Granular Instrumental Variables*

Xavier Gabaix and Ralph Koijen

May 9, 2018. Incomplete. Please do not circulate – we really mean it!

Abstract

In many settings, there is a dearth of instruments, which hampers economists' ability to understand causal relations. In this paper we propose a general way to construct instruments, the "granular instrumental variables" (GIVs). In the economies we study, many decisions are taken by a few large firms or countries. As they are large, their idiosyncratic shocks (e.g., productivity shocks) affect the aggregate. This makes those idiosyncratic shocks valid instruments for the aggregate shocks. We provide a methodology to tease apart aggregate vs idiosyncratic effects, this way creating GIVs. Those GIVs allow then to estimate parameters of interest, including causal relations.

We first illustrate the idea in a basic static setup. Then, we show how the procedure can be adapted to handle many enrichments, such as feedback loops, heterogeneity, and autocorrelations (including VARs).

We illustrate the procedure in detail in some applications. (i) We study the impact of mega-firms on aggregates, such as aggregate competition (as proxied by the Herfindahl index). This allows us to find a causal relationship between the rise of mega firms and lower investment, and a lower labor share. (ii) Using country-level supply shocks, we identify supply and demand elasticity for oil. (iii) We also show how in principle GIVs allow to solve outstanding issue: going from local to general-equilibrium impact in cross-state regressions (à la Mian-Sufi); how foreign inflows affect investment; how we can get causality in macro VARs, without the usual identification schemes based on orderings in causality; the interest-rate elasticity of investment. As this paper grows, those analytical results will be supplemented by empirical results. We conclude that the GIVs are a good candidate to supplement the empirical economists' toolbox.

*xgabaix@fas.harvard.edu, rkoijen@stern.nyu.edu. We are grateful to German Gutiérrez and Thomas Philippon for sharing their data with us. This is the very first draft, for submission to the "Rise of mega-firms" conference. Please do not circulate – we are not distributing the paper at this stage. We are confident that the paper will be ready to circulate much before the conference date.

1 Introduction

In many settings, there is a dearth of instruments, which hampers the economists' understanding of causal relations. In this paper we propose a general way to construct instruments, the "granular instrumental variables" (GIVs). Those instruments in turn allow to tease apart causal relations in a wide variety of economic contexts.

In the economies we study, many decisions are taken by a few large firms (we use "firms", but soon we will generalize to large entities such as industries or, in an international context, countries). As has been observed before, those firms are often large, and their idiosyncratic shocks (e.g., productivity shocks) affect the aggregate[1]. The starting observation of this paper is that those firm or country level shocks are valid instruments for the aggregate shocks. We provide a methodology to tease apart aggregate vs idiosyncratic effects, this way creating GIVs. Those GIVs allow then to estimate parameters of interest.

We first illustrate the idea in a basic static setup. Then, we show how the procedure can be enriched to handle many enrichments, such as feedback loops, heterogeneity, autocorrelations (including VAR), and multidimensionality.

Once the basic idea is in place, it is easy to think about many uses for it. Here is a potential list of applications:

1. What's the impact of industry concentration on outcomes? There is a widespread suspicion that this leads to bad things, e.g. low investment, low productivity growth (Gutiérrez and Philippon (2017))? Relatedly, the growth of mega-firms has been suspected to lower the labor share (Autor et al. (2017)). Here the GIV is made of firm-specific growth shocks, that shock the concentration (e.g., Herfindahl) of the economy.

2. What's the impact of output of inflation? Industry-level idiosyncratic TFP growth shocks serve as valid GIV instruments for aggregate TFP, and inflation. More generally, firm or industry-level shocks serve as instrument for macroeconomic VARs, i.e. to see the impact of output on inflation, and then in term of inflation on output.

3. What's the micro-to-macro multiplier? For instance, the work of Mian and Sufi identifies local (country-level) effects of housing booms and busts on local employment (Mian and Sufi (2009)). One important outstanding question is: what is the economy-wide general equilibrium impact of those shocks? Extracting the state or county-specific idiosyncratic shocks again give valid GIVs that allow to estimate this multiplier.

4. What's the impact of foreign flows on the exchange rates? We can look at the idiosyncratic effects of fund-level (or, perhaps less cleanly, country-level) shocks on the aggregate flows as

[1] Hence economies are "granular", i.e. and their shocks made of incompressible "grains" of economic shocks, those firm or industry level shocks. This theme is laid out in Gabaix (2011), and expanded in Acemoglu et al. (2012); Di Giovanni and Levchenko (2012); **?**.

a GIV (using the methodology in Koijen and Yogo (forthcoming)).

5. What's the impact of interest rate on investment? Now, the idiosyncratic investment demand of various firms or sectors lead to shocks to the equilibrium corporate interest rate; this instrument for the change in the interest rate allows to estimate the interest-rate sensitivity of firms' investments.

We think that this methodology will be useful in other contexts, e.g. to study the impact of bank capitalization on lending, the impact on uncertainty on aggregate outcomes (Bloom (2009)), the reflection problem on social science (Manski (1993)), or structural IO as in Berry et al. (1995). We are working on some of those specific examples, and they will be included in this paper as it progresses.

We provide preliminary empirical results for two applications. First, we study the impact of industry concentration on investment and the labor share. Using data compiled by Gutiérrez and Philippon (2017) from 1994 to 2015 covering over 40 industries, we show how to construct GIVs for 5-year changes in industry concentration. We find that changes in industry concentration are significantly negatively related to investment rates and the labor share, even if we control for lagged investment rates or the labor share, lagged concentration and Tobin's Q as well as industry and year fixed effects.

Second, we show how to construct GIVs for demand and supply shocks in oil markets using country-level data on crude oil production and petroleum consumption from January 1985 to December 2017. We find that supply shocks have a significant negative impact on oil price growth, while demand shocks have a positive impact as expected. In future versions of this paper, we will extend this application to a dynamic model of demand and supply in oil markets.

Related literature A productive literature studies the impact of the entry of China in the WTO to study various outcomes, the "China shock" (David et al. (2013); Acemoglu et al. (2016)). In some sense, GIVs are a generalization of China shocks. There are lots idiosyncratic shocks, and GIVs allows to construct them quite systematically.

2 Granular Instrumental Variables: The basics

2.1 Notations

Notations For a vector of $(X_i)_{i=1...N}$, and a series of weight relative weights S_i with with $\sum_i S_i = 1$, we call

$$X_E := \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1}$$

$$X_S := \sum_{i=1}^{N} S_i X_i \tag{2}$$

$$X_\Gamma := X_S - X_E \tag{3}$$

so that X_E is the equal weighted average, X_S is the size weighted average, and X_Γ will be the "granular residual", the difference between size weighted and equal weighted average.

We use notation $\widehat{\mathbb{E}}_T$ for the sample temporal mean; V^e the estimator of V, for some generic variable V; C_t for a vector of controls (it is observable, rather than latent), $\langle X, Y\rangle = \sum_i X_i Y_i$ is the dot product.

Herfindahl We define the herfindahl (strictly speaking, this is an "excess Herfindahl") as:

$$h := \sqrt{-\frac{1}{N} + \sum_{i=1}^{N} S_i^2}. \tag{4}$$

A higher h means that the industry is more concentrated: an industry where all the firms have the same size features $h = 0$, and in general $h \in \left[0, \sqrt{1 - \frac{1}{N}}\right]$.

We will repeatedly use the fact that with $(u_i)_{i=1...N}$ is a series of uncorrelated random variables with mean 0 and common variance:

$$\mathbb{E}\left[u_\Gamma^2\right] = \mathbb{E}\left[u_S u_\Gamma\right] = h^2 \sigma_u^2 \tag{5}$$

Hence, the standard deviation of the granular residual is proportional to the herfindahl.

2.2 A Simple Example

We consider a very simple example. Suppose that demand for some commodity (say, oil) is:

$$\delta_i = \eta + u_i \tag{6}$$

There is an outcome variable of interest, say a price p

$$p = \alpha\delta_S + \varepsilon \tag{7}$$

where η, ε can be correlated. This is, the outcome p depends on the size-weighted average demand, δ_S.

The classic problem is that we cannot estimate α by OLS. Indeed, in a regression, $p = \alpha\delta_S + \varepsilon^e$, ε^e and δ_S are correlated.

