Granular Instrumental Variables*

Xavier Gabaix and Ralph S.J. Koijen

April 4, 2023

Abstract

We propose a new method to construct instruments in a broad class of economic environments. In the economies we study, a few large firms, industries or countries account for an important share of economic activity. As the idiosyncratic shocks to these large players affect aggregate outcomes, they are valid and often strong instruments. We provide a methodology to extract idiosyncratic shocks from the data and create "granular instrumental variables" (GIVs), which are size-weighted sums of idiosyncratic shocks. These GIVs allow us to estimate causal parameters of interest, including elasticities and multipliers. For instance, in a basic supply and demand framework, GIVs provide a novel approach to identify both supply and demand elasticities via idiosyncratic shocks to either supply or demand. We then show how to extend the basic procedure to cover a range of settings that are relevant to empirical research.

*`xgabaix@fas.harvard.edu`, `ralph.koijen@chicagobooth.edu`. We thank Manav Chaudhary, Aditya Chaudhry, Antonio Coppola, Zhiyu Fu, Rishab Guha, Dong Ryeol Lee, Constantin Schesch, Lingxuan Wu, and Yingxuan Yang for excellent research assistance. For helpful comments we thank the editor and referees, Isaiah Andrews, Miriam Arden, Jushan Bai, Stéphane Bonhomme, Gary Chamberlain, Jiafeng Chen, Raj Chetty, Gabriel Chodorow-Reich, Mark Gertler, Paul Goldsmith-Pinkham, Robin Greenwood, James Hamilton, Nathan Hendren, Guido Imbens, Oleg Itskhoki, Larry Katz, Lutz Kilian, Sofonias Korsaye, Jens Kvaerner, Karen Lewis, Erik Loualiche, Matteo Maggiori, Isabelle Méjean, Serena Ng, Ashesh Rambachan, Jean-Marc Robin, Kelly Shue, Jim Stock, Elie Tamer, Otto Tizian, Harald Uhlig, Thomas Wiemann, Christian Wolf, Tiemen Woutersen, Dacheng Xiu, and seminar participants at various conferences. We gratefully acknowledge financial support from the Ferrante Economics Research Fund and the Sloan Foundation (Gabaix) and the Center for Research in Security Prices and the Fama Research Fund at the University of Chicago Booth School of Business (Koijen).

1 Introduction

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

We first illustrate the idea in a basic static setup with supply and demand (Section 2). In this classic setting, we show how GIVs allow for a novel estimation procedure: they yield an instrument that allows us to estimate the price elasticities of both supply and demand. Intuitively, idiosyncratic demand shocks to large firms or countries provide valid instruments for changes in demand – and thus allow us to estimate the elasticity of supply. They also allow us to estimate the elasticity of demand: the idiosyncratic demand shock of a large firm impacts the price, which changes the demand of other firms. We formalize these ideas and present a way to extract and optimally aggregate idiosyncratic shocks, thus constructing optimal GIVs: the optimal GIV is the size-weighted sum of the idiosyncratic shocks. In this basic setup, we also show that some parameters that are of interest to economists, such as the pass-through of aggregate shocks to prices, can simply be estimated via OLS using GIVs. We establish the consistency of the OLS and IV estimators using GIVs and we derive their asymptotic distributions.

We analyze the instrument strength of GIVs in Section 3. The main insight is that GIVs are strong instruments when, first, a few large actors account for a substantial fraction of aggregate economic activity and, second, when idiosyncratic shocks are volatile relative to the volatility of aggregate shocks.

In the same section, we also discuss the robustness of GIVs to various forms of misspecification. The IV estimator using GIVs is robust to including only a subset of the large actors, mismeasurement of the size distribution, and heterogeneity in demand elasticities. The OLS estimator using GIVs does require that the size distribution is correctly measured. The main threat to identification when using GIVs is that we do not correctly isolate the idiosyncratic shocks. In this case, the estimated idiosyncratic shocks are affected by aggregate shocks, which can lead to bias. We provide several concrete solutions to mitigate that concern. First, we show that we can add factors for which we need to estimate the loadings. Second, we can "narratively check" GIVs: as the GIV procedure provides the dates and magnitudes of idiosyncratic and aggregate shocks, the largest idiosyncratic

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

shocks can be analyzed in their historical context to confirm their idiosyncratic nature. Third, when a large idiosyncratic shock coincides with an unusual aggregate event (which we label a sporadic factor realization), it is prudent to remove those dates. Fourth, we can perform over-identification tests: for instance, split countries into two groups (e.g., developing versus developed countries), form two GIVs using the idiosyncratic shocks of each group, and test whether the resulting estimates are statistically different.

Section 4 shows how the basic procedure extends to a range of settings that are relevant to empirical research that include (but are not limited to) heterogeneous demand elasticities, heteroskedasticity, time-varying factor loadings, and multidimensional outcomes. This generality stems from the fact that the basic intuitive idea is general, namely to use large, idiosyncratic shocks as primitive disturbances to the system.

Uses of GIVs Several recent papers have already applied GIV procedures to identify key parameters and elasticities of interest. Chodorow-Reich et al. (2021) study the multiplier of idiosyncratic shocks to an insurer's asset portfolio on the insurer's equity valuation. Camanho et al. (2022) study the impact of currency flows on exchange rates, using idiosyncratic shocks to fund-level rebalancing. Galaasen et al. (2021) use GIVs to study how idiosyncratic shocks to firms impact banks, and how this spills over to other (small, non-granular) firms borrowing from the same bank. Schubert et al. (2022) study the impact of concentration on wages and use idiosyncratic firm-level shocks to instrument for concentration. Kundu and Vats (2021) estimate how idiosyncratic firm-level shocks in one state affect economic activity in other states via their transmission through the banking system. Adrian et al. (2022) use bank-level idiosyncratic shock to estimate growth-at-risk. Ma et al. (2022) estimate how lenders' expectations about a city affect GDP growth in the same geography. Dong et al. (2022) study the impact of flows on factor returns using GIVs. Flynn and Sastry (2022) use GIVs to study how narratives spread across firms over time. We developed the GIV while working on Gabaix and Koijen (2022), where we use it to measure the elasticity of the aggregate stock market using idiosyncratic demand shocks to large investor sectors.

Related literature We relate to a number of strands. An active literature discusses identification strategies in macroeconomics (Ramey (2016); Nakamura and Steinsson (2018); Chodorow-Reich (2019); Huber (2023)). We add to it by proposing GIVs, which provide a systematic candidate approach to identification.

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

The idea of using idiosyncratic shocks as instruments to estimate spillover effects has been explored in several creative papers, as we discuss in more detail in Section D.1, such as Leary and Roberts (2014), 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 when extracting idiosyncratic shocks. Section D.1 contains a fuller discussion of the literature.

Outline Section 2 introduces the GIV framework, centered around a classic model of supply and demand. Section 3 gives an economic discussion of the robustness and potential threats to identification, and proposes diagnostic tests. Section 4 presents a number of extensions. Section 5 concludes. Proofs are in the Appendix, or in the Online Appendix, which also presents additional extensions.

Notations For a vector $X = (X_i)_{i=1\ldots N}$ and a series of weights w_i, we define $X_w = \sum_i w_i X_i$. With size weights S_i that satisfy $\sum_{i=1}^N S_i = 1$, we define $X_E := \frac{1}{N}\sum_{i=1}^N X_i$ and $X_S := \sum_{i=1}^N S_i X_i$ so that X_E is the equal-weighted average of the vector's elements and X_S is their size-weighted average. We also commonly use the notation u_i for shocks that are uncorrelated across i's and with variance $\sigma_{u_i}^2$. We use the notation θ_T^e for an estimator, based on observations from T periods, of a parameter θ whose true value is θ_0. We use C_t for a vector of controls; I for the identity matrix, $\iota = (1, \ldots, 1)'$ for a vector of ones, all of the appropriate dimension given the context; V^Y for the variance-covariance matrix of a random vector Y_t (so $V^Y = \mathbb{E}[Y_t Y_t']$ if Y_t is a mean zero column vector with constant variance). We use notation (A, B) for the concatenation of two matrices A, B. We use the notation $a_t, b_t \perp x_t, y_t$ to signify that variables a_t and b_t are uncorrelated with x_t and y_t.

2 GIVs: Main results

In this section, we formally introduce GIVs at a fairly high level of generality. Section 3.1 contains an instructive simple case. We also establish consistency of the estimators and derive their asymptotic distributions.

2.1 The baseline GIV

Setup Our goal is to estimate the scalar parameters ψ and ϕ^d in the following prototypical model:

$$p_t = \psi y_{St} + C_t^p m^p + \varepsilon_t, \tag{1}$$

$$y_{it} = \phi^d p_t + C_{it}^y m^y + \lambda_i \eta_t + u_{it}, \tag{2}$$

where $y_{St} \coloneqq \sum_i S_i y_{it}$, with S_i the weight on entity $i \in \{1, \ldots, N\}$, with $\sum_i S_i = 1$. An economic interpretation is that of demand and supply for a given good, e.g. oil.[2] Each entity's log demand y_{it} depends on the log price p_t of that good as in (2), with demand elasticity ϕ^d, and some shocks. The price in turn depends on aggregate demand, y_{St}, which is the size-weighted demand, with a sensitivity ψ that is the inverse of the supply elasticity. We impose that $\psi\phi^d \neq 1$, which ensures the existence of a unique solution of our system (1)-(2).

The econometrician observes S_i, p_t, y_{it} and the controls C_t^p and C_{it}^y, which include constants and entity fixed effects. C_t^y is the vector of C_{it}^y's. The parameters ψ, ϕ^d, m^p, and m^y have to be estimated, and we are chiefly interested in ψ and ϕ^d. The shocks ε_t, η_t and $u_t = (u_{it})_{i=1\ldots N}$ are not observed. The demand-side aggregate shocks are captured by a factor model of dimension r, $\lambda_i \eta_t = \sum_{f=1}^r \lambda_i^f \eta_t^f$, so that λ_i and η_t have dimensions $1 \times r$ and $r \times 1$, respectively. We assume that the first component of λ_i is 1, and the other components have been normalized to have mean 0, so we can write $\lambda_i = \left(1, \check{\lambda}_i\right)$—this is just a normalization. We call $\lambda = (\lambda_i)_{i=1\ldots N}$ and $\check{\lambda} = \left(\check{\lambda}_i\right)_{i=1\ldots N}$ the $N \times r$ and $N \times (r-1)$ matrices collecting the λ_i's and $\check{\lambda}_i$'s respectively. The normalization implies $\iota' \check{\lambda} = 0_{1 \times (r-1)}$.

Throughout the paper, the number of entities N is fixed, and we study estimators in the limit where the number of periods $T \to \infty$. We view the data $(y_t, p_t, C_t^y, C_t^p)_{t=1\ldots T}$ as sampled i.i.d. across periods,[3] and take the vector of weights S_i to be fixed (i.e. non-random).

