V_{u(n+1)} = \frac{1}{T}diag\left\{\sum_t \delta_t\delta'_t - \lambda_{(n+1)}\mathbb{E}_{(n)}[\eta_t \mid \delta_t]\delta'_t\right\}.$

91

The log likelihood can be computed as (omitting constants that do not depend on the parameters)

$$\mathbb{P}_{(n)}(\delta_t) = -\frac{1}{2}\ln\left|\Sigma_{(n)}\right| - \frac{1}{2}\delta_t'\Sigma_{(n)}^{-1}\delta_t,$$

and we iterate until convergence. To initialize the algorithm, we start from estimates of $\lambda$ and $V_u$ based on PCA.

# F    Microfoundations for the model of sovereign spillovers

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 arbitrageurs. Still, the specification that this model delivers might be broadly similar, as we shall see.

## F.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 $df(X_t) = f'(X_t)\,dX_t$.[106],[107]

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

$$Q_{it} = \mathbb{E}_t\left[1 - L_{iT}^+\right] = e^{-(T-t)y_{it}}, \tag{191}$$

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

$$L_{iT} = \frac{F_{iT}}{B_i}, \tag{192}$$

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

$$F_{iT} = \psi_{iT}G_{iT}, \tag{193}$$

where $\psi_{iT} \in [0, 1]$ is a propensity to pass on raw government fiscal losses $G_{iT}$ to bondholders. A

---

[106]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.

[107]We only care, for the regressions, about the "$dZ_t$" terms, that depends on innovation to underlying Brownian shocks $dZ_t$, as those are the loading detected by the regressions.

financially virtuous country (say Germany) has $\psi_{iT}$ close to 0, and a laxer country has a high $\psi_{iT}$. To gain intuition, it is useful to think that most variation in yield spreads comes from the political willingness to not pay bondholders, $\psi_{iT}$.

This raw position $G_{iT}$ is in turn:

$$G_{iT} = V_{iT} - \phi F_{iT}^+ + \phi m_i F_T, \tag{194}$$

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

$$F_T = \sum_i F_{iT}^+. \tag{195}$$

Debts are partially mutualized with intensity $\phi \in [0,1]$: a fraction $\phi$ 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.

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.

## F.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.

$$L_T = \sum_i \frac{B_i}{B} L_{iT}. \tag{196}$$

We call $\rho_i = \frac{m_i}{B_i/B}$, the ratio between country $i$'s mutualization share $m_i$ and its debt share.[108,109]. This leads to:

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

---

[108]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.

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

So, if we define

$$\Psi_{iT} = \frac{\psi_{iT}}{1 + \phi\psi_{iT}}, \tag{197}$$

we have:

$$L_{iT} = \Psi_{iT}\left(\frac{V_{iT}}{B_i} + \phi\rho_i L_T\right). \tag{198}$$

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 (191) gives:

$$y_{it} = a_t\mathbb{E}_t\left[L_{iT}\right], \tag{199}$$

where

$$a_t = \frac{1}{T - t} \tag{200}$$

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 "quasi-static" regime, where all noises are small—see Section F.3 for details. Hence, (198) becomes, in yield space:

$$y_{it} = \Psi_{it}\left(v_{it} + \phi\rho_i y_{St}\right), \tag{201}$$

where

$$y_{St} = \frac{\sum_i B_i y_{it}}{B}. \tag{202}$$

This shows that the yield spread depends on a country-specific fundamental $v_{it}$ and a "spillover" proportional to $\phi$. 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

$$\frac{dy_{it}}{y_{it}} = dw_{it} + \gamma_{it}\frac{dy_{St}}{y_{St}} \tag{203}$$

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 $\Psi_{it}$ (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

$$\tilde{y}_{it} := \ln y_{it}. \tag{204}$$

Then, we have

$$d\tilde{y}_{it} = dw_{it} + \gamma_{it} d\tilde{y}_{\tilde{S}t}, \tag{205}$$

where

$$\tilde{S}_{it} = \frac{B_i y_{it}}{\sum_j B_j y_{jt}}, \tag{206}$$

$$d\tilde{y}_{\tilde{S}t} = \sum_i \tilde{S}_{it} d\tilde{y}_{it} = \sum_i \frac{B_i y_{it} \frac{dy_{it}}{y_{it}}}{\sum_j B_j y_{jt}} = \frac{dy_{St}}{y_{St}}. \tag{207}$$

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 (58). In particular, if $\Psi_{it} = 0$, then the change is $dy_{it} = 0$ always. 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.

## F.3 Quasi-static regime of stochastic processes

Suppose a stochastic process, governed by some noise size $\sigma$, 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 $\sigma$ 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 $t \le T$)

$$X_t = \mathbb{E}_t [F(Y_T)] \tag{208}$$

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

$$X_t = F(\mathbb{E}_t [Y_T]) + O(\sigma^2) \tag{209}$$

i.e. we swap $\mathbb{E}_t$ and $F$.[110] So, that, assuming now that $Y_t$ is a martingale,

$$X_t = F(Y_t) + O(\sigma^2) \tag{210}$$

and

$$dX_t = F'(Y_t) dY_t + O(\sigma^2) \tag{211}$$

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

$$dX_t \simeq F'(Y_t) dY_t. \tag{212}$$

---

[110]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 F.7: Bloomberg identifiers of countries included in the sovereign yield model.

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

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:

$$X_t = e^{Y_t + \frac{\sigma^2}{2}(T-t)}, \qquad dX_t = X_t dY_t \tag{213}$$

and the quasi-static approximation gives

$$X_t = e^{Y_t} + O\left(\sigma^2\right), \qquad dX_t = e^{Y_t} dY_t + O\left(\sigma^2\right). \tag{214}$$

## F.4 Details on the data

Table F.7 describes the tickers of the yields that we use in our empirical analysis.

# 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.

**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.

**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 Miguel and Basile Grassi**, "Large Firm Dynamics and the Business Cycle," *American Economic Review*, 2019.

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

**Ghahramani, Zoubin and Geoffrey E. Hinton**, "The EM Algorithm for Mixtures of Factor Analyzers," *Unpublished working paper, University of Toronto.*, 1997.

**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.

**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.