However, suppose that we form the GIV:

$$z \coloneqq \delta_\Gamma = \sum_{i=1}^{N} \left(S_i - \frac{1}{N} \right) \delta_i \tag{8}$$

Then, we have

$$z = u_\Gamma. \tag{9}$$

We assume that the shock u_i are idiosyncratic:

$$\mathbb{E}\left[u_i \varepsilon\right] = 0 \text{ for all } i \tag{10}$$

Then, we have

$$\mathbb{E}\left[z\varepsilon\right] = 0: \text{ Exogeneity}$$

and

$$\mathbb{E}\left[z\delta_S\right] \neq 0: \text{ Relevance}$$

Indeed, let us now assume, for the purposes of concrete calculations, that $\mathbb{E}\left[u_i \eta\right] = 0$ for all i (in some sense, this is just a normalization). Then $\mathbb{E}\left[z\delta_S\right] = \mathbb{E}\left[u_S u_\Gamma\right] = h^2\sigma_u^2$, and indeed $\frac{\mathbb{E}[z\delta_S]}{var(z)^{1/2}} = h\sigma_u$. We see that we need a large enough excess herfindahl h to have a strong instrument. Hence, we have $p - \alpha\delta_S = \varepsilon$, so that:

$$\mathbb{E}\left[\left(p - \alpha\delta_S\right) z\right] = 0$$

which gives the supply elasticity α, by

$$\alpha = \frac{\mathbb{E}\left[sz\right]}{\mathbb{E}\left[pz\right]} \tag{11}$$

Indeed, in practice, we form the temporal average

$$\hat{\alpha}_T \coloneqq \frac{\frac{1}{T}\sum_t p_t z_t}{\frac{1}{T}\sum_t \delta_{St} z_t} = \frac{\mathbb{E}_T\left[p_t z_t\right]}{\mathbb{E}_T\left[\delta_{St} z_t\right]} \tag{12}$$

The next proposition states when we have a precise estimator (its proof is in Section 6).

Proposition 2.1 (Precision of the GIV estimate). The GIV estimator has the standard error

$$\sqrt{T}\left(\hat{\alpha}_T - \alpha\right) \sim \mathcal{N}\left(0, \sigma_\alpha^2\right), \qquad \sigma_\alpha = \frac{\sigma_\varepsilon}{h\sigma_u} \tag{13}$$

So to have a precise estimate (low σ_α) we need: a few large firms (to have a large herfindhal), and that demand shocks do affect the price importantly, compared to aggregate shocks (large $\sigma_u/\sigma_\varepsilon$).

This simple example illustrates the basics. In the full version, we can handle a much richer setup, including with feedback loop, richer heterogeneity, autocorrelations (and indeed, vector autocorrelation (VAR)), and multidimensionality.

2.3 Estimators

We next consider a setup with richer heterogeneity. Consider the following model of stationary demands (or in general, actions by firm or entity i):

$$\delta_{it} = k_i + m'C_{it} + \beta p_t + \lambda_{it}\eta_t + u_{it}, \tag{14}$$

where η_t are common shocks, u_{it} are idiosyncratic shocks, and C_{it} is a vector of controls that may include lagged demands and other characteristics. Section 8.3 extends this to heterogeneous β_i, and multidimensional actions.

We suppose also that in equilibrium

$$p_t = \alpha\delta_{St} + \phi C_t + \varepsilon_t, \tag{15}$$

where ε_t is a shock that may be correlated with η_t, C_t is a vector of controls (uncorrelated with ε_t, η_t) which does include a constant, and we assume:

$$\mathbb{E}\left[\left(\varepsilon_t, C_t\right) u_{it}\right] = 0.$$

We assume that $\alpha \neq 0$, but we could have $\beta = 0$. We are interested in estimating α and sometimes to estimate β. We call "no feedback loop" the cases where we have a priori knowledge that $\beta = 0$, and "feedback loop" the cases where β might be non-zero.

In the simplest case, we would have λ_i a constant, say normalized to 1. In a more general model, we model the exposures to the common factor as

$$\lambda_{it} = 1 + \lambda x_{it} + \zeta_i, \tag{16}$$

with $x_{it} \in \mathbb{R}$ is an observable (e.g. x_{it} is the de-meaned log size of entity i) and we normalize $x_{Et} = 0_k$ and $\zeta_E = 0$, implying $\lambda_{Et} = 1$.

We refer to combined error as

$$\xi_{it} = \lambda_{it}\eta_t + u_{it}. \tag{17}$$

We make the following identifying assumptions

$$\begin{aligned} \mathbb{E}[\xi_{it} \mid C_{it}] &= 0, & (18)\\ \mathbb{E}[\eta_t u_{it} \mid x_{it}] &= 0. & (19) \end{aligned}$$

We consider the following procedure:

1. We estimate a panel regression with time and "agent" fixed effects

$$\delta_{it} = m'C_{it} + a_t + k_i + b_t x_{it} + v_{it}, \tag{20}$$

which yields

$$\begin{aligned} a_t &= \beta p_t + \eta_t,\\ b_t &= \lambda\eta_t,\\ v_{it} &= \zeta_i\eta_t + u_{it}. \end{aligned}$$

2. We regress

$$v_{it} = \gamma_i b_t + u_{it}, \tag{21}$$

and collect the residuals. We form the instrument $z_t \coloneqq u_{\Gamma t}$. In the special case where $\zeta_i = 0, \forall i$, we directly have $v_{it} = u_{it}$.

3. To estimate α and ϕ, we use the moment condition:

$$\mathbb{E}\left[\left(p_t - \alpha\delta_{St} - \phi C_t\right)\left(u_{\Gamma t}, C_t\right)\right] = 0 \tag{22}$$

For instance, when there is no control C_t, we use the estimator:

$$\alpha = \frac{\widehat{\mathbb{E}}_T\left[p_t u_{\Gamma t}\right]}{\widehat{\mathbb{E}}_T\left[\delta_{St} u_{\Gamma t}\right]}. \tag{23}$$

4. Suppose next that there is a parameter β to estimate. One way to do that is to use the estimates of step 2 for m, k_E, and form the "purified" δ_{Et}, defined as: $\check{\delta}_{Et} \coloneqq \delta_{Et} - k_E - m'C_{Et}$. Given that $\check{\delta}_{Et} - \beta p_t = \varepsilon_t$, we use the moment condition

$$\mathbb{E}\left[\left(\check{\delta}_{Et} - \beta p_t\right) u_{\Gamma t}\right] = 0 \tag{24}$$

which gives the estimator:[2]

$$\beta = \frac{\widehat{\mathbb{E}}_T \left[\check{\delta}_{Et} u_{\Gamma t}\right]}{\widehat{\mathbb{E}}_T \left[p_t u_{\Gamma t}\right]}. \tag{25}$$

The procedure can be simplified in some cases. *If the λ_i are the same.* Then, we normalize them to $\lambda_i = 1$. In step 1 we just regress:

$$\delta_{it} = m' C_{it} + a_t + k_i + v_{it}, \tag{26}$$

without x_{it}, and in step 2 there is no regression: we just set $u^e_{it} = v_{it}$.

When the weights Γ_{it} are time-invariant. Then, we can simplify Step 2 by just running one regression, rather N regressions (21)

$$v_{\Gamma t} = \gamma_\Gamma b_t + u_{\Gamma t} \tag{27}$$

Robust instruments in the presence of outliers In many applications, smaller countries, firms or funds have more volatile growth rates or flows, and outliers are more common. In such cases, we can compute the equally-weighted average using the largest $N^* < N$. Asymptotic arguments directly extend to this case as long as $\frac{N^*}{N}$ remains constant as $N \to \infty$.

How big are those herfindahls? The GIV instrument is valid as long as $h > 0$, i.e. as long as there is heterogeneity. However, for it to be strong, we need high herfindhals. In estimates for firms, we typically have $h \in [0.02, 0.3]$ (see Gabaix (2011)). One can have an a priori estimate of its size (for theory purposes). In practice, many size distributions follow a power law with fat tails, $\mathbb{P}(S > x) \sim k x^{-\zeta}$ for large x, with $\zeta \in (1, 2]$ — something also explained via random growth behavior (see e.g. Gabaix (1999) and the references therein). In that case one can show that (as in Gabaix (2011), Proposition 2)

$$h \sim k' N^{-\psi}, \qquad \psi = 1 - \frac{1}{\zeta} \in (0, \frac{1}{2}] \tag{28}$$

for k' a non-zero random variable independent of N. The traditional variance case gives corresponds to $\zeta = 2$, which confirms $h \sim k' N^{-1/2}$ (and then k' is a constant) a very weak instrument. But when $\zeta \in (1, 2)$, we have a decay in $N^{-\psi}$ with $\psi \in \left(0, \frac{1}{2}\right)$. A fatter tail in the distribution of large firms (lower ζ) creates scaling in N that decays more slowly (ψ is lower) as N grows large. In the

[2] A variant is to use only the time-series information as one would typically do in macro. In this variant, form new estimators of (m, k_S, k_E) and estimate β using the moment conditions

$$\begin{aligned} k_S, m : \mathbb{E}\left[(\delta_{St} - k_S - m' C_{St} - \beta p_t)(1, C_{St})\right] &= 0, \\ k_E, \beta : \mathbb{E}\left[(\delta_{Et} - k_E - m' C_{Et} - \beta p_t)(1, u_{\Gamma t})\right] &= 0. \end{aligned}$$

limit of Zipf's law (i.e., $\zeta \to 1$), we find $\psi \to 0$ (indeed, one can show that we have $h \sim \frac{k'}{\log N}$), a stronger instrument.