Assumptions made throughout the paper We maintain the following assumptions throughout the paper. All random variables have finite fourth moments. The aggregate and idiosyncratic shocks, η_t, ε_t, and u_t, are uncorrelated across periods, and have mean 0. The controls C_t^p and C_t^y are mutually uncorrelated with η_t and ε_t: $C_t^p, C_t^y \perp \eta_t, \varepsilon_t$.

The central assumption is that shocks u_{it} are idiosyncratic, that is, they are uncorrelated with the aggregate shocks, η_t and ε_t, and with controls, C_t^y and C_t^p. Formally,

$$u_t \perp \eta_t, \varepsilon_t, C_t^y, C_t^p. \tag{3}$$

Assumptions made for expositional convenience We sometimes make assumptions that simplify the exposition or the reasoning, typically with minimal impact on the economics. We will indicate in the remainder of the paper when we make these auxiliary assumptions.

Assumption 1 *The λ_i are known.*

Assumption 1 is weaker than it seems. For instance, it holds when the λ_i simply have a parametric form, $\lambda_i = X_i \dot{\lambda}$, where λ_i and X_i are r-dimensional row vectors and $\dot{\lambda}$ is an $r \times r$ matrix. The vector of characteristics X_i are observed. This implies that $\lambda = X\dot{\lambda}$ where λ and X are $N \times r$

[2]Section D.2 of the Online Appendix contains a simple economic model leading to this structure.

[3]One could extend this, e.g. to a setting with stochastic volatility, but we keep the i.i.d. assumption for simplicity.

matrices. This assumption simplifies the exposition and reasoning, but it will be relaxed in Section 4.2. As one can change η_t into $\dot{\lambda}\eta_t$, one might normalize $\dot{\lambda} = I_r$.

The following three assumptions also simplify the exposition and we will relax each of them later in the paper.

Assumption 2 *The controls have the shape:* $C_{it}^y = \lambda_i c_t^y$.

Assumption 3 *The* u_{it} *are uncorrelated across* i*'s and homoskedastic with variance* $\sigma_u^2 > 0$.

Assumption 4 *The* u_{it} *are heteroskedastic with variance-covariance matrix* $V^u = \mathbb{E}\left[u_t u_t'\right]$ *that is non-singular and constant over time.*

The core idea of GIVs The idea of GIVs is to use idiosyncratic shocks u_{it} as instruments. We construct GIVs as follows, where we impose Assumption 2 for now. Suppose that we have a set of weights $\Gamma \in \mathbb{R}^N$ orthogonal to factor loadings λ but not to the size vector S:

$$\Gamma'\lambda = 0, \qquad \Gamma' S \neq 0. \tag{4}$$

This is possible when S is not spanned by the loadings λ.[4] We provide a concrete and optimal construction of Γ in (17). Then the GIV is defined as:

$$z_t := \Gamma' y_t = \sum_{i=1}^{N} \Gamma_i y_{it}. \tag{5}$$

Hence, the GIV z_t is constructed from observables, y_{it}. As $\Gamma'\lambda = 0$, we have $\Gamma'\iota = 0$ (recall that $\iota = (1, \ldots, 1)'$ and that the first column vector of λ is a 1). The key observation is that (using Assumption 2 for now),[5] $z_t := \Gamma' y_t = \Gamma'\left(\iota \phi^d p_t + \lambda c_t^y + \lambda \eta_t + u_t\right) = \Gamma' u_t$, so

$$z_t = \Gamma' u_t. \tag{6}$$

As a result, the GIV z_t is a linear combination of idiosyncratic shocks. The GIV satisfies the exogeneity condition by (3):

$$\text{Exogeneity: } z_t \perp \eta_t, \varepsilon_t, \tag{7}$$

and the relevance condition because $\Gamma' S \neq 0$:

$$\text{Relevance: } \mathbb{E}\left[y_{St} z_t\right] \neq 0. \tag{8}$$

[4] This is, when there is no vector b such that $S = \lambda b$.

[5] Alternatively, the reader is encouraged to think of the even simpler case where there are no controls, $C_t^y = 0$.

Then, (1) implies the moment condition:[6]

$$\mathbb{E}\left[\left(p_t - \psi y_{St}\right) z_t\right] = 0, \tag{9}$$

and we can estimate ψ using the GIV z_t as an instrument, as $\psi = \frac{\mathbb{E}[p_t z_t]}{\mathbb{E}[y_{St} z_t]}$. Its natural empirical counterpart is the estimator $\psi_T^e = \frac{\sum_{t=1}^T p_t z_t}{\sum_{t=1}^T y_{St} z_t}$.

To estimate ϕ^d, we take a vector $\tilde{E}$ such that

$$\mathbb{E}\left[u_{\tilde{E}t} u_{\Gamma t}\right] = 0, \qquad \iota' \tilde{E} = 1. \tag{10}$$

For instance, with homoskedastic residuals, we can take $\tilde{E}_i = \frac{1}{N}$, which leads us to call $\tilde{E}$ a "quasi-equal weight" vector more generally.[7] As $y_{\tilde{E}t} := \sum_i \tilde{E}_i y_{it} = \phi^d p_t + C_{\tilde{E}t}^y m^y + \lambda_{\tilde{E}} \eta_t + u_{\tilde{E}t}$, with $u_{\tilde{E}t} := \sum_i \tilde{E}_i u_{it}$, (10) implies

$$\mathbb{E}\left[\left(y_{\tilde{E}t} - \phi^d p_t\right) z_t\right] = 0. \tag{11}$$

The relevance condition is then

$$\text{Relevance: } \mathbb{E}\left[p_t z_t\right] \neq 0. \tag{12}$$

This condition is satisfied if $\psi \neq 0$, that is, when demand shocks influence the price. Then, we can estimate ϕ^d using z_t as an instrument, via $\phi^d = \frac{\mathbb{E}\left[y_{\tilde{E}t} z_t\right]}{\mathbb{E}[p_t z_t]}$. Its empirical counterpart is $\phi_T^{d,e} = \frac{\sum_{t=1}^T y_{\tilde{E}t} z_t}{\sum_{t=1}^T p_t z_t}$. The following proposition establishes the consistency of the GIV estimator.

Proposition 1 (Consistency of the GIV estimator). *Let Assumptions 1 and 2 hold, and assume that* $\mathbb{E}\left[u_{St} u_{\Gamma t}\right] \neq 0$. *Define the GIV estimators* $\psi_T^e = \frac{\sum_{t=1}^T p_t z_t}{\sum_{t=1}^T y_{St} z_t}$ *and* $\phi_T^{d,e} = \frac{\sum_{t=1}^T y_{\tilde{E}t} z_t}{\sum_{t=1}^T p_t z_t}$. *Then,* ψ_T^e *is a consistent estimator of* ψ*: as* $T \to \infty$, $\psi_T^e \xrightarrow{a.s.} \psi$. *Likewise, if* $\psi \neq 0$, *then* $\phi_T^{d,e}$ *is a consistent estimator of* ϕ^d*: as* $T \to \infty$, $\phi_T^{d,e} \xrightarrow{a.s.} \phi^d$.

To think about the variance of the estimators, and, not coincidentally, the economics, we define

$$M := \frac{1}{1 - \psi \phi^d}, \qquad \mu := \psi M. \tag{13}$$

Those quantities are the pass-through from idiosyncratic and aggregate shocks to the aggregate quantity and the price, respectively, see (33) and (34).

Proposition 2 (Variance of the GIV estimator). *Let Assumptions 1 and 2 hold, and assume that* $\mathbb{E}\left[u_{St} u_{\Gamma t}\right] \neq 0$. *Assume that* $\varepsilon_t^\psi := C_t^p m^p + \varepsilon_t$ *is homoskedastic conditionally on* u_t *with variance*

[6] The error term, $p_t - \psi y_{St}$, includes aggregate shocks and controls. This moment condition holds, even though it is more efficient to account for controls in estimating ψ, see Section 2.3.

[7] See (**??**) for a proof. The heteroskedastic case will lead to (54).

$\sigma^2_{\varepsilon\psi}$. *Then, as* $T \to \infty$,

$$\sqrt{T}\left(\psi^e_T - \psi\right) \xrightarrow{d} N\left(0, \sigma^2_\psi\right), \qquad \sigma_\psi = \frac{\mathbb{E}\left[u^2_{\Gamma t}\right]^{1/2}}{\left|\mathbb{E}\left[u_{St}u_{\Gamma t}\right]\right|}\frac{\sigma_{\varepsilon\psi}}{|M|}. \tag{14}$$

Likewise, if $\mu \neq 0$, *then, assuming that* $\varepsilon^\phi_t := C^y_{\tilde{E}t}m^y + \lambda_{\tilde{E}}\eta_t + u_{\tilde{E}t}$ *has variance* $\sigma^2_{\varepsilon^\phi}$ *conditionally on* u_t,

$$\sqrt{T}\left(\phi^{d,e}_T - \phi^d\right) \xrightarrow{d} N\left(0, \sigma^2_{\phi^d}\right), \qquad \sigma_{\phi^d} = \frac{\mathbb{E}\left[u^2_{\Gamma t}\right]^{1/2}}{\left|\mathbb{E}\left[u_{St}u_{\Gamma t}\right]\right|}\frac{\sigma_{\varepsilon\phi}}{|\mu|}. \tag{15}$$

We provide further economic intuition for the asymptotic variances of the GIV estimators in Section 3.1. The condition $\mathbb{E}\left[u_{St}u_{\Gamma t}\right] \neq 0$ is generically true. For instance, with homoskedastic u_{it}, $\mathbb{E}\left[u_{St}u_{\Gamma t}\right] = \sigma^2_u\Gamma' S$, it is guaranteed by (4).

2.2 Optimal GIV weights and the precision of the GIV estimator

We now derive the optimal GIV weights Γ that minimize the asymptotic variances σ^2_ψ and $\sigma^2_{\phi^d}$ in Proposition 2. For this, we use the $N \times N$ projection matrix Q that is orthogonal to the factor loadings, that is, $Q\lambda = 0$:

$$Q := I - \lambda\left(\lambda'\lambda\right)^{-1}\lambda'. \tag{16}$$

Proposition 3 (Optimal weights Γ for the GIV $y_{\Gamma t}$). *Let Assumptions 1, 2, and 3 hold. The optimal weights*

$$\Gamma^{*\prime} = S'Q \tag{17}$$

minimize the asymptotic variances σ^2_ψ *and* $\sigma^2_{\phi^d}$ *of the GIV estimators in Proposition 2, with* $\frac{\mathbb{E}\left[u^2_{\Gamma t}\right]}{\mathbb{E}\left[u_{St}u_{\Gamma t}\right]^2} = \frac{1}{\mathbb{E}\left[u^2_{\Gamma^* t}\right]}$. *For any other* Γ *that is not proportional to* Γ^*, *the asymptotic variances* $\sigma^2_\psi(\Gamma)$ *and* $\sigma^2_{\phi^d}(\Gamma)$ *are strictly larger than for* Γ^*.

We summarize the intuition behind Proposition 3. Each entity i affects the price proportionally to its size S_i, see (1). Hence, the economically appropriate weights are S. However, we need to satisfy $\Gamma'\lambda = 0$ to remove the influence from aggregate shocks on the GIV. Proposition 3 shows that the optimal weights vector is Γ^*: it is the vector closest to S, while being orthogonal to the factor loadings λ.