3 Examples of applications

3.1 Impact of market concentration on industry outcomes, e.g. productivity or investment

A common concern is that too much market concentration leads to a variety of bad outcomes, e.g. too little investment or productivity growth. We capture that here. The growth rate of firm i is:

$$g_{it} = \beta p_t + \xi_{it}$$

where

$$c_t = \sum_i \alpha_i S_{it} \tag{29}$$

is the market concentration. The exact specific functional form varies. One example is that concentration can be

$$c_t = \frac{1}{2} \sum_i S_{it}^2 \tag{30}$$

and then

$$\Delta c_t = \sum S_{it,t-1}^2 \frac{\Delta S_{it,}}{S_{i,t-1}} \tag{31}$$

In general, we shall take

$$\Delta c_t = \sum_i \alpha_i g_{it} \tag{32}$$

and in the square share example,

$$\alpha_i = S_{i,t-1}^2.$$

There are microfoundations for that particular measure of concentration. For instance, under Cournot competition, the markup is proportional to the Herfindhal. In term, with endogenous growth, profits, and incentives for higher profits, are proportional to that markup, hence to the Herfindhal. Other models (e.g. with CES aggregators) likely generate more generate measures $c_t = \frac{1}{\psi} \sum_i S_{it}^{\psi}$ with $\psi > 1$, which in turn give $\alpha_i = S_{i,t-1}^{\psi}$.

We can capture (locally) the ad hoc but popular "Ck" concentration ratios (market share of the top k firms), by ordering first by size (1 being the largest), and have

$$\alpha_i = 1_{i \leq k}. \tag{33}$$

Optimal instrument We will use an instrument of the form

$$z_t = \sum_i \Gamma_i g_{it} \tag{34}$$

with $\sum_i \Gamma_i = 0$. Let us pause about the optimal weight Γ_i in that instrument. To maximize precision, we want to maximize

$$\max_{\Gamma} corr\left(z_t\left(\Gamma\right), c_t\right)^2 \text{ s.t. } \sum_i \Gamma_i = 0 \tag{35}$$

That gives (up to a proportionality factor):

$$\Gamma_i = \alpha_i - \frac{1}{N}\sum_{j=1}^{N} \alpha_j \tag{36}$$

Hence, the basic concentration example with $\alpha_i = S_i^2$, the optimal granular instrument is not size-weighted (as in most our examples), but rather squared-size weighted.

Outcome variables The ultimate outcome variable is Y_t, which could be: TFP growth in the industry; mark up in the industry (which is rather hard to measure, see De Loecker and Eeckhout (2017), Gutiérrez and Philippon (2017)); labor share Autor et al. (2017)) growth, or labor share; investment rate (e.g. investment over sales (or over capital stock). This is an aggregate from firm-level decisions:

$$Y_t = y_{St} \tag{37}$$

where the firm-level decisions are:

$$y_{it} = \alpha c_t + \delta C_t + \tilde{\xi}_i \tag{38}$$

$$\tilde{\xi}_i = \tilde{\eta} + \tilde{u}_i$$

Remark that $\mathbb{E}\left[\eta\tilde{\eta}\right]$ and $\mathbb{E}\left[u_i\tilde{u}_i\right]$ could be non-zero.

We take

$$y_{Et} = \alpha\Delta c_t + \delta' C_t' + \tilde{\eta}_t + O\left(\frac{1}{N^{1/2}}\right) \tag{39}$$

where the controls $C_t' = (C_t, c_{t-1})$ include lagged concentration c_{t-1}. We use the moment condition (true up to $O\left(\frac{1}{N^{1/2}}\right)$ terms):

$$\mathbb{E}\left[\left(y_{Et} - \alpha\Delta c_t - \delta' C_t'\right)\left(z_t, C_t'\right)'\right] = 0 \tag{40}$$

That allows us to estimate α, the impact of concentration on the outcome of interest.

3.2 From micro to macro: calculating the GE implications of Mian-Sufi style estimates

Here i is a state, p_{it} the housing price (compared to some benchmark), L_{it} employment growth. We have:

$$p_{it} = \alpha_i h_t + \eta_t^p + u_{it}^p \tag{41}$$
$$L_{it} = \beta p_{it} + \gamma L_{St} + \eta_t^L + u_{it}^L \tag{42}$$

where h_t is some housing shock (coming e.g. from taste, or anticipation of future rents), η's are aggregate shocks, u_{it}'s are idiosyncratic shocks. The GE effect is via L_{St}.

3.2.1 Let us first recall what Mian-Sufi do to estimate partial equilibrium effects

Mian-Sufi posit

$$\alpha_i = \alpha^0 + \alpha^1 X_i \tag{43}$$

where X_i is a local sensitivity, e.g. Saiz (2010)-style supply elasticity, normalized so that $X_E = 0$. Then estimate α^1 via:

$$p_{it} - p_{Et} = \alpha^1 X_i h_t \tag{44}$$

where we approximate $u_{Et}^p = 0$, and h_t is proxied by the increase in housing growth, $p_{Et} = \alpha^0 h_t + \eta_t^p$.

Then, they form:

$$L_{it} - L_{Et} = \beta \left(p_{it} - p_{Et}\right) + u_{it}^L \tag{45}$$
$$= \beta \left(\alpha^1 X_i h_t + u_{it}^p\right) + u_{it}^L \tag{46}$$

and they estimate β by

$$\mathbb{E}\left[\left(L_{it} - L_{Et} - \beta\alpha^1 X_i h_t\right) h_t\right] = 0 \tag{47}$$

which gives β. This parameter β is the partial-equilibrium impact of house prices on employment, estimated from the cross-section.

3.2.2 The GE impact

Aggregate employment growth is:$L_{St} = \beta p_{St} + \gamma L_{St} + \eta_t^L + u_{St}^H$, so

$$L_{St} = \frac{1}{1-\gamma}\left(\beta p_{St} + \eta_t^L + u_{St}^H\right) \tag{48}$$

Hence, the multiplier is $\frac{1}{1-\gamma}$, and the GE effect is $\frac{\beta}{1-\gamma}$. That literature is not good at estimating γ.

We now show how to estimate the general equilibrium parameter γ via GIVs.

First, the Mian-Sufi data is at the ZIP code level. We aggregate it up to something bigger, e.g. cities, counties or states. Then we form the instrument:

$$z_t \coloneqq L_{\Gamma t}$$

where we impose

$$\langle \Gamma, 1 \rangle = \langle \Gamma, \alpha \rangle = 0 \tag{49}$$

where $\langle X, Y \rangle = \sum_i X_i Y_i$ is the dot product. This is, in addition to the usual $\sum_i \Gamma_i = 0$, we impose $\sum_i \Gamma_i \alpha_i = 0$. Let us pause to think about this instrument. Like in (35), we we still to maximize $\frac{\langle \Gamma, S \rangle}{\langle \Gamma, \Gamma \rangle^{1/2}}$ s.t. $\langle \Gamma, 1 \rangle = \langle \Gamma, \alpha \rangle = 0$. A bit of linear algebra gives $\Gamma = \hat{S} + b\alpha$ (using $\alpha_E = 0$), where $\hat{S}_i = S_i - \frac{1}{n}$, and b is chosen so that $\langle \Gamma, \alpha \rangle = 0$. This gives the optimal weight

$$\Gamma = \hat{S} - \frac{\left\langle \alpha, \hat{S} \right\rangle}{\langle \alpha, \alpha \rangle} \alpha \tag{50}$$

This implies (using $\langle \Gamma, \alpha \rangle = 0$):

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

We also have

$$L_{Et} = \beta p_{Et} + \gamma L_{St} + \eta_t^L$$

so

$$\mathbb{E}\left[\left(L_{Et} - \beta p_{Et} - \gamma L_{St} \right) L_{\Gamma t} \right] = 0$$

which allows to estimate γ (recall that p_{Et} is known, and β has already been estimated)

The coefficient β comes from the Mian-Sufi cross-sectional analysis, and the coefficient γ from the GIV. Together, they give the GE impact, $\frac{\beta}{1-\gamma}$.

3.3 Interest-rate elasticity of investment demand: an instrument for the interest rate

Suppose that firm i has investment growth (or Investment over assets)

$$X_i = \beta \left(r_t + s_i \right) + \delta^D C + \delta_2 F_i + x_i \tag{51}$$

where $-\beta$ is the interest-rate semi-elasticity of investment, r_t is the safe interest rate, and s_i is the yield spread faced by firm i (that depends e.g. on its credit rating), C is a vector of controls, and F_i is a vector of firm-level controls.

Note: there could be t's everywhere. They're not here to lighten up the notation. Likewise, we drop the F_i term in the exposition.

Here, to streamline the analysis (and zoom in on issue of identification), we do the analysis imagining that β and δ^D do not depend on i. It is reasonably easy to extend to the case where they do depend on i. We posit a structure:

$$x_i = \eta + u_i \tag{52}$$

The demand for loans is:

$$L^{Dem} = \sum_i S_i X_{it} =: X_t^S \tag{53}$$

The supply of loans is

$$L^{Supply} = \alpha r_t + \delta^S C_t + \varepsilon_t \tag{54}$$

with presumably $\alpha > 0$, $\beta < 0$. So equilibrium entails: $L^{Supply} = L^{Dem}$, so

$$r_t = \frac{1}{\alpha - \beta}\left(X_t^S + \left(\delta^D - \delta^S\right) C_t + \varepsilon_t\right) \tag{55}$$

We set the instrument

$$z \coloneqq X^\Gamma = u^\Gamma, \tag{56}$$

for the interest (perhaps a somewhat weak one). We have $cov\left(r_t, z_t\right) \neq 0$.

We have

$$\mathbb{E}\left[\left(X_t^E - \beta\left(r_t + s_t^E\right) - \delta^D C_t\right)\left(z_t, C_t\right)'\right] = 0 \tag{57}$$

which gives an estimate of β, δ^D.

Next, define, for given values of α, δ^D:

$$y_t \coloneqq \left(\alpha - \beta\right) r_t + \left(\delta^S - \delta^D\right) C - X^S \tag{58}$$

Then, under the true values of α, δ^D, $y_t = \varepsilon_t$. So we have

$$\mathbb{E}\left[y_t\left(z_t, C_t\right)'\right] = 0 \tag{59}$$

which identifies α, δ^S.

Here is a perhaps neater way to estimate the supply elasticity – that doesn't use that y_t device. We have $L^{Supply} - \alpha r_t - \delta C_t = \varepsilon_t$, so we form:

$$\mathbb{E}\left[\left(L^{Supply} - \alpha r_t - \delta C_t\right)\left(z_t, C_t\right)'\right] = 0 \tag{60}$$

which allows to estimate α, δ^S.

Remarks Enrichments to this little model: model the market for all yield spread.

How weak is this instrument? it's: how much does investment demand drive the real interest rate. Of course, the Fed tends to smoothly adjust interest rates (as in $i_t = \alpha \sum_{k \geq 0} \gamma^k x_{t-k}^S + \varepsilon_t$,

where i_t is the nominal interest rate), so that the situation is more complex

So it's more: via yield spreads, we measure β (or β_i); which gives also an estimate

We could also estimate α_I, β_I for different classes of firms, e.g. investment-grade or not.

3.4 General equilibrium macro

3.4.1 Identification in aggregate macro: static case first

To illustrate the economics, let's first do a static (really, 2 period with frictionless second period) New Keynesian model. With y output (as a deviation from steady state) and i nominal interest rate, we have

$$y = -\sigma i + \varepsilon^y + u_S \tag{61}$$

$$i = \phi y + \varepsilon^i \tag{62}$$

where the ε are shocks, and u_j^y are the firms-level shocks (i.e., at the firm level, we have $y_j = -\sigma i + \varepsilon^y + u_j$, and output it $y = y_S$), so that u_S is the size-weighted idiosyncratic firm-level shocks

First, with just one instrument $z = u_\Gamma$, we get identification. The moments are $\mathbb{E}\left[(y + \sigma i - u_\Gamma)\, z\right] = 0$, which gives σ, and same thing for i: $\mathbb{E}\left[(i - \phi y)\, z\right] = 0$ gives ϕ.

Generalizing the thought, suppose that we have a system $Y \in \mathbb{R}^n$ (e.g. $Y = (y, i)$) with n variables

$$Y = BY + \varepsilon + U_S \tag{63}$$

with

$$B_{jj} = 0 \text{ for all } j \tag{64}$$

and ε is a mean-zero shock and U_S is a vector of size-weighted idiosyncratic shocks. In the example above, $v' = (\varepsilon^y, \varepsilon^i)$ and $U_S' = (u_S, 0)$ in the example above). Then we can identify the system if we have

$$k \geq n - 1 \tag{65}$$

instruments. Indeed, the moment conditions are:

$$\mathbb{E}\left[\left(Y_i - \sum_j B_{ij} Y_j - U_{S,i}^e\right) z'\right] = 0 \tag{66}$$

with $z \in \mathbb{R}^k$ the instrument. We set

$$U_S^e \coloneqq Y_{St} - Y_{Et} \tag{67}$$

Note that if there are firm-level controls, we need to remove them. For instance, if firm f has:

$$y_{ft} = \delta C_{ft} - \sigma i_t + u_{ft}$$

then we need to control for C_{ft} when doing the estimation.

A 4-dimensional example Suppose that we have investment I_{ft} and employment L_{ft} in our enriched example. Then $Y_t = (y_{St}, I_{St}, L_{St}, i_t)$. The framework above works, with $U_S = \left(u^y_{St}, u^I_{St}, u^L_{St}, 0\right)$ and $z = U_\Gamma = \left(u^y_{\Gamma t}, u^I_{\Gamma t}, u^L_{\Gamma t}, 0\right)$.

Connection to the SVAR literature Now, let's connect this to the SVAR literature.

Static case We first focus on the contemporaneous term. Suppose that we have Y_t. One can always write this as:

$$Y_t = Aw_t \tag{68}$$

where w_t is a vector of shocks with unit covariance, $\mathbb{E}ww' = I$. But the only data we have is Y_t. We can identify A in the following way.

We first estimate (63):

$$Y = BY + v \tag{69}$$

where v is the total shock:

$$v = \varepsilon + U_S \tag{70}$$

with $B_{jj} = 0$ for all j, where U_S is a vector made of firm-level idiosyncratic shocks ($U_S = (u_S, 0)$ in the previous example).

Hence, we can answer:

(Q1 – Causality) "What does an interest shock imply for output?", "What does a shock to output imply for the interest rate?" The answer is in matrix B. This is the main macro question.

Often economists would like to answer a question of the type: "what's the share of fluctuations of variable X coming from shocks of type i?" Let us inspect the situation, and see how we can answer a form of that question.

Consider an outcome X of the type:

$$X = b'Y = b'\left(I - B\right)^{-1} v = a'v \tag{71}$$

with $a' \coloneqq b'\left(I - B\right)^{-1}$, then we have the decomposition:

$$1 = \sum_i \frac{cov\left(X, a_i v_i\right)}{var\left(X\right)} \tag{72}$$

This gives a decomposition "a share $\frac{cov(X, a_i v_i)}{var(X)}$ of the variance of X comes from shock v_i".[3]

Hence, we have a clear answer to the question:

[3]There is another decomposition:

(Q2 –Share of variances) "we can decompose the variance of outcome X into a $\frac{cov(X,a_i v_i)}{var(X)}$ coming from shocks to i."

Dynamic case How, go back to the more general SVAR. We first estimate

$$Y_t = G(L) Y_{t-1} + w_t \tag{74}$$

by OLS, which gives a series of innovation w_t. Then, we say that truly $Y_t - \mathbb{E}_{t-1} Y_t = w_t$ reflects a deeper structure (63). So, we use the procedure above (the "static case") to estimate B, and possibly the variance decomposition.

4 Preliminary Empirical Results

4.1 Industry concentration, investment, and the labor share

We first apply our GIV estimator to measure the causal impact of industry concentration on investment and the labor share. We use the data constructed by Gutiérrez and Philippon (2017) (GT). We use the Herfindahl index based on sales as our measure of industry competition. Formally, we index firms by i, industries by j, and time by t. The Herfindahl index is defined as

$$C_{jt} = \sum_i S_{ijt}^2,$$

where $S_{ijt} = y_{ijt}/y_{jt}$ with y_{ijt} denoting firm-level sales. All variables that are only indexed by j are the sum within each industry.