2.3 Controlling for observed and latent factors

So far, we only used z_t to estimate the elasticities without accounting for observed and latent factors. We now show how to account for those in the procedure.

We proceed in two steps. First, in Section 2.3.1, we show that we can rewrite the demand equation in (2) in a more convenient form

$$y_{it} = \phi^d p_t + C^y_{it} m^y + \eta_{1t} + \check{\lambda}_i \check{\eta}_t + \check{u}_{it}, \tag{18}$$

where we replace $\lambda\eta_t = \eta_{1t} + \lambda_i \eta_t$ by $\eta_{1t} + \check{\lambda}_i \check{\eta}_t$, and u_{it} by $\check{u}_{it}$, where $\check{\eta}_t$ and $\check{u}_{it}$ are defined below. This reformulation has two advantages. First, when m^y and $\check{\lambda}_i$ are known, $\check{\eta}_t$ and $\check{u}_{it}$ can be recovered without error, unlike the η_t and u_{it} of the original factor model (2). Second, it holds that $\check{\eta}_t \perp \check{u}_t$. These two convenient features are then used in Proposition 4, which is the main result of this section.

2.3.1 Introducing recoverable factors

Before turning to this section's main result, we introduce the concept of "recoverable factors" that will simplify the idea and proofs that follow. We start from the factor model part of our system (calling it Y_t rather than y_t):

$$Y_t = \lambda \eta_t + u_t, \qquad \eta_t \perp u_t, \tag{19}$$

and we maintain Assumption 1 that we know λ. In this case, η_t cannot be recovered without error when the number of entities N is fixed, as it is in our case (one would need $N \to \infty$, see e.g. Bai (2003)). The main insight is that the estimates and residuals that we can recover without error, $\check{\eta}_t$ and $\check{u}_t$, are uncorrelated, just as the true values, η_t and u_t. These estimates can be computed even though N is fixed, and have useful properties that Lemma 1 spells out.

Lemma 1 (Recoverable factors and their properties) *Let Assumption 3 hold. Given a factor model (19), define the $r \times N$ and $N \times N$ matrices:*

$$R^\lambda := (\lambda'\lambda)^{-1} \lambda', \qquad Q^\lambda := I - \lambda R^\lambda \tag{20}$$

and

$$\check{\eta}_t := \eta_t + R^\lambda u_t, \qquad \check{u}_t := Q^\lambda u_t. \tag{21}$$

Then we have

$$Y_t = \lambda \check{\eta}_t + \check{u}_t, \qquad \check{\eta}_t, \eta_t \perp \check{u}_t. \tag{22}$$

In addition, $\check{\eta}_t$ and $\check{u}_t$ can be exactly recovered from Y_t and λ, via

$$\check{\eta}_t = R^\lambda Y_t, \qquad \check{u}_t = Q^\lambda Y_t. \tag{23}$$

In the remainder of this section, we work with the recoverable shocks $(\check{\eta}_t, \check{u}_t)$ rather than the non-recoverable shocks (η_t, u_t). Just as η_t and u_t are uncorrelated, $\check{\eta}_t$ and $\check{u}_t$ are uncorrelated as well—compare (19) and (22). However, proofs and concepts are easier working with $(\check{\eta}_t, \check{u}_t)$ rather than with (η_t, u_t) as $(\check{\eta}_t, \check{u}_t)$ are directly recovered from Y_t (see (23)).[8] By a slight abuse of language, we will refer to $\check{\eta}_t$ as common factors and $\check{u}_t$ as a vector of idiosyncratic shocks.[9]

Computing $\check{\eta}_t$ and $\check{u}_t$ in Lemma 1 corresponds to running cross-sectional regressions of Y_{it} on λ_i, for every date t:

$$Y_{it} = \lambda_i \check{\eta}_t + \check{u}_{it}, \tag{24}$$

which yields an exact recovery of $\check{\eta}_t$ and $\check{u}_{it}$.

2.3.2 GIV, controlling for observable and recovered controls

Before stating the procedure with controls, we need a bit more notation. We decompose $\eta_t = (\eta_{1t}, \eta_{2t})$, where η_t, η_{1t}, and η_{2t} have dimensions r, 1, and $(r-1)$. We apply the idea of Lemma 1 to $\check{\lambda}$ (recall we decomposed $\lambda_i = \left(1, \check{\lambda}_i\right)$ with $\iota' \check{\lambda} = 0_{1\times(r-1)}$), hence we define:

$$\check{\eta}_t := R^{\check{\lambda}} \left(\eta_t + \lambda u_t\right) = \eta_{2t} + R^{\check{\lambda}} u_t, \qquad \check{u}_t := Q^{\lambda} u_t. \tag{25}$$

Then, we can recover $\check{\eta}_t$ and $\check{u}_t$, and we still have $\check{\eta}_t \perp \check{u}_t$. [10] We decompose $\eta_{1t} = b^1 \check{\eta}_t + \eta_{1t}^{\perp}$ and $\varepsilon_t = b^{\varepsilon} \check{\eta}_t + \varepsilon_t^{\perp}$, where $\eta_{1t}^{\perp}$ and $\varepsilon_t^{\perp}$ are uncorrelated with $\check{\eta}_t$: they are the residuals after projecting on $\check{\eta}_t$. We call $\check{y}_{it} = y_{it} - y_{\tilde{E}t}$ the cross-sectionally demeaned value with quasi-equal weights $\tilde{E}$ satisfying (10). Defining $b^y := \check{\lambda}_{\tilde{E}} + b^1$ and $\varepsilon_t^y := \eta_{1t}^{\perp} + u_{\tilde{E}t}$, we write $y_{\tilde{E}t}$ as

$$y_{\tilde{E}t} = \phi^d p_t + b^y \check{\eta}_t + C_{\tilde{E}t}^y m^y + \varepsilon_t^y. \tag{26}$$

Proposition 4 (GIV estimation with controls) *Let Assumptions 1 and 3 hold. Define $\check{y}_{it} := y_{it} - y_{Et}$ and similarly $\check{C}_{it}^y := C_{it}^y - C_{Et}^y$. Given a candidate value m^y, we construct $\check{\eta}_t\left(m^y\right) := R^{\check{\lambda}}\left(y_t - C_t^y m^y\right)$ and the GIV*

$$z_t\left(m^y\right) := \Gamma'\left(y_t - C_t^y m^y\right). \tag{27}$$

with Γ satisfying (4). Define θ to be the collection of parameters $\left(\psi, \phi^d, m^p, b^p, m^y, b^y\right)$ and θ_0 the true value of θ. Define θ_T^e to be the estimator of θ_0 that solves the following sample moments:

$$\sum_t \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y\right) + \check{C}_t^y m^y\right)' \check{C}_t^y = 0, \tag{28}$$

$$\sum_t \left(-p_t + \psi y_{St} + b^p \check{\eta}_t\left(m^y\right) + C_t^p m^p\right)\left(z_t\left(m^y\right), \check{\eta}_t\left(m^y\right)', C_t^{p\prime}\right) = 0. \tag{29}$$

[8]Even when λ is not known, $(\check{\eta}_t, \check{u}_t)$ are still very useful. See for instance Proposition 7.

[9]Even if the u_{it} are uncorrelated across i's, the $\check{u}_{it}$ will have some slight correlation across i's since$V^{\check{u}} = Q\sigma_u^2$ This sometimes requires a bit of care, which we take.

[10]Running the regression $y_{it} - C_{it}^y m^y = a_t + \check{\lambda}_i \check{\eta}_t + \check{u}_{it}$ on the known $\check{\lambda}_i$ yields $\check{\eta}_t$ and $\check{u}_{it}$, see (28). This holds because $\check{\eta}_t = R^{\check{\lambda}}\left(y_t - C_t^y m^y\right)$, $\check{u}_t = Q^{\check{\lambda}}\left(y_t - C_t^y m^y\right)$.

To estimate ϕ^d (assuming $\psi \neq 0$), we use the additional moment condition:

$$\sum_t \left(-y_{\tilde{E}t} + \phi^d p_t + b^y \check{\eta}_t (m^y) + C^y_{\tilde{E}t} m^y\right) \left(z_t (m^y), \check{\eta}_t (m^y)', C^{y\prime}_{\tilde{E}t}\right) = 0. \tag{30}$$

Under the additional regularity Assumption 5 from Appendix A, θ^e_T is consistent and $\sqrt{T}-$asymptotically normal. Assuming that $\varepsilon^\perp_t$ and ε^y_t are homoskedastic conditionally on u_t, the variances of the estimators ψ^e_T and $\phi^{d,e}_T$ are as in Proposition 2, with the new values $\sigma^2_{\varepsilon^\psi} = var\left(\varepsilon^\perp_t\right)$ and $\sigma^2_{\varepsilon^\phi} = var\left(\varepsilon^y_t\right)$. In addition, the optimal Γ remains as in Proposition 3.

The upshot of Proposition 4 is that we can estimate the parameters as in our basic Proposition 2, except that the controls "soak up" more variance.[11] In addition, under Assumption 1, the standard errors of ψ^e_T and $\phi^{d,e}_T$ are the asymptotic standard errors that a "naive" IV estimator would report that ignores that m^y and $\check{\eta}_t$ have been estimated.

In summary, when there are no controls and under the assumptions of Proposition 4, the optimal GIV is

$$z_t := \Gamma^{*\prime} y_t = S' Q y_t = S' \check{u}_t, \tag{31}$$

with $\check{u}_t := Q u_t$. In other terms, $\check{u}_{it}$ is the residual from a cross-sectional regression of $\check{y}_{it} := y_{it} - y_{Et}$, on the demeaned factor loadings $\check{\lambda}_i$:

$$\check{y}_{it} := \check{\lambda}_i \check{\eta}_t + \check{u}_{it}. \tag{32}$$

This regression also recovers the aggregate shocks $\check{\eta}_t$, see Lemma 1. When there are controls, we just replace y_t by $y_t - C^y_t m^y$.

2.4 Using the GIV via OLS

So far, we used the GIV z_t in an IV form. We now show how to use it in estimating μ and M using OLS, where we defined μ and M in (13). First, we solve for y_{St}, which leads to:

$$p_t = \mu u_{St} + b^p c_t + \varepsilon^p_t, \tag{33}$$

$$y_{St} = M u_{St} + b^y c_t + \varepsilon^y_t, \tag{34}$$

where $c_t = (C^p_t, C^y_{St}, \check{\eta}_t)$ is a vector of controls, and $(\varepsilon^p_t, \varepsilon^y_t)$ are uncorrelated with u_{St}. If y_{it} is log demand and p_t is the log price, the interpretation is that a 1% demand shock leads to a μ% price increase and an M% supply increase. The intuition is that the size-weighted idiosyncratic shock u_{St} affects p_t with a strength μ and y_{St} with a strength M.

[11] The assumption $\psi \neq 0$ is only needed to estimate ϕ^d via (28) and (30). If $\psi = 0$, we can still estimate ψ by the moments (28)-(29).