GT start from Compustat data and make a series of adjustments as Compustat only contains a subset of all firms and does not account for imports. For manufacturing, they use data from Feenstra and Weinstein (2017) based on the U.S. Economic Census. For other industries, GT adjust the Herfindahl using the share of Compustat sales relative to U.S. output and, when available, imports. We follow their definition of the net investment rate, ni_{jt} (essentially, net investment divided by previous period assets), and we construct a measure of the the labor share as the compensation of employees relative to the sum of compensation of employees and the gross operating surplus, ls_{jt}. Both the net investment rate and the labor share are constructed using data from the U.S. Bureau of Economic Analysis (BEA).

$$1 = \sum_i \frac{a_i^2 var(v_i)}{var(X)} + 2 \sum_{i<j} \frac{a_i a_j cov(v_i, v_j)}{var(X)} \tag{73}$$

Then we can say "a share $\frac{a_i^2 var(v_i)}{var(X)}$ of the variance from from shocks i" – but then there's a covariance term. The two decompositions are the same if the component of v are uncorrelated.

For net investment, we consider the following specification

$$ni_{jt} = \alpha_0 + \alpha_1 \Delta_5 C_{jt} + \alpha_2 C_{j,t-5} + \alpha_3 ni_{j,t-5} + \alpha_3 Q_{j,t-1} + \epsilon_{jt}, \tag{75}$$

where $Q_{j,t-1}$ is Tobin's Q, which is constructed from Compustat. Hence, in our regression, we explain investment with investment five years ago, $ni_{j,t-5}$, Tobin's Q in the previous year, $Q_{j,t-1}$, concentration five years ago, $C_{j,t-5}$, and the change in concentration during the last five years. We are interested in estimating α_1. We estimate an analogous model for the labor share, in which case we omit Tobin's Q.

The main concern is that other, omitted, factors affect both industry concentration and investment rates or the labor share. To mitigate such concerns, we also estimate the model with year fixed effects and with industry fixed effects.

We first report OLS estimates as a point of reference in Table 1. We find that a change in industry concentration is negatively related to net investment, even if we account for industry and time fixed effects. For the labor share, the OLS results are more fragile. Without any fixed effects, or with only year fixed effects, changes in industry concentration are negatively related to the labor share. This result disappears when we include industry fixed effects.

Next, we use GIV to instrument for changes in competition. We first rewrite the definition of the change in the Herfindahl index $\Delta_5 C_{jt}$ as

$$\Delta_5 C_{jt} = \sum_{i \in I^*_{jt}} \alpha_{ijt} g_{ijt}, \tag{76}$$

where I^*_{jt} is the set of firms that exist in period $t-5$ or period t and

$$\begin{aligned} \alpha_{ijt} &= \frac{\hat{y}^2_{ijt}}{\hat{y}^2_{jt}}, \\ g_{ijt} &= \frac{1}{\alpha_{ijt}} \left(S^2_{ijt} - S^2_{ij,t-5} \right), \\ \hat{y}_{ijt} &= (y_{ijt} + y_{ij,t-5})/2. \end{aligned}$$

Note that for small shocks, g_{ijt} is approximately equal to the log growth rate of S^2_{ijt}. Empirically, we prefer to use the average of sales between $t-5$ and t as it is less sensitive to outliers.

We construct the instrument for $\Delta_5 C_{jt}$ as

$$z_{jt} = \sum_{i \in I^*_{jt}} \Gamma_{ijt} g_{ijt},$$

where

$$\Gamma_{ijt} = \alpha_{ijt} - \frac{1}{N_{jt}} \sum_{i \in I^*_{ijt}} \alpha_{ijt}.$$

	(1)	(2)	(3)	(4)	(5)	(6)
	Net investment rate			Labor share		
dHerf	-0.29	-0.27	-0.18	-0.26	-0.29	0.06
	(-7.20)	(-6.81)	(-4.51)	(-2.27)	(-2.58)	(0.57)
Herf(-5)	0.05	0.05	0.01	-0.10	-0.10	0.34
	(2.94)	(3.33)	(0.24)	(-2.10)	(-2.24)	(4.49)
NI(-5)	0.27	0.32	-0.07			
	(9.37)	(10.74)	(-1.90)			
Q(-1)	0.02	0.01	0.02			
	(8.97)	(7.52)	(6.63)			
LS(-5)				0.90	0.90	0.12
Year FE		X	X		X	X
Industry FE			X			X
R-sqr	0.32	0.41	0.59	0.75	0.77	0.88

Table 1: Changes in industry concentration, net investment, and the labor share: OLS estimates. We report the results of estimating (75) using OLS in Column (1) to (3). In Column (4) to (6), we estimate an analogous model for the labor share, in which case we omit Tobin's Q. dHerf is the 5-year change in the Herfindahl index of industry j, Herf(-5) is the 5-year lag in the Herfindahl index of industry j, NI(-5) and LS(-5) are the 5-year lags of net investment or the labor share of industry j, respectively, and Q(-1) corresponds to Tobin's Q in industry j in the previous year. In Column (2) and (5), we add year fixed effects, and in Column (3) and (6) we also add industry fixed effects. The sample is from 1994 to 2015 and covers 42 industries.

	(1)	(2)	(3)	(4)	(5)	(6)
	Net investment rate			Labor share		
dHerf	-0.46	-0.36	-0.28	-1.98	-1.92	-1.33
	(-3.55)	(-2.89)	(-2.78)	(-3.13)	(-3.27)	(-3.38)
Herf(-5)	0.01	0.03	-0.04	-0.54	-0.52	-0.27
	(0.17)	(0.91)	(-0.57)	(-2.43)	(-2.50)	(-2.07)
NI(-5)	0.26	0.31	-0.07			
	(4.96)	(5.52)	(-1.34)			
Q(-1)	0.02	0.01	0.02			
	(7.24)	(6.33)	(5.25)			
LS(-5)				0.87	0.87	0.11
				(13.61)	(14.44)	(1.18)
Year FE		X	X		X	X
Industry FE			X			X
R-sqr	0.31	0.41	0.59	0.69	0.72	0.85

Table 2: Changes in industry concentration, net investment, and the labor share: IV estimates. We report the results of estimating (75) using IV in Column (1) to (3). In Column (4) to (6), we estimate an analogous model for the labor share, in which case we omit Tobin's Q. dHerf is the 5-year change in the Herfindahl index of industry j, Herf(-5) is the 5-year lag in the Herfindahl index of industry j, NI(-5) and LS(-5) are the 5-year lags of net investment or the labor share of industry j, respectively, and Q(-1) corresponds to Tobin's Q in industry j in the previous year. In Column (2) and (5), we add year fixed effects, and in Column (3) and (6) we also add industry fixed effects. The sample is from 1994 to 2015 and covers 42 industries.

We scale the Herfindahl index for the non-manufacturing using the same factor as in GT. We average the factor across the entire sample from 1994 to 2016 so that variation in the instrument is not due to variation in the scaling. Lastly, we truncate the changes so the instrument Herfindahl remains between zero and one.

We report the results in Table 2. The results for net investment rates are similar to the OLS estimates. However, for the labor share, we now find that the change in concentration significantly predicts a decline of the labor share in all three specifications, even when we account for industry fixed effects. The point estimates are also more negative than their OLS counterparts.

Using our GIVs, we conclude that increases in industry concentration significantly lower net investment rates and the labor share.

4.2 Demand and supply shocks in oil markets

We use monthly data from the U.S. Energy Information Administration (EIA) on country-level crude oil production (including OPEC and non-OPEC countries) and petroleum consumption. The production data are from 21 countries and the consumption data from eight countries. Data on

oil prices is also from the U.S. EIA. Following Kilian (2009), we use the refiner acquisition cost of imported crude oil as the nominal oil price and deflate it using the U.S. CPI.

Denote the production in country i at t as Q_{it}^S and the demand by Q_{it}^D. We model, with $j = S, D$,

$$\Delta Q_{it}^j = c + \alpha^j \Delta P_t + \eta_t^j + \epsilon_{it}^j,$$

where $\Delta x_t = (x_t - x_{t-1})/x_{t-1}$. Future versions of this paper will consider a dynamic demand and supply model, consistent with the structural VAR literature (e.g., Kilian (2009) and Caldara et al. (2016)), and use more complete data so we can impose market clearing. We use this simple model to construct granular instrumental variables for demand and supply shocks. We denote the share of country i by S_{it} and define our instrument as before

$$z_t^j = \sum_i \Gamma_{it}^j \Delta Q_{it}^j,$$

where

$$\Gamma_{it}^j = S_{i,t-1}^j - 1/N_i^j.$$

Our sample starts in January 1985, as in Caldara et al. (2016), and ends in December 2017. During turbulent periods such as wars, the supply from some countries falls to (close to) zero and often recovers quickly. This leads to extreme positive growth rates. Given the relatively small cross section, we compute the supply GIV using the eight countries (the same number as we have for the demand side) using the largest eight producers of the previous month. We cannot make the same adjustments for the demand GIV, but we plan to extend our data in future version to follow the same procedure.

In Figure 1, we plot the supply growth alongside the supply GIV. The supply GIV captures some of the large swings. For instance, the global supply of oil reduced in 1990 as Iraq and Kuwait cut their production, while Saudi-Arabia increased their production in subsequent months.

In Table 3, we report the results of a regression of ΔP_t on GIV supply and demand shocks. We find that supply shocks have a significantly negative impact on oil price growth and demand shocks have a positive impact, as expected. Future versions will extend these results using a dynamic demand and supply model.

5 Conclusion

In this paper we introduce the GIVs. We remark that idiosyncratic shock offer a rich source of instruments, and we develop econometric procedures to extract them from aggregate shocks. As this paper progresses, we shall report more concrete applications of GIVs. We think that GIVs are likely to be helpful in a variety of settings beyond those studied in this paper.

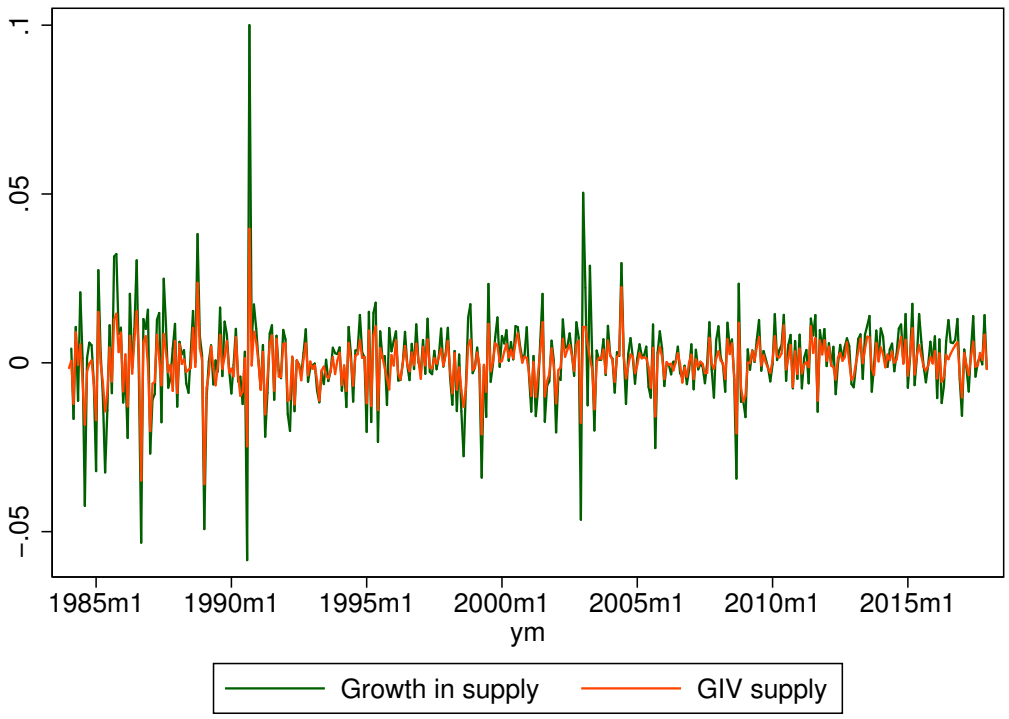


Figure 1: Growth in supply and the supply GIV. The sample is from January 1985 until December 2017.

	(1) ΔP	(2) ΔP	(3) ΔP
GIV supply	-1.19		-1.16
	(-2.06)		(-2.02)
GIV demand		0.35	0.33
		(1.54)	(1.47)
Constant	0.00	0.00	0.00
	(0.74)	(0.87)	(0.88)

Table 3: The impact of demand and supply shocks on oil price growth. The first column reports the results of a regression of oil price growth, ΔP_t, on GIV supply shocks. The second column reports the results of a regression of oil price growth on GIV demand shocks. The third column reports the results of a regression on both shocks. We report the standard errors in parentheses. The sample is from January 1985 to December 2017.

References

Acemoglu, Daron, Ufuk Akcigit, and William Kerr, "Networks and the macroeconomy: An empirical exploration," *NBER Macroeconomics Annual*, 2016, *30* (1), 273–335.

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

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, John Van Reenen et al., "The fall of the labor share and the rise of superstar firms," 2017.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica*, 1995, pp. 841–890.

Bloom, Nicholas, "The impact of uncertainty shocks," *Econometrica*, 2009, *77* (3), 623–685.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello, "Oil Price Elasticities and Oil Price Fluctuations," Technical Report, Board of Governors of the Federal Reserve System (US) 2016.

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.

Feenstra, Robert C. and David E. Weinstein, "Globalization, markups and U.S. welfare," *Journal of Political Economy*, 2017, *125* (4), 1040–1074.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, 1999, *114* (3), 739–767.

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

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

Gutiérrez, Germán and Thomas Philippon, "Declining Competition and Investment in the US," Technical Report, National Bureau of Economic Research 2017.

Hamilton, James D., "Causes and Consequences of the Oil Shock of 2007-08," *Brookings Papers on Economic Activity*, 2009, *40* (1 (Spring), 215–283.

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.

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

Loecker, Jan De and Jan Eeckhout, "The rise of market power and the macroeconomic implications," Technical Report, National Bureau of Economic Research 2017.

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

Mian, Atif and Amir Sufi, "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis," *The Quarterly Journal of Economics*, 2009, *124* (4), 1449–1496.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

6 Appendix: Proofs

Proof of Proposition 2.1 We have

$$\hat{\alpha_T} - \alpha = \frac{\widehat{\mathbb{E}}_T \left[\left(\alpha \delta_{St} + \varepsilon_t \right) u_{\Gamma t} \right]}{\widehat{\mathbb{E}}_T \left[\delta_{St} u_{\Gamma t} \right]} - \alpha = \frac{\widehat{\mathbb{E}}_T \left[\varepsilon_t u_{\Gamma t} \right]}{\widehat{\mathbb{E}}_T \left[\delta_{St} u_{\Gamma t} \right]} = \frac{A_T}{D_T}$$

Next,

$$D_T = \widehat{\mathbb{E}}_T \left[\delta_{St} u_{\Gamma t} \right] \to D = \widehat{\mathbb{E}} \left[\delta_S u_\Gamma \right]$$

$$D = \mathbb{E} \left[\left(\eta + u_S \right) u_\Gamma \right] = \mathbb{E} \left[u_S u_\Gamma \right] = h^2 \sigma_u^2$$

For the numerator, we have:

$$T^{1/2} A_T \sim \mathcal{N} \left(0, \sigma_A^2 \right)$$

where

$$\sigma_A^2 = \mathbb{E} \left[\varepsilon^2 u_\Gamma^2 \right] = \mathbb{E} \left[\varepsilon^2 \right] \mathbb{E} \left[u_\Gamma^2 \right] = \sigma_\varepsilon^2 h^2 \sigma_u^2$$

so that

$$\frac{\sigma_A}{D} = \frac{\sigma_\varepsilon h \sigma_u}{h^2 \sigma_u^2} = \frac{\sigma_\varepsilon}{h \sigma_u} =: \sigma_\beta$$

Hence,

$$\sqrt{T} \left(\hat{\alpha_T} - \alpha \right) \sim N \left(0, \sigma_\alpha^2 \right)$$

□

7 Appendix: Asymptotic Properties of GIV Estimators

Here we analyze the econometric performance of GIV estimators.

7.1 Precision of GIV estimators in simple contexts

7.1.1 Situations without feedback loop

Homogenous exposure case ($\lambda_i = 1$ for all i) We form, as an instrument for the price

$$z := \delta_\Gamma \tag{77}$$

so that under the assumptions

$$z = u_\Gamma$$

and we have the moment condition to identify α:

$$\mathbb{E} \left[\left(p - \alpha \delta_S \right) z \right] = 0 \tag{78}$$

Hence the estimator is:

$$\hat{\alpha}_T = \frac{\frac{1}{T}\sum_t p_t z_t}{\frac{1}{T}\sum_t \delta_{St} z_t} \tag{79}$$

Heterogeneous exposure case (λ_i need not be all equal to 1)

Parametric heterogeneous exposures We posit:

$$\lambda_i = 1 + \gamma' X_i \tag{80}$$

for some vector γ', with $X_E = 0$. Then, we estimate γ with a precision $O\left(\frac{1}{N^{1/2}}\right)$, and we can show that (13) still holds.