Next, we introduce Γ satisfying (4), and such that

$$\frac{\mathbb{E}\left[u_{St}u_{\Gamma t}\right]}{\mathbb{E}\left[u_{\Gamma t}^2\right]} = 1. \tag{35}$$

Defining $\tilde{E} := S - \Gamma$, condition (35) is equivalent to having (10).[12] Then $u_{St} = u_{\Gamma t} + u_{\tilde{E}t}$, and

$$\begin{aligned} p_t &= \mu z_t + b^p c_t + e_t^p, & (36)\\ y_{St} &= M z_t + b^y c_t + e_t^y, & (37) \end{aligned}$$

with $e_t^p := \varepsilon_t^p + \mu u_{\tilde{E}t}$ and $e_t^y := \varepsilon_t^y + M u_{\tilde{E}t}$, which are orthogonal to z_t and c_t (recall (10)). It follows that we can estimate μ and M by an OLS regression of p_t and y_{St} on z_t.[13] The next proposition records that result and derives the precision of the estimators.

Proposition 5 (GIV estimator for OLS). *Let Assumptions 1 and 3 hold, and assume (35). Consider the procedure that (i) estimates m^y by the regression (28), where $\check{\eta}_t(m^y) := R^{\check{\lambda}}\left(y_t - C_t^y m^y\right)$ and $R^{\check{\lambda}} = \left(\check{\lambda}'\check{\lambda}\right)^{-1}\check{\lambda}'$; (ii) defines the GIV z_t by (27); (iii) estimates μ and M by the OLS regression coefficients of, respectively, (36) and (37), where $c_t = \left(C_t^p, C_{St}^y, \check{\eta}_t(m^y)\right)$ is a vector of controls. Then, under the additional regularity Assumption 5 from Appendix A, the OLS estimators μ_T^e and M_T^e are consistent. In addition, $\sqrt{T}\left(\mu_T^e - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\mu^2\right)$ and $\sqrt{T}\left(M_T^e - M\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_M^2\right)$. Assuming that e_t^p and e_t^y are homoskedastic conditionally on u_t, $\sigma_\mu^2 = \frac{var\left(e_t^p\right)}{var(z_t)}$ and $\sigma_M^2 = \frac{var\left(e_t^y\right)}{var(z_t)}$, and these asymptotic standard errors are those that an OLS procedure would report that did not account for the fact that the values of m^y and $\check{\eta}_t(m^y)$ were estimated. In addition, the optimal Γ remains as in Proposition 3.*

The fact that the standard errors are correct is a bit surprising at first as z_t is a generated regressor. When there are no other controls, $C_t^p = C_t^y = 0$, the reason is that the GIV is directly obtained from an exact formula ($z_t := S'Qy_t$ as in (31)) and can thus be constructed without error.[14] For this result, Assumption 1, and the fact that we know the σ_{u_i}, are important. If one needs to estimate λ or the σ_{u_i}'s, then that estimation error needs to be taken into account when calculating the standard errors of the elasticities, ψ and ϕ^d, and pass-throughs, M and μ (see Sections 4.2 and 4.3).

[12] So for instance, we can use $\Gamma = S - \tilde{E}$ with $\tilde{E}_i = \frac{1}{N}$ in the homoskedastic case, and (54) in the heteroskedastic case.

[13] Alternatively, in the regressions (36) and (37) one can replace z_t by $Z_t := y_{St} - y_{\tilde{E}t}$. Then, the result is the same, including the standard error of the estimators. This is simply because with $z_t = Z_t - \check{\lambda}_\Gamma \check{\eta}_t$, we control for $\check{\eta}_t$.

[14] When there is an m^y, the reason is more subtle, and revealed by the proof: an error in m^y creates an error proportional to C_t^y times random variables uncorrelated to it, hence a negligible term $o_p\left(T^{-1/2}\right)$.

3 Discussion

3.1 The intuition behind GIV estimators

3.1.1 GIVs in a simple model

To explain the intuition, we specialize our analysis to the case where there is only a single factor and all entities have the same loading on the factor, $\lambda_i = 1_{1\times r}$, and there are no other controls, $C_t^p = C_t^y = 0$. The single factor is then absorbed by a time fixed effect. It allows us to develop the main intuition in a transparent way. The system is

$$p_t = \psi y_{St} + \varepsilon_t, \tag{38}$$

$$y_{it} = \phi^d p_t + \eta_t + u_{it}. \tag{39}$$

To take advantage of the great analytical simplicity of that example, we retrace the derivation steps in an elementary manner. We cannot estimate ψ and ϕ^d by OLS as ε_t and η_t are typically correlated, implying that y_{St} is correlated with ε_t in (38), and p_t with η_t in (39).

We construct the GIV as the "size-weighted" average outcome, $y_{St} = \sum_i S_i y_{it}$, minus the "equal-weighted" average outcome, $y_{Et} = \frac{1}{N}\sum_{i=1}^N y_{it}$:

$$z_t := y_{\Gamma t} = y_{St} - y_{Et}. \tag{40}$$

Given that[15]

$$y_{St} = \phi^d p_t + \eta_t + u_{St}, \qquad y_{Et} = \phi^d p_t + \eta_t + u_{Et}, \tag{41}$$

the GIV is also

$$z_t = u_{\Gamma t} = u_{St} - u_{Et}, \tag{42}$$

and it is only made of idiosyncratic shocks. Given the exogeneity condition (3), we have $\mathbb{E}[u_t\varepsilon_t] = 0$, and hence $\mathbb{E}[z_t\varepsilon_t] = 0$. This gives (using (38)): $\mathbb{E}[(p_t - \psi y_{St}) z_t] = 0$ and thus $\psi = \frac{\mathbb{E}[p_t z_t]}{\mathbb{E}[y_{St} z_t]}$. Its empirical counterpart is the estimator $\psi_T^e = \frac{\sum_{t=1}^T p_t z_t}{\sum_{t=1}^T y_{St} z_t}$.

Likewise, given the exogeneity condition (3), we have $\mathbb{E}[z_t\eta_t] = 0$. In addition, under the Assumption 3 that the u_{it} are homoskedastic we have $\mathbb{E}[u_{Et}u_{\Gamma t}] = 0$.[16] So, using (41), we have $\mathbb{E}\left[\left(y_{Et} - \phi^d p_t\right) z_t\right] = 0$, hence, $\phi^d = \frac{\mathbb{E}[y_{Et} z_t]}{\mathbb{E}[p_t z_t]}$. Its empirical counterpart is $\phi_T^{d,e} = \frac{\sum_{t=1}^T y_{Et} z_t}{\sum_{t=1}^T p_t z_t}$.

Hence, the same instrument, the GIV z_t, can be used to estimate both the demand elasticity $\phi^d = \frac{\mathbb{E}[y_{Et} z_t]}{\mathbb{E}[p_t z_t]}$ and the supply elasticity $\phi^s = \frac{1}{\psi} = \frac{\mathbb{E}[y_{St} z_t]}{\mathbb{E}[p_t z_t]}$. Intuitively, an idiosyncratic shock to, for instance, a large country affects both world prices and quantities, so it allows us to estimate the elasticity of demand of the other countries, and the elasticity of supply, which is equal (as supply

[15] As is clear in this simple example, any average (other than size-weighted) of the y_{it} can be used to remove $\phi^d p_t + \eta_t$, not just y_{Et}. What is most efficient, however, is to use the equal-weighted average in this case.

[16] Indeed, we have $\Gamma' E = \sum_i \left(S_i - \frac{1}{N}\right)\frac{1}{N} = 0$. So $\mathbb{E}[u_{\Gamma t}u_{Et}] = \mathbb{E}[(\Gamma' u_t)(u_t' E)] = \Gamma'\mathbb{E}[u_t u_t'] E = \sigma_u^2 \Gamma' E = 0$.

equals demand) to the size-weighted demand of all countries.[17]

The general Propositions 1 and 2 show that those estimators are $\sqrt{T}$–consistent. Proposition 3 implies that this GIV (40) is the optimal one, as stated below.

Corollary 1 (GIV estimator in the case with a single factor and common loadings). *Let Assumption 3 hold. In the case with a single factor and common loadings, (38)-(39), the optimal GIV estimator of Proposition 3 takes the form* $\Gamma = S - E$*, that is,* $\Gamma_i = S_i - \frac{1}{N}$*:*

$$z_t := y_{\Gamma t} = y_{St} - y_{Et}, \tag{43}$$

so that $z_t = u_{\Gamma t} = u_{St} - u_{Et}$*.*

Hence, in this simplest of cases, the GIV is the "size-weighted" minus the "equal-weighted" value of outcomes, which recovers the the "size-weighted" minus the "equal-weighted" value of idiosyncratic shocks. In addition, $\check{u}_{it} = u_{it} - u_{Et}$, so that $z_t = \sum_i S_i \check{u}_{it}$. In the general case, intuitively, the GIV weights Γ mimic this structure, but they need to be adjusted to remove common factors if the factor structure is more complex, see (17).

3.1.2 GIV estimators are more precise in concentrated economies with volatile idiosyncratic shocks

We next study when the GIV achieves a good precision.

Corollary 2 (Precision of the GIV estimator in the case with a single factor and common loadings). *Let Assumption 3 hold. Consider the case with a single factor and common loadings, (38)-(39). The GIV estimators are consistent, and converge as* $\sqrt{T}\left(\psi_T^e - \psi\right) \xrightarrow{d} N\left(0, \sigma_\psi^2\right)$ *and (assuming* $\psi \neq 0$*)* $\sqrt{T}\left(\phi_T^{d,e} - \phi^d\right) \xrightarrow{d} N\left(0, \sigma_{\phi^d}^2\right)$*, where the asymptotic standard deviations of the scaled and centered GIV estimators are (assuming that* ε_t *and* $\varepsilon_t^\phi := \eta_t + u_{Et}$ *are homoskedastic conditionally on* u_t*):*

$$\sigma_z = \sigma_{u_\Gamma} = h\sigma_u, \qquad \sigma_\psi = \frac{\sigma_\varepsilon}{\sigma_z \left|M\right|}, \qquad \sigma_{\phi^d} = \frac{\sigma_{\varepsilon^\phi}}{\sigma_z \left|\mu\right|}, \tag{44}$$

where h *is the excess Herfindahl:*

$$h := \sqrt{\left(\sum_{i=1}^N S_i^2\right) - \frac{1}{N}}. \tag{45}$$

[17] As the supply elasticity $\phi^s = \frac{\mathbb{E}[y_{St} z_t]}{\mathbb{E}[p_t z_t]}$ traces out how the supply side reacts to the price, we need to use aggregate supply y_{St}, which is, in equilibrium, the size-weighted average of the demands. On the other hand, the demand elasticity $\phi^d = \frac{\mathbb{E}\left[y_{\tilde{E}t} z_t\right]}{\mathbb{E}[p_t z_t]}$ indicates how the demand side of an individual country reacts to the price, so one can take many weights $\tilde{E}$, provided that they satisfy (10). It is most efficient, however, to take the equal-weighted average of demands of individual countries, $\tilde{E}_i = E_i = \frac{1}{N}$, to maximally smooth out idiosyncratic noise.