Nonparametric heterogeneous exposures We analyze the precision in the method of Section 2.3 when we have heterogeneous, non-parametric λ_i. We note V^e the estimator of V, for some generic variable V. Given that

$$\delta_{\Gamma t} = \lambda_\Gamma \eta_t + u_{\Gamma t} = \lambda^e_{\Gamma t}\eta^e_t + u^e_{\Gamma t}$$

we get for the error in the estimation of $u_{\Gamma t}$:

$$u^e_{\Gamma t} - u_{\Gamma t} = \lambda_{\Gamma t}\eta_t - \lambda^e_{\Gamma t}\eta^e_t = \left(\lambda_\Gamma - \lambda^e_\Gamma\right)\eta_t + \left(\eta_t - \eta^e_t\right)\lambda^e_\Gamma$$

and we evaluate the magnitude of each term

We can show that

$$\lambda^e_\Gamma - \lambda_\Gamma = \frac{\sigma_u h}{\sigma_\eta}\frac{n_\lambda}{\sqrt{T}} \tag{81}$$

with for some $N(0,1)$ variable n_λ indicating the measurement error.

As η_t is estimated from the cross-section,

$$\eta_t - \eta^e_t = \frac{\sigma_u}{\lambda\sigma_x}\frac{n_{\eta_t}}{\sqrt{N}}$$

with again n_{η_t} a $N(0,1)$ variable. In this analysis we assume

$$N \gg T \tag{82}$$

so we have, to the leading order,

$$u^e_{\Gamma t} - u_{\Gamma t} = \left(\lambda_\Gamma - \lambda^e_\Gamma\right)\eta_t \tag{83}$$

So, we get some systematic error in $u^e_{\Gamma t} - u_{\Gamma t}$, correlated with η_t.

Next, consider the "no loop case", with

$$\mathbb{E}\left[\left(p-\alpha\delta_S\right)z\right]=0 \tag{84}$$

and $z=u_\Gamma^e$. Then,

$$\hat{\alpha}_T-\alpha=\frac{\mathbb{E}_T\varepsilon_t z_t}{\mathbb{E}_T\delta_{St}z_t}$$

The numerator is

$$A_T=\mathbb{E}_T\varepsilon_t u_{\Gamma t}^e=\mathbb{E}_T\varepsilon_t\left(u_{\Gamma t}+\left(\lambda_\Gamma-\lambda_\Gamma^e\right)\eta_t\right)=\mathbb{E}_T\varepsilon_t u_{\Gamma t}+\left(\lambda_\Gamma-\lambda_\Gamma^e\right)cov\left(\eta_t,\varepsilon_t\right) \tag{85}$$

The first term on the right-hand side, $\mathbb{E}_T\varepsilon_t u_{\Gamma t}$, is the one that comes about even without measurement error for λ_Γ: it has standard deviation $\frac{\sigma_\varepsilon\sigma_u h}{\sqrt{T}}$, as we saw above. The second one, $\left(\lambda_\Gamma-\lambda_\Gamma^e\right)cov\left(\eta_t,\varepsilon_t\right)$, is specific to the measurement error of λ_Γ. Equation (81) shows that has S.E. $\frac{\sigma_\varepsilon\sigma_u h}{\sqrt{T}}corr\left(\eta,\varepsilon\right)$, and it is uncorrelated with the first one. (The same holds by inspection of (89), replacing ε_t by η_t, so that the correlation is 1). Hence, we obtain the following.

Proposition 7.1 (Arbitrary heterogeneities in λ multiply standard error of estimates of α and β by $\sqrt{2}$ at most). *Consider the case where the λ_i are heterogeneous, and estimated by the method in section 2.3 (with cross-section then time-series). In the No Feeback loop (resp. Feedback loop) case, the standard error on α (resp. β) is the one in the $\lambda_i=1$ seen in (13) (resp. seen in (88)) times $\sqrt{1+corr\left(\varepsilon,\eta\right)^2}$ (resp. times $\sqrt{2}$). In other terms, one can estimate the standard error as if we there was no measurement error in u_Γ^e – simply, one must then multiply the standard error on α (resp. β) by the above factor ($\sqrt{1+corr\left(\varepsilon,\eta\right)^2}$ (resp. $\sqrt{2}$), which is bounded above by $\sqrt{2}$.*

In other terms, arbitrary heterogeneities in λ_i (i.e., when we use the procedure of section 2.3 with cross-section then time-series) multiply standard error of estimates of α and β by at most $\sqrt{2}$, compared to the naive S.E. that would ignore that the instrument $z=u_\Gamma^e$ has measurement error due to the estimation of λ_i.

7.1.2 Situations with feedback loop

First, to estimate α, the same procedure and instrument work as in the No feedback loop case. Equation (13) holds. The difficulty is just to estimate β.

We use the same instrument for the price

$$z\coloneqq\delta_\Gamma=u_\Gamma \tag{86}$$

and as (up to an $O\left(\frac{1}{\sqrt{N}}\right)$ an term)

$$\delta_E=\beta p+\eta$$

we have

$$\mathbb{E}\left[\left(\delta_E - \beta p\right) z\right] = 0$$

which gives β:

$$\hat{\beta}_T = \frac{\frac{1}{T}\sum_t \delta_{Et} z_t}{\frac{1}{T}\sum_t p_t z_t} \tag{87}$$

Now, what's the error in $\hat{\beta}_T$? One can show that

$$\sqrt{T}\left(\hat{\beta}_T - \beta\right) \sim \mathcal{N}\left(0, \sigma_\beta^2\right), \qquad \sigma_\beta = \frac{\sigma_\eta}{|\alpha|\,\sigma_u h} \tag{88}$$

So to have a precise estimate we need: a few large firms (to have a large herfindhal), and that demand shocks do affect the price (a large α). However, perhaps surprisingly, the variance of f does not figure in (88).

Proof of (88) We have,

$$\hat{\beta}_T - \beta = \frac{\frac{1}{T}\sum_t \delta_{Et} z_t}{\frac{1}{T}\sum_t p_t z_t} - \beta = \frac{\widehat{\mathbb{E}}_T\left[\left(\beta p_t + \eta_t + u_{Et}\right) u_{\Gamma t}\right]}{\widehat{\mathbb{E}}_T\left[p_t u_{\Gamma t}\right]} = \frac{\widehat{\mathbb{E}}_T\left[\left(\eta_t + u_{Et}\right) u_{\Gamma t}\right]}{\widehat{\mathbb{E}}_T\left[p_t u_{\Gamma t}\right]} = \frac{A_T}{D_T} \tag{89}$$

Next,

$$D_T = \widehat{\mathbb{E}}_T\left[p_t u_{\Gamma t}\right] \to D = \widehat{\mathbb{E}}\left[p u_\Gamma\right]$$

$$D = \widehat{\mathbb{E}}\left[p u_\Gamma\right] = \mathbb{E}\left[\left(\alpha\left(\eta + u_S\right) + f\right) u_\Gamma\right] = \alpha \mathbb{E}\left[u_S u_\Gamma\right] = \alpha h^2 \sigma_u^2$$

For the numerator, we take the large N limit, where $u_{Et} = O\left(\frac{\sigma_u}{N^{1/2}}\right) \ll \sigma_\eta$, so to the leading order:

$$T^{1/2} A_T \sim \mathcal{N}\left(0, \sigma_A^2\right)$$

$$\sigma_A^2 = \mathbb{E}\left[\eta^2 u_\Gamma^2\right] = \mathbb{E}\left[\eta^2\right]\mathbb{E}\left[u_\Gamma^2\right] = \sigma_\eta^2 h^2 \sigma_u^2$$

so that

$$\frac{\sigma_A}{D} = \frac{\sigma_\eta h \sigma_u}{\alpha h^2 \sigma_u^2} = \frac{\sigma_\eta}{\alpha h \sigma_u} =: \sigma_\beta$$

Hence,

$$\sqrt{T}\left(\hat{\beta}_T - \beta\right) \sim N\left(0, \sigma_\beta^2\right)$$

□

8 Appendix: Extensions

8.1 Outliers and non-linearities

Suppose that we wish to construct the GIV for an aggregate growth

$$G_t = \sum_i S_{i,t-1} \phi\left(\delta_{it}\right) \tag{90}$$

where

$$\delta_{it} = \eta_t + u_{it} \tag{91}$$

and ϕ is a potentially non-linear growth rate function. In our main application, $\phi\left(\delta\right) = \delta$, but we could have e.g. $\phi\left(\delta\right) = e^{\delta} - 1$, or $\phi\left(\delta\right) = \tanh \delta$, which is between -1 and 1, or some other such function.