Corollary 2 highlights what it takes to obtain a strong instrument and a precise estimate of ψ (and similarly of ϕ^d): we need one or several large units (in order to have a large excess Herfindahl h) and we need idiosyncratic shocks to be volatile compared to the volatility of aggregate shocks (large σ_u compared to $\sigma_{\varepsilon^\psi}$ or $\sigma_{\varepsilon^\phi}$).[18]

3.2 Instrument strength

Two-stage least squares One can view regression (36) as the first-stage regression of the price on the GIV and controls c_t, which yields the instrumented price $p_t^e := \mu^e z_t + b^p c_t$. In the second stage, we regress supply on the instrumented price and controls c_t:

$$y_{St} = \frac{1}{\psi} p_t^e + \beta^{y_S} c_t + \varepsilon_t^{y_S,\mathrm{2SLS}}, \tag{46}$$

which gives an estimate of the supply elasticity $\frac{1}{\psi}$. To estimate the demand elasticity, we regress the equal-weighted (not size-weighted) demand on the instrumented price, p_t^e, and controls c_t in the second stage: $y_{\tilde{E}t} = \phi^d p_t^e + \beta^{y_{\tilde{E}}} c_t + \varepsilon_t^{y_{\tilde{E}},\mathrm{2SLS}}$. This yields estimates that are the same as the IV estimators from Proposition 4.

GIVs and weak instruments In case of the OLS estimator, a low power of the GIV simply manifests itself as large standard errors, while in case of the IV estimators, we can encounter weak instrument problems. For the parameters that can be estimated using OLS, that is, μ and M, the standard errors obtained via standard OLS inference are valid when T is large (as per Proposition 5, and under its assumptions). When a ratio is implicitly performed, for instance to estimate ϕ^d or ψ by instrumental variables, the two-stage least squares (2SLS) procedure 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-statistic (which is the squared t-statistic on μ) on the first stage (37) should be greater than a threshold around 16 to 19, and this advice is being progressively enhanced in current IV research (see Montiel Olea and Pflueger (2013) and Andrews et al. (2019)).

Proposition 3 and Corollary 2 show that we have a precise estimator if concentration is high and if the idiosyncratic shocks are volatile relative to the volatility of aggregate shocks. Hence, before conducting a large-scale study and data collection effort, researchers can perform an *ex ante power analysis*. Recall that in the most basic case, the standard error of the estimator is $s.e.\,(\psi_T^e) = \frac{\sigma_\varepsilon}{h\sigma_u |M| \sqrt{T}}$ (see (44)). To get a quick sense of the power of GIVs, we can have a rough estimate of σ_ε using the volatility of the left-hand side variable (here, p_t) as an order of magnitude, σ_u using $\sigma_{\check{y}_i}$ so that a simple common factor is removed as well as the component that depends

[18] This intuition extends to the model with a more complex factor structure: the expression $\sigma_z = h\sigma_u$ now entails a modified Herfindhal $h = \sqrt{S'QS}$, which is less transparent (see Proposition 3) but has broadly similar behavior. See the proof of Corollary 2 for an example.

on prices, and the average excess Herfindahl h. These inputs can be used to calibrate an order of magnitude for $\sigma_{\psi_T^e}$, given the researcher's prior view on M. This gives a sense of the t-statistics $t = \frac{\psi}{\sigma_{\psi_T^e}}$ that can be obtained with T periods of data. Hence, before refining the empirical model and perhaps collecting additional data, one has a sense of whether the GIV will be sufficiently powerful. The results of such a power analysis can be reported alongside the final estimates.

3.3 Robustness to misspecification and threats to identification

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

Forms of misspecification to which GIVs are robust First, it is possible to construct GIVs based on a subset of entities, I_t, that is, $z_t = \sum_{i \in I_t} S_i \check{u}_{it}$. This can be useful in practice as we can select the top K entities, the entities for which we have data, or to omit entities for which the data may contain measurement error. In this case, all results go through, although a rescaling may be required to ensure that Assumption 7 holds.[19] Hence, the estimator remains valid, although it is not the optimal GIV estimator.

Second, suppose that we misspecify the vector S of size weights, for example, by defining $z_t = \sum_i S_i^\circ \check{u}_{it}^e$ using a wrong vector S°. Then, the parameters estimated using IV are still consistently estimated, but the parameters estimated using OLS can be biased.[20] After all, (9) still holds, namely $\mathbb{E}\left[(p_t - \psi y_{St}) z_t\right] = 0$, so that the IV procedure in Proposition 1 remains valid.

Third, if we assume that the elasticities are homogeneous across entities (of demand in our model in (1)-(2)), while they are actually heterogeneous, then the IV and OLS estimates are still useful, as the parameters that we estimate are equal-weighted averages of coefficients. For instance, the IV estimators yield estimates of ψ and ϕ_E^d, and the OLS estimators estimate coefficients corresponding to the interpretation that the elasticity of demand is ϕ_E^d instead of ϕ_S^d. Section D.8 provides the derivations.

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

Conditional on GIVs being sufficiently powerful, the most important threat to identification is that we do not properly control for common factors. Indeed, calling λ_Γ^e and η_t^e the estimated values, $z_t = u_{\Gamma t} + \check{\lambda}_\Gamma \check{\eta}_t - \check{\lambda}_\Gamma^e \check{\eta}_t^e$, so there is a danger that, even after controlling for $\check{\eta}_t^e$ in the regression,

[19] For instance, we still have $u_{St} = z_t + \varepsilon_t^{u_S}$ with $z_t \perp \varepsilon_t^{u_S}$. Section D.10 gives for a formal analysis.

[20] Calling $\xi = \frac{\mathbb{E}[z_t u_{S^\circ t}]}{\mathbb{E}[z_t^2]}$ (which is 1 when $S^\circ = S$, assuming (35)), then the OLS above gives the estimates (in expectation) $\mu^e = \mu\xi$ and $M^e = M\xi$. For some selection procedures (e.g. selecting the shocks to some pre-specified entities as we discussed), we still have that $\xi = 1$, so that OLS is still valid. However, if $S^\circ = S + \epsilon^S$, where ϵ^S is a vector of measurement error, then typically $\xi \neq 1$ and the OLS procedure is biased.

we will not completely eliminate the error term $\check{\lambda}_\Gamma \check{\eta}_t - \check{\lambda}^e_\Gamma \check{\eta}^e_t$.[21] This bias is larger when $\left|\check{\lambda}_\Gamma\right|$ is greater, that is, when loadings are correlated with size. Indeed, omitting factors for which $\check{\lambda}_\Gamma = 0$ is inconsequential. We provide four concrete suggestions to mitigate this concern.

First, we can add factors for which we estimate the loadings, see Section 4.2. As we show, when T is sufficiently large, we accurately measure $\check{\lambda}$ and $\check{\eta}_t$ and this procedure detects (strong) missing factors. While one cannot rule out weak factors, it is reassuring if the GIV estimates remain stable when adding factors. A related concern is that the factor loadings are unstable over time. In this case, it is possible to use more advanced methods to extract factors, for instance by modeling $\lambda_{it} = X_{it}\dot{\lambda}$, as we discuss in Section D.12. Identification then depends on the correct specification of X_{it}.

Second, we can "narratively check" the idiosyncratic shocks that are being used in the GIV. Indeed, if granular shocks have aggregate consequences, it is often the case that the researcher can "label" them and understand them—especially if we have high-frequency data. This conforms to the spirit of the credibility revolution that the researcher is very clear about the underlying "assignment-to-treatment process"; here we can be clear about what the "treatments" are. The researcher can check the top, say, 10 events, and confirm using additional information that those shocks indeed are valid idiosyncratic shocks. It is then possible to construct the GIV based only on those narratively-checked shocks only.[22] This analysis also helps researchers to provide economic content to the variation used in estimating the parameters of interest, which lends further credibility to the estimates.

Third, the narrative check may reveal that a large idiosyncratic shock coincides with a large aggregate event. For instance, there may be an important policy announcement in the sample. If this happens once in the sample, it is hard to detect using standard factor models. Such "sporadic factor realizations" can lead to bias, and it is therefore prudent to remove these dates.

Fourth, we can do an overidentification test. We can construct two GIVs based on two types of entities, e.g. developing and developed countries. Then form the GIVs $z_t^{(1)} = S^{(1)\prime}\check{u}_t$ and $z_t^{(2)} = S^{(2)\prime}\check{u}_t$ based on the size-weighted sum of idiosyncratic shocks of each type (i.e., with two different sets of weights $S^{(1)}$ and $S^{(2)}$), and test whether the estimates using either instrument are the same. If the estimates are significantly different, then this points to misspecification, for instance, of the factor model or that the elasticities are heterogeneous across entities. In this case, we can explore generalization in terms of the factor model (see Section 4.2) or the model of elasticities (see Section 4.1).

[21] As we do control for $\check{\eta}^e_t$ in the regression, the bias is due to the residual of $\check{\lambda}_\Gamma \check{\eta}_t - \check{\lambda}^e_\Gamma \check{\eta}^e_t$ after controlling for $\check{\eta}^e_t$.

[22] This is roughly what the "narrative" approach in the literature does (e.g., Caldara et al., 2019). But the GIV procedure helps researchers even in the narrative context, since it automates the "pre-selection" of the top K (perhaps $K = 10$) shocks, by selecting the events with the largest K values in $S_i\,|\check{u}_{it}|$. Hence, researchers don't need to know the whole history before selecting their main events – the GIV gives them the most promising candidate events, and the detailed historical search is simply restricted to K events. In addition, the factor analysis in the GIV gives controls η^e_t that are usable when running regressions, which increases the precision of estimators.

What is an idiosyncratic shock? Mathematically, an idiosyncratic shock is plainly a random variable u_{it} such that $\mathbb{E}_{t-1}\left[\left(\eta_t', \varepsilon_t\right) u_{it}\right] = 0$. That said, it is useful to discuss different types of economic settings that map into this definition. In some cases it is quite clear – for example, a random productivity or demand shock. But there are more subtle types of idiosyncratic shocks. One is an "unexpected change in the loading on a common shock." For instance, if China decreases oil consumption by more than anticipated in response to a global economic downturn, then it is an idiosyncratic shock. Formally, if demand is $y_{it} = \phi^d p_t + \left(\lambda_i + \tilde{\lambda}_{it}\right)\eta_t + v_{it}$, with $\mathbb{E}_{t-1}\left[\left(1, \eta_t', \varepsilon_t\right)\tilde{\lambda}_{it}\right] = 0$, then $u_{it} := \tilde{\lambda}_{it}\eta_t + v_{it}$ is a valid idiosyncratic shock.

The volatility of idiosyncratic shocks can depend on the common shocks. For instance, suppose that $u_{it} = \sigma_t v_{it}$ where σ_t and $\left(\eta_t', \varepsilon_t\right)$ could be correlated (for instance, σ_t could increase when $|\varepsilon_t|$ is high), but $\mathbb{E}_{t-1}\left[\left(\eta_t', \varepsilon_t\right)\sigma_t v_{it}\right] = 0$ (a sufficient condition is that v_{it} independent of $\sigma_t\left(\eta_t', \varepsilon_t\right)$); then, u_{it} is an idiosyncratic shock because $\mathbb{E}_{t-1}\left[\left(\eta_t', \varepsilon_t\right) u_{it}\right] = 0$.[23]

4 Extensions

We now present a succession of extensions of the basic GIV procedure that cover a range of empirically relevant cases. The Online Appendix gives a number of other extensions.

4.1 Heterogeneous demand elasticities

We have assumed so far that demand elasticities are constant across entities. We now extend the model to the case where demand elasticities vary across entities and are a function of characteristics x_i.[24] x_i is a k-dimensional vector and the first entry is equal to 1.[25] We represent the demand elasticities as:

$$\phi_i^d = x_i \dot{\phi}^d = \sum_{\ell=1}^{k} x_{i\ell}\dot{\phi}_\ell^d, \tag{47}$$

for some k-dimensional vector $\dot{\phi}^d = \left(\dot{\phi}_\ell^d\right)_{\ell=1\ldots k}$ that is to be estimated. With x the $N \times k$ matrix of characteristics, we summarize the elasticities as $\phi^d = x\dot{\phi}^d$. We assume that λ spans x. For instance, we can have $\lambda = \left(x, \hat{\lambda}\right)$, where $\hat{\lambda}$ comprises loadings orthogonal to x. The following proposition describes how we can consistently estimate ψ and $\dot{\phi}^d$.

Proposition 6 (Estimation of heterogeneous parametric elasticities). *Let Assumptions 1 and 3 hold. Consider the model with heterogeneous elasticities of demand following* (47)*. Define* $R^x := \left(x'x\right)^{-1} x'$*,* $\dot{y}_t := R^x\left(y_t - C_t^y m^y\right)$ *(which has dimension k),* $\check{u}_t := Q^\lambda\left(y_t - C_t m^y\right)$*, and* $z_t := \Gamma'\check{u}_t$*,*

[23] A related concern is that some (small) clusters of the u_{it}'s are correlated. In that case, we can either aggregate the entities within each cluster or, alternatively, we estimate V^u from Assumption 4 to be block diagonal.

[24] This x_i might be different from X_i in Assumption 1, e.g. be a strict subset of it.

[25] Section D.9 considers a non-parametric version, which is more involved.

where Γ *satisfies (4). We then identify* ψ *and* $\dot{\phi}^d$ *using the moment conditions:* $\mathbb{E}\left[(p_t - \psi y_{St})\, z_t\right] = 0$ *and* $\mathbb{E}\left[\left(\dot{y}_t - \dot{\phi}^d p_t\right) z_t\right] = 0_{K\times 1}$*, i.e.*

$$\mathbb{E}\left[\left(\dot{y}_{\ell t} - \dot{\phi}^d_\ell p_t\right) z_t\right] = 0 \qquad \textit{for } \ell = 1 \dots k \tag{48}$$

This procedure extends the model with homogeneous demand elasticities, where the moment condition (48) simplifies to $\mathbb{E}\left[\left(y_{Et} - \phi^d p_t\right) z_t\right] = 0$.[26]

Computing $\dot{y}_t$ and $\check{u}_t$ in Proposition 6 corresponds to running cross-sectional regressions of $y_{it} - C^y_{it} m^y$ on x_i and $\check{\lambda}_i$, for every date t (and given m^y):

$$y_{it} - C^y_{it} m^y = x_i \dot{y}_t + \check{\lambda}_i \check{\eta}_t + \check{u}_{it}, \tag{49}$$

which yields an exact recovery (as in Lemma 1) of $\dot{y}_t$, $\check{\eta}_t$, and $\check{u}_{it}$. Then, we form the GIV $z_t := \sum_i S_i \check{u}_{it}$ and use the moment conditions in Proposition 6 to recover the elasticities. One can add controls, as in Proposition 4, and the procedure is otherwise the same.

4.2 When the factor loadings are estimated

The results so far are derived by imposing Assumption 1 that we know λ. We now show how the procedure extends when we relax this assumption and estimate the factor loadings. We remain in the case where $T \to \infty$ and N is fixed.[27]

Recall that we have $\lambda = \left(\iota, \check{\lambda}\right)$. As in all factor models, we need a normalization. We choose the normalization that $\frac{1}{N}\lambda'\lambda = I_r$ and that $V^{\check{\eta}}$ is diagonal.[28] To ensure uniqueness of the representation, we also impose that the diagonal terms of $V^{\check{\eta}}$ are distinct and in decreasing order; this is a purely technical condition without economic substance, that could be relaxed at the cost of distracting notations. This leads to the following system of moment conditions.

Proposition 7 (GIV estimation with controls and estimation of factor loadings λ) *Let Assumption 3 hold. Define* $\check{y}_{it} := y_{it} - y_{Et}$ *and similarly* $\check{C}^y_{it} := C^y_{it} - C^y_{Et}$*. Given candidate values* m^y *and* $\lambda = \left(\iota, \check{\lambda}\right)$*, we construct estimates of the factors using* $\check{\eta}_t\left(m^y, \check{\lambda}\right) := R^{\check{\lambda}}\left(y_t - C^y_t m^y\right)$ *and the associated GIV* $z_t\left(m^y, \check{\lambda}\right) := S'Q^\lambda\left(y_t - C^y_t m^y\right)$*. Define* θ *to be* $\left(\psi, \phi^d, m^p, b^p, m^y, b^y, \check{\lambda}\right)$ *and* θ^e_T *to be the GMM estimator of* θ *associated with the following moment conditions:*

$$\sum_t \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y, \check{\lambda}\right) + \check{C}^y_t m^y\right) \check{\eta}_t\left(m^y, \check{\lambda}\right)' = 0 \tag{50}$$

$$\sum_t \left(-\check{y}_t + \check{C}^y_t m^y\right)' \check{C}^y_t = 0, \tag{51}$$

[26]Indeed, when the elasticity is the same across entities, $x = \iota$ so that $R^x = E'$ and $\dot{y}_t = y_{Et}$.

[27]Banafti and Lee (2022) extend the results in this paper by studying the the case where $T, N \to \infty$.

[28]To remove the classic indeterminacy that the f-th factor λ^f could be changed into $-\lambda^f$, we can impose e.g. that for each f, the first non-zero λ^f_i is positive.

and moments (29)-(30), making $\check{\eta}_t$ and z_t depend on $\left(m^y, \check{\lambda}\right)$. Then, assuming $\psi \neq 0$, and the regularity Assumption 5 from Appendix A, the vector θ_0 is consistently estimated (with $T \to \infty$ and fixed N), and we have $\sqrt{T}\left(\theta_T^e - \theta_0\right) \to N\left(0, V^\theta\right)$ for a matrix V^θ.

The new moment (50) identifies $\check{\lambda}$.[29,30] The advantage of the formulation in Proposition 7 is that the variance V^θ is derived using standard GMM theory, and implemented numerically using standard GMM routines. An analogous proposition can be stated for the OLS procedure, as in Proposition 5, but adding the moment (50) to estimate $\check{\lambda}$. Then, unlike the positive message of Proposition 5, when λ is estimated, the fact that z_t is a constructed regressor does affect the asymptotic variance of the estimator and we cannot rely on the OLS standard error. Instead, one must rely on the variance matrix V^θ, which is still easy to compute numerically.

4.3 Heteroskedasticity

We now discuss the case where the u_{it} are heteroskedastic. We call V^u their variance-covariance matrix. We assume for now that this is known, at least up to a factor of proportionality, and constant over time. Given a positive definite matrix W, define the $r \times N$ and $N \times N$ matrices,

$$R^{\lambda,W} := \left(\lambda' W \lambda\right)^{-1} \lambda' W, \qquad Q^{\lambda,W} := I - \lambda R^{\lambda,W}, \tag{52}$$

so that $Q^{\lambda,W}$ is a projection on the space orthogonal to λ and $R^{\lambda,W}$ is a projection on λ.[31]

Proposition 8 (GIV with heteroskedastic idiosyncratic shocks) *Let us replace Assumption 3 by Assumption 4, assuming that we know V^u, and define the corresponding matrices $Q^{\lambda,W}$ and $R^{\lambda,W}$ in (52), with $W = \left(V^u\right)^{-1}$. Then Propositions 1-7 and Lemma 1 hold, provided that we change $\Gamma' S \neq 0$ into $\Gamma' V^u S \neq 0$ in (4); change $Q^{\lambda,W}$ and R^{λ} into $Q^{\lambda,W}$ and $R^{\lambda,W}$ everywhere, for all occurrences of factors λ (e.g., in* (17) *we change $\Gamma^{*\prime} = S'Q$ into $\Gamma^{*\prime} = S'Q^{\lambda,W}$; in (20) we change R^λ, Q^λ into $R^{\lambda,W}$, $Q^{\lambda,W}$; in Proposition 6, we change R^x into $R^{x,W}$); change equal weights E into quasi-equal weights $\tilde{E} := \frac{W\iota}{\iota' W \iota}$ (e.g. in* (43)*); change (50) into $\sum_t \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y, \check{\lambda}\right) + \check{C}_t^y m^y\right) W \check{\eta}_t\left(m^y, \check{\lambda}\right)' = 0$.*

When idiosyncratic shocks are homoskedastic, $\tilde{E}_i = \frac{1}{N}$, while if they are uncorrelated but het-

[29] Moment (50) means that the $r-1$ column vectors $\check{\lambda}$ are eigenvectors of the empirical variance-covariance matrix of $\check{y}_t - \check{C}_t^y m^y$, so that a PCA can be equivalently performed. See the proof of Proposition 7.

[30] Instead of (51), one could use the moment $\sum_t \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y, \check{\lambda}\right) + \check{C}_t^y m^y\right)' \check{C}_t^y = 0$, but (51) simplifies the proof.

[31] These matrices have convenient properties that we record here (dropping the superscripts for simplicity):

$$Q\lambda = 0, \quad R\lambda = I, \quad Q'W\lambda = 0, \quad (I-Q)\, W^{-1} Q' = 0, \quad (I - Q')\, WQ = 0, \quad Q^2 = Q, \quad RW^{-1}Q' = 0. \tag{53}$$

eroskedastic (i.e. $W = Diag\left(1/\sigma_{u_i}^2\right)$), we have[32]

$$\tilde{E}_i := \frac{1/\sigma_{u_i}^2}{\sum_j 1/\sigma_{u_j}^2}. \tag{54}$$

In addition, in the cross-sectional regressions, we use GLS with weights $W = (V^u)^{-1}$ instead of OLS (e.g. in (32)).

While estimating ψ could be done without knowing the $\sigma_{u_i}^2$, estimating ϕ^d, M, and μ does require the conditions (10) or (35), hence some knowledge of the heteroskedasticity of the u_i's.[33,34]

Estimation of the degree of heteroskedasticity In the central case where the u_{it} are uncorrelated across i's, we can estimate the volatilities σ_{u_i}'s consistently, via the moments $\mathbb{E}\left[\check{u}_{it}^2\right] = Q_{ii}^{\lambda,W}\sigma_{u_i}^2$, where $\check{u}_t = Q^{\lambda,W}u_t$ is the vector of residuals; this fits in the GMM structure of the rest of the estimation.[35] This is detailed in Section D.3 of the Online Appendix. When the σ_{u_i} are estimated, the standard errors of the estimators of ψ, ϕ^d, μ, and M adjust accordingly.

4.4 Generalization of the GIV to other setups

While we focus on the demand and supply setup for our main analysis, we discuss in this section how the basic ideas extend to a variety of settings.

Time-varying size weights Size weights could vary over time, $S_{i,t}$, so that in our leading example (2) the aggregate demand disturbance y_{St} becomes $y_{St} = \sum_i S_{i,t-1}y_{it}$. We then make the additional assumption that u_t is independent of S_{t-1}. Then, our identifying moments are still correct, replacing S_i by $S_{i,t-1}$ and Γ by Γ_t, with the natural changes (e.g. in (4) we replace S and Γ by S_t and Γ_t). [36]

Time-varying factors We can relax Assumption 1 and replace $\lambda_i = X_i\dot{\lambda}$ by $\lambda_{it} = X_{it}\dot{\lambda}$. This would allow for time variation in the loadings on the aggregate shocks. In general, we can then have factor loadings that are observed and time varying and factor loadings that are unknown (see

[32] Generically, $\Gamma^* \neq 0$. But in cases where the variance is inversely proportional to size, $V^u S = a\iota$ for some scalar a, so that $\Gamma^{*\prime} = S'Q^{\lambda,W} = 0$ and the GIV would fail. This would be detected in practice via extremely large standard errors. Fortunately, in most contexts, variance may decay a bit with size S_i, but not as fast as $1/S_i$ (see e.g. Lee et al. (1998) and the discussion in Gabaix (2011)).

[33] Recall that those conditions are identical when $\tilde{E} := S - \Gamma$.

[34] If we misspecify the variance of the u_{it} (keeping them uncorrelated), the impact on the estimates is typically quite small. For instance, if the $\sigma_{u_i}^2$ are heteroskedastic and we ignore this fact, then $\mathbb{E}\left[u_{\Gamma t}u_{Et}\right] = \frac{1}{N}\sum_i \Gamma_i\sigma_{u_i}^2$ will be non-zero, but will still be small when we have a large cross-section, of order $O\left(\frac{1}{N}\right)$. The relative bias $\frac{\mathbb{E}[u_{\Gamma t}u_{Et}]}{\mathbb{E}\left[u_{\Gamma t}^2\right]} = O\left(\frac{1}{Nh_N^2}\right)$ more generally goes to 0 when we are in a "granular" case where $h_N^2 \gg \frac{1}{N}$, where h_N is the Herfindahl (45)—see Gabaix (2011), Proposition 2 for details.

[35] The moment does features $Q_{ii}^{\lambda,W}\sigma_{u_i}^2$, not $\left(Q_{ii}^{\lambda,W}\right)^2\sigma_{u_i}^2$ as one might think at first.

[36] The expressions of the variances will change, for instance replacing σ_z^2 by its average value, which should remain bounded away from 0.

Section 4.2). The combined loadings are then given by $\lambda_{it} = \left(X_{it}\dot{\lambda}, \lambda_i^F\right)$, where $X_{it}\dot{\lambda}$ fluctuates over time and λ_i^F is constant.

Multidimensional GIV The basic model can be extended to cover multidimensional outcomes y_{it} and shocks u_{it}. For instance, a firm i could have two-dimensional shocks u_{it}, one to productivity and one to labor demand. We show how GIVs can be used in this setting in Section D.6.

Beyond supply and demand for one good In Section D.5, we show how GIVs can be used to estimate parameters of interest in models that feature multiple general equilibrium channels.

Estimating structural vector autoregressions with GIVs We can estimate vector autoregressions and impulse responses with GIVs. If $Y_t = AY_{t-1} + X_t$ for vectors X_t, Y_t and matrix A, we 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 we can follow the methods spelled out in Stock and Watson (2018).[37] We can also estimate Jordà (2005)-style local projections, regressing $\mathbb{E}_t\left[Y_{t+h}\right] = \beta_h X_t$, and instrumenting some of the regressors X_t by GIVs. This shows how GIVs can be used to identify parameters in structural VARs, complementing an active literature that uses sign restrictions, as in Uhlig (2005), or narrative restrictions combined with sign restrictions, as in Antolín-Díaz and Rubio-Ramírez (2018) and Ludvigson et al. (2020).

Comparison with Bartik instruments Bartik (1991) instruments are widely used in economics and we discuss the link between Bartik and GIV in Online Appendix D.4, and clarify the difference in identifying assumptions using the econometric framework of Borusyak et al. (2022).[38] We briefly summarize the main insights. First, in a number of cases where a cross-section is studied, Bartik applies but GIV cannot be used as there is no large idiosyncratic shock that one can use (e.g., Autor et al., 2013). On the other hand, in a number of cases GIV applies naturally, particularly when there are large idiosyncratic shocks that affect aggregate outcomes. We therefore view Bartik and GIVs as complements in the toolkit of economists, and it depends on the setting which empirical strategy is most appropriate.

When aggregate shocks are at least partially made of idiosyncratic shocks GIVs extend to economies where aggregate shocks η_t are themselves (at least partially) made of idiosyncratic shocks u_{it} (as in Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019); Brownlees and Mesters (2021)). We develop this in Sections D.17-D.18. These sections show that we can identify important parameters even if we have only crude proxies for the primitive shocks such as TFP.

[37] See Plagborg-Møller and Wolf (2022) and Sarto (2022) for recent developments in this area.

[38] Bartik instruments (also known as shift-share estimators) have been rigorously studied in recent econometric work, see also Adao et al. (2019); Goldsmith-Pinkham et al. (2020); Borusyak et al. (2022).

5 Conclusion

We developed granular instrumental variables (GIVs). The generative insight is that idiosyncratic shocks offer a rich source of instruments in concentrated economic environments with large idiosyncratic shocks. We lay out econometric procedures to extract them from panel data and efficiently aggregate them to obtain the most powerful instruments. We discuss various econometric extensions that might be useful in applied work. Given the ubiquity of concentration in various economic settings and the volatility of idiosyncratic shocks, we hope that GIVs provide a valuable addition to the toolkit of economists to estimate 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.

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

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.

Andrews, Isaiah, James H Stock, Liyang Sun et al., "Weak instruments in instrumental variables regression: Theory and practice," *Annual Review of Economics*, 2019, *11* (1), 727–753.

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

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

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

Banafti, Saman and Tae-Hwy Lee, "Inferential theory for granular instrumental variables in high dimensions," *arXiv preprint arXiv:2201.06605*, 2022.

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

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

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

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

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

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

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

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

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

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

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

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

Dong, Xi, Namho Kang, and Joel Peress, "Fast and slow arbitrage: The predictive power of capital flows for factor returns," *Available at SSRN 3675163*, 2022.

Flynn, Joel P and Karthik Sastry, "The Macroeconomics of narratives," *Available at SSRN 4140751*, 2022.

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

Gabaix, Xavier and Ralph SJ Koijen, "In search of the origins of financial fluctuations: The inelastic markets hypothesis," *NBER Working Paper No. 28967*, 2022.

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

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

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

Hansen, Lars, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–54.

Huber, Kilian, "Estimating general equilibrium spillovers of large-scale shocks," *The Review of Financial Studies*, 08 2023, *36* (4), 1548–1584.

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

Kundu, Shohini and Nishant Vats, "Banking networks and economic growth: From idiosyncratic shocks to aggregate fluctuations," *Available at SSRN 3556299*, 2021.

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

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

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," *NBER Working Paper No. 23225*, 2020.

Ma, Yueran, Teodora Paligorova, and José-Luis Peydro, "Expectations and bank lending," *Working Paper*, 2022.

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 Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 1994, *4*, 2111–2245.

Plagborg-Møller, Mikkel and Christian K Wolf, "Instrumental variable identification of dynamic variance decompositions," *Journal of Political Economy*, 2022, *130* (8), 2164–2202.

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

Sarto, Andres, "Recovering macro elasticities from regional data," *Working Paper*, 2022.

Schubert, Gregor, Anna Stansbury, and Bledi Taska, "Employer concentration and outside options," *Available at SSRN 3599454*, 2022.

Stock, James H and Mark W Watson, "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," *Handbook of Macroeconomics*, 2016, *2*, 415–525.

Stock, James H and Mark W Watson, "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *The Economic Journal*, 2018, *128* (610), 917–948.

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

A Appendix: Proofs

This section shows some of the main proofs. Further proofs are in the Online Appendix, Section C.

A.1 Variance facts

We will repeatedly use a number of facts that we record here. For two deterministic vectors X and Y of dimensions $n \times 1$, defining $u_X := X'u$ and $u_Y := Y'u$, we have

$$\mathbb{E}\left[u_X u_Y\right] = \mathbb{E}\left[\left(X'u\right)\left(Y'u\right)\right] = X'\mathbb{E}\left[uu'\right]Y = X'V^u Y. \tag{55}$$

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

$$\mathbb{E}\left[u_X u_Y\right] = X'Y\sigma_u^2. \tag{56}$$

With $\Gamma = S - E$ (where $E_i = \frac{1}{N}$), we have with $h = \sqrt{\sum_{i=1}^N S_i^2 - \frac{1}{N}}$:[39]

$$\mathbb{E}\left[u_\Gamma u_E\right] = 0, \qquad \mathbb{E}\left[u_\Gamma^2\right] = \mathbb{E}\left[u_S u_\Gamma\right] = h^2\sigma_u^2. \tag{57}$$

In the general heteroskedastic case for V^u, defining $W = \left(V^u\right)^{-1}$, the quasi-equal weight vector is $\tilde{E} = \frac{W\iota}{\iota' W\iota}$. Then, for any Γ such that $\iota'\Gamma = 0$, we have:[40]

$$\mathbb{E}\left[u_\Gamma u_{\tilde{E}}\right] = 0. \tag{58}$$

A.2 Proof of Propositions 1 and 2

First, we note that (2) implies

$$y_{St} = \phi^d p_t + u_{St} + \varepsilon_{1t},$$

where ε_{1t} (and later ε_{2t}, ε_{3t}, etc.) is a linear function of $\varepsilon_t, \eta_t, C_t^y$, and C_t^p, and thus uncorrelated with u_t by (3). Using (1) gives

$$y_{St} = \psi\phi^d y_{St} + u_{St} + \varepsilon_{2t}.$$

Hence, using the notation introduced in (13), we have

$$y_{St} = Mu_{St} + \varepsilon_{3t}, \qquad p_t = \mu u_{St} + \varepsilon_{4t}, \qquad y_{\tilde{E}t} = \left(M-1\right)u_{St} + u_{\tilde{E}t} + \varepsilon_{5t}. \tag{59}$$

This gives how idiosyncratic shocks affect y_{St}, p_t, and y_{Et}.

The rest of the proof uses well-known ingredients. We use $\mathbb{E}_T$ for the sample temporal mean,

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

[40]Indeed, as $\tilde{E} = kW\iota$, and $k = \frac{1}{\iota W\iota}$, we have $\mathbb{E}\left[u_\Gamma u_{\tilde{E}}\right] = \tilde{E}'\mathbb{E}\left[uu'\right]\Gamma = \tilde{E}'V^u\Gamma = k\iota'WV^u\Gamma = k\iota'\Gamma = 0$.

$\mathbb{E}_T[Y_t] := \frac{1}{T}\sum_{t=1}^{T} Y_t$. We have $p_t - \psi y_{St} = \varepsilon_t^{\psi} := C_t^p m^p + \varepsilon_t$, so

$$\psi_T^e - \psi = \frac{\mathbb{E}_T[p_t z_t]}{\mathbb{E}_T[y_{St} z_t]} - \psi = \frac{\mathbb{E}_T[(p_t - \psi y_{St}) z_t]}{\mathbb{E}_T[y_{St} z_t]} = \frac{\mathbb{E}_T\left[\varepsilon_t^{\psi} u_{\Gamma t}\right]}{\mathbb{E}_T[y_{St} u_{\Gamma t}]} = \frac{A_T}{D_T}, \tag{60}$$

where $A_T = \mathbb{E}_T\left[\varepsilon_t^{\psi} u_{\Gamma t}\right]$ and $D_T = \mathbb{E}_T[y_{St} u_{\Gamma t}]$. For the denominator, the law of large number gives, as $T \to \infty$,

$$D_T = \mathbb{E}_T[y_{St} u_{\Gamma t}] \xrightarrow{a.s.} D = \mathbb{E}[y_{St} u_{\Gamma t}].$$

Using (59), we have

$$D = \mathbb{E}[y_{St} u_{\Gamma t}] = \mathbb{E}[(M u_{St} + \varepsilon_{3t}) u_{\Gamma t}] = M\mathbb{E}[u_{St} u_{\Gamma t}]$$

which is non-zero because we assumed that $\mathbb{E}[u_{St} u_{\Gamma t}] \neq 0$, and because of the definition of M in (13). For the numerator, the central limit theorem gives the convergence in distribution $\sqrt{T} A_T \xrightarrow{d} \mathcal{N}(0, \sigma_A^2)$,where, using the assumption that ε_t^{ψ} is homoskedastic conditionally on u_t,

$$\sigma_A^2 = \mathbb{E}\left[\left(\varepsilon_t^{\psi}\right)^2 u_{\Gamma t}^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\varepsilon_t^{\psi}\right)^2 \mid u_t\right] u_{\Gamma t}^2\right] = \mathbb{E}\left[\sigma_{\varepsilon^{\psi}}^2 u_{\Gamma t}^2\right] = \sigma_{\varepsilon^{\psi}}^2 \sigma_{u_{\Gamma}}^2.$$

This implies (by Slutsky's theorem) that $\sqrt{T}(\psi_T^e - \psi) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\psi}^2\right)$, where $\sigma_{\psi} = \frac{\sigma_A}{|D|} = \frac{\sigma_{\varepsilon^{\psi}} \sigma_{u_{\Gamma}}}{|M\mathbb{E}[u_{St} u_{\Gamma t}]|}$. The derivation of $\sigma_{\phi^d}^2$ follows along the same lines. It is detailed in Section C of the Online Appendix.

A.3 Proof of Proposition 3

We solve for the optimal Γ, which minimizes $\sigma_{\psi}(\Gamma)$ (and then, automatically, $\sigma_{\phi^d}(\Gamma)$) subject to $\Gamma'\lambda = 0$. Given Proposition 2, we want to maximize the squared correlation:

$$\max_{\Gamma} C(\Gamma) := corr(u_{St}, u_{\Gamma t})^2 \text{ subject to } \Gamma'\lambda = 0. \tag{61}$$

We next solve this problem. By Assumption 3 and (56) we have $\mathbb{E}[u_{St} u_{\Gamma t}] = \sigma_u^2 \Gamma' S$, hence:

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

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

$$\mathcal{L} = \frac{1}{2}\Gamma'\Gamma - \Gamma'\lambda b - c(\Gamma' S - v), \tag{62}$$

with some Lagrange multipliers b (of dimension $r \times 1$), and c (a scalar). The first order condition in Γ' is:

$$0 = \Gamma - \lambda b - cS \qquad \Longleftrightarrow \qquad \Gamma = cS + \lambda b. \tag{63}$$

We next use the projection operator $Q = I - \lambda (\lambda'\lambda)^{-1} \lambda'$ (see (16)) which satisfies $Q\lambda = 0$. As $\lambda'\Gamma = 0$, we have $Q\Gamma = \Gamma$. So

$$\Gamma = Q\Gamma = cQS + Q\lambda b = cQS. \tag{64}$$

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

A.4 Proof of Lemma 1

We prove the lemma in the more general heteroskedastic case, with $W = (V^u)^{-1}$. We define, as in (52), $R := (\lambda' W \lambda)^{-1} \lambda' W$ and $Q := I - \lambda R$, dropping the superscripts for concision. We have $Q\lambda = 0$ and $\lambda \check{\eta}_t + \check{u}_t = \lambda (\eta_t + R u_t) + (I - \lambda R) u_t = \lambda \eta_t + u_t = Y_t$. Finally,

$$\begin{aligned} \mathbb{E}[\check{\eta}_t \check{u}_t'] &= \mathbb{E}\left[(\eta_t + R u_t)(Q u_t)'\right] = \mathbb{E}[R u_t u_t' Q'] = R\mathbb{E}[u_t u_t'] Q' = R V^u Q' \\ &= (\lambda' W \lambda)^{-1} \lambda' W V^u Q' = (\lambda' W \lambda)^{-1} \lambda' Q' = (\lambda' W \lambda)^{-1} (Q\lambda)' = 0. \end{aligned}$$

A.5 Proof of Proposition 4

We will use the following regularity assumption.

Assumption 5 *The matrices $\mathbb{E}[\check{\eta}_t \check{\eta}_t']$, $\mathbb{E}\left[(Q^\lambda C_t^y)'(Q^\lambda C_t^y)\right]$, $\mathbb{E}\left[C_{St}^{y\prime} C_{St}'\right]$ and $\mathbb{E}\left[C_t^{p\prime} C_t^p\right]$ have full rank. The true parameter θ_0 is in the interior of a compact set Θ.*

These conditions are mild and technical.[41] We also recall that we maintain the assumption that at the true values, $\psi\phi^d \neq 1$.

We define the function $g_t(\theta) = \left(g_t^k(\theta)\right)_{k=1,2,3}$ as

$$\begin{aligned} g_t^1(\theta) &:= \left(-\check{y}_t + \check{\lambda}\check{\eta}_t(m^y) + \check{C}_t^y m^y\right)' \check{C}_t^y, && (65) \\ g_t^2(\theta) &:= \left(-p_t + \psi y_{St} + b^p \check{\eta}_t(m^y) + m^{p\prime} C_t^{p\prime}\right)\left(z_t(m^y), \check{\eta}_t(m^y)', C_t^p\right) && (66) \\ g_t^3(\theta) &:= \left(-y_{\tilde{E}t} + \phi^d p_t + b^y \check{\eta}_t(m^y) + C_{\tilde{E}t}^y m^y\right)\left(z_t(m^y), \check{\eta}_t(m^y)'\right) && (67) \end{aligned}$$

We again use $\mathbb{E}_T$ for the sample temporal mean, $\mathbb{E}_T[Y_t] := \frac{1}{T}\sum_{t=1}^T Y_t$. So that, given a sample of size T, our estimator θ_T^e solves $\mathbb{E}_T[g_t(\theta_T^e)] = 0$. Amongst other things, the proof will verify that, for large T, the solution exists and is unique.

[41] For instance, if $\mathbb{E}[\check{\eta}_t \check{\eta}_t']$ doe not have full rank $r - 1$, one can just use a lower-dimensional factor model, and achieve full rank.

At the true value, $\mathbb{E}\left[g_t\left(\theta_0\right)\right]=0$. Indeed, $g_t^1\left(\theta_0\right)=-\check{u}_t'\check{C}_t^y$, as spelled out in the Online Appendix, equation (73). This implies that $\mathbb{E}\left[g_t^1\left(\theta_0\right)\right]=0$, by (3), and the fact that $\check{u}_t=Q^\lambda u_t$. Likewise, $g_t^2\left(\theta_0\right)=-\varepsilon_t^\perp\left(z_t,\check{\eta}_t',C_t^{p\prime}\right)$, so $\mathbb{E}\left[g_t^2\left(\theta_0\right)\right]=0$, by the maintained assumptions of Section 2.1, which imply $\varepsilon_t,\eta_t\perp C_t^{p\prime},u_t,\check{u}_t,z_t$ as $z_t=\Gamma' u_t$; Lemma 1, which implies $\check{\eta}_t:=\eta_t+R^\lambda u_t\perp C_t^{p\prime},\check{u}_t,z_t$ as additionally $z_t=\Gamma'\check{u}_t$; the fact that $\varepsilon_t^\perp=\varepsilon_t-b^\varepsilon\check{\eta}_t\perp\check{\eta}_t$ which implies $\varepsilon_t^\perp\perp C_t^{p\prime},z_t,\check{\eta}_t$. And similarly $\mathbb{E}\left[g_t^3\left(\theta_0\right)\right]=0$.

The solution of $\mathbb{E}\left[g_t\left(\theta\right)\right]=0$ is unique. The high-level idea is that the equation $\mathbb{E}\left[g_t^1\left(\theta\right)\right]=0$ is linear equation in m^y, and identifies m^y given our full-rank Assumption 5. Then, given m^y, the other equations ($\mathbb{E}\left[g_t^k\left(\theta\right)\right]=0$ for $k=2,3$) are linear in all the other parameters in θ, so that unicity of the solution is guaranteed as those linear equations have full rank. We spell out the details in the Online Appendix, Section C.

The estimator is consistent. Our estimator θ_T^e is the solution $\mathbb{E}_T\left[g_t\left(\theta\right)\right]=0$.[42] Note that θ_T^e is equivalent to the GMM estimator that minimizes $\mathbb{E}_T\left[g_t\left(\theta\right)\right]'\mathbb{E}_T\left[g_t\left(\theta\right)\right]$, as indeed is achieves $\mathbb{E}_T\left[g_t\left(\theta\right)\right]=0$. We thus proceed to apply general GMM results (Hansen (1982); Newey and McFadden (1994)). In particular under the regularity conditions and i.i.d. sampling, by Newey and McFadden (1994), Theorem 2.6, the estimator is consistent.

The estimator is $\sqrt{T}$-normally distributed. Under the regularity conditions we assumed (and in particular that $\psi\phi^d\neq 1$, which ensures bounded gradients), by Newey and McFadden (1994), Theorem 3.4, $\sqrt{T}\left(\theta_T^e-\theta\right)$ converges in distribution to a normal distribution with mean 0 and positive variance covariance matrix V^θ.

Derivation of the variance of ψ_T^e. It turns out that it is enough to zoom in on first component of (66), i.e. the moment $\mathbb{