We can do the following. First, we estimate η_t. In the simplest case, we can just take $\eta_t^e \coloneqq \delta_{Et}$. In the simple case above, if there are outliers in u_{it}, we might take a median-based estimator of η_t^e

$$\eta_t^e = \mathcal{M}\left(\delta_{it}\right) - \frac{1}{T}\sum_{t'=1}^{T} \mathcal{M}\left(\delta_{it'}\right) \tag{92}$$

where $\mathcal{M}$ is the median.

Given η_t^e, we construct the proxy idiosyncratic shock:

$$u_{it}^e \coloneqq \delta_{it} - \eta_t^e \tag{93}$$

When the models are model complicated, as in (14), we can also by the procedure of Section 2.3, perhaps adapted to handle outliers. We form likewise the proxy for the idiosyncratic shock, $u_{it}^e = \delta_{it} - (k_i + m'C_{it} + \beta p_t + \lambda_{it}\eta_t)$.

Given the proxy u_{it}^e, we form the instrument:

$$z_t \coloneqq \sum_{i \in J_t} S_{i,t-1} \phi\left(u_{it}^e\right) \tag{94}$$

When $J_t = \{1, \ldots, N\}$is the set of all firms, what we have an instrument made of only exogenous shocks. As a refinement, to avoid potentially problematic outliers, we can set e.g.

$$J_t = \{i : |u_{it}^e| \leq q_t\}$$

where q_t is some suitable quantile, e.g. the middle 90% of the distribution.

8.2 Forward and Backward Looking Behavior

In some applications (e.g. finance or oil) it is sensible to have a backward or forward looking behavior. For concreteness, let us us the example of oil supply. We might model oil supply as:

$$\delta_{it} = \phi\left(p_t - \mathbb{E}_t p_{t+1}\right) + b\left(p_t - p_{t-1}\right) + \beta p_t + \eta_t + u_{it} \tag{95}$$

This is the same supply function as in the main text except for the first two terms on the right-hand side. The first one, $\phi\left(p_t - \mathbb{E}_t p_{t+1}\right)$, capture forward looking behavior: less oil is produced (or, released on oil markets) if current oil prices are low compared to future oil prices (with $\phi > 0$). The second one, $b\left(p_t - p_{t-1}\right)$, captures a smoothing behavior: if prices go up, one produces more oil, to smooth prices. In the context of oil, the term $b\left(p_t - p_{t-1}\right)$ seems important (see Hamilton (2009), Figure 5).[4]

Let us discuss the identification. First, suppose that $\phi = 0$. Then, we can proceed as before, the coefficient b estimated by OLS, with p_{t-1} as a regressor. We just need one GIV, as is standard in this paper.

Next, suppose that $\phi > 0$. Then, we need *two* instruments (two GIVs): morally, one for p_t, one for $p_t - \mathbb{E}_t p_{t+1}$. One way to do that would be to divide countries e.g. two types $k = 1, 2$ of u_{it}, some with high ρ_u (call them the "stable" countries), some with low ρ_u (call them the "unstable" countries). (A variant is to have an instrument for demand shock, another for supply shocks; this works if they are strong enough and they have different autocorrelations). Then, we form two $u_{\Gamma t,k}$, and use the two instruments for the regression. More rigorously, the moment condition is:

$$\mathbb{E}\left[\left(\delta_{Et} - \left(\phi\left(p_t - \mathbb{E}_t p_{t+1}\right) + b\left(p_t - p_{t-1}\right) + \beta p_t\right)\right)\left(u_{\Gamma 1t}, u_{\Gamma 2t}, p_{t-1}\right)\right] = 0$$

Constructing those $u_{\Gamma t,k}$ requires a bit of care. If things are i.i.d., we can take: $u_{\Gamma t,k} = \frac{\sum_{i \in I_k} S_i \delta_{it}}{\sum_{i \in I_k} S_i} - \delta_{Et}$. In an iteration of this paper, we'll discuss GIVs when shocks are autocorrelated.

8.3 Extension of the framework to heterogeneous β_i

Here we consider an extension of the procedure in Section 2.3 when we have heterogeneous β_i, and possibly actions and observables p_t of identical dimension possibly greater than 1 (for instance, a firm shock can comprise an increase in productivity, and in labor demand). We have:

$$\delta_{it} = k_i + m' C_{it} + \beta_{it} p_t + \lambda_{it} \eta_t + u_{it}, \tag{96}$$

[4] In the case of oil, one might replaced $b\left(p_t - p_{t-1}\right)$ by $-B\Delta\delta_{St}$, i.e. if the aggregate supply is lower, then more supply is released (e.g. by Saudi Arabia). The term $b\left(p_t - p_{t-1}\right)$ may be advantageous, as it smooths both supply and demand shocks.

where p_t is a vector of endogenous variables. Sensitivity β_i can be heterogeneous, and we suppose:

$$\beta_{it} = \beta_0 + \beta_x x_{it} + \zeta_i^{\beta}, \tag{97}$$

We also suppose:

$$p_t = \alpha \delta_{St} + \phi C_t + \varepsilon_t$$

where C_t can be correlated with ε_t, η_t but not with u_{it}. We assume

$$\mathbb{E}\left[(\varepsilon_t, C_t)\, u_{it}\right] = 0.$$

We also suppose that the dimension of u_{it} and p_t have the same dimension, K, which could be greater than 1. Then, α, β_{it} are $K \times K$ matrices.

We procedure is then as follows.

1. We estimate a panel regression with time and "agent" fixed effects

$$\delta_{it} = m' C_{it} + a_t + k_i + B_t x_{it} + v_{it}, \tag{98}$$

   where

$$\begin{aligned} a_t &= \beta_0 p_t + \eta_t, & (99)\\ B_t &= \beta_x p_t + \lambda \eta_t, & (100)\\ v_{it} &= \zeta_i^{\beta} p_t + \zeta_i \eta_t + u_{it}. & (101) \end{aligned}$$

2. We regress:

$$B_t = \Phi p_t + b_t, \tag{102}$$

   which gives, with $\kappa \coloneqq \mathbb{E}\left[p_t p_t'\right]^{-1} \mathbb{E}\left[p_t \eta_t\right]$ the coefficient of projection of η_t on p_t,

$$\begin{aligned} \Phi &= \beta_x + \lambda \kappa & (103)\\ b_t &= \lambda\left(\eta_t - \kappa p_t\right) & (104) \end{aligned}$$

3. We regress

$$v_{it} = \gamma_i^{\beta} p_t + \gamma_i b_t + u_{it}, \tag{105}$$

   using the instrument $p_{it} \coloneqq p_t - \left(1 - \alpha \beta_{S,t}\right)^{-1} u_{it}^e$ for p_t.[5] Then, we collect the residuals. This

[5] To purge the impact of u_{it} on p_t we use the fact that the full impact of u_{it} on p_t is:

$$\frac{dp_t}{du_{it}} = \left(1 - \alpha \beta_S\right)^{-1} \alpha S_i \tag{106}$$

also yields $\gamma_i = \frac{\zeta_i}{\lambda}$. Note there is a loop there, as we need a prior estimate for α, β_{St} and u^e_{it}. The instrument is $u_{\Gamma t}$. In the special case where $\zeta_i = \zeta_i^\beta = 0, \forall i$, we directly have $v_{it} = u_{it}$.

4. To estimate α and ϕ, we use the moment condition:

$$\mathbb{E}\left[\left(p_t - \alpha\delta_{St} - \phi C_t\right)\left(u'_{\Gamma t}, C_t\right)\right] = 0 \tag{107}$$

For instance, when there is no control C_t, we use the estimator:

$$\alpha = \widehat{\mathbb{E}}_T\left[\delta_{St} u_{\Gamma t}\right]^{-1} \widehat{\mathbb{E}}_T\left[p_t u_{\Gamma t}\right]. \tag{108}$$

5. We finish with the estimation of the $\beta_i's$. We use first $a_t = \beta_0 p_t + \eta_t$, which gives:

$$\mathbb{E}\left[\left(a_t - \beta_0 p_t\right) u'_{\Gamma t}\right] = 0$$

which gives β_0, and $\eta^e_t \coloneqq a_t - \beta_0 p_t$. Next, we regress $B_t = \beta_x p_t + \lambda\eta^e_t$ to get β_x and λ. Finally, we regress (101), using the instrument p_{it} for p_t, which gives ζ_i^β. Together, they give 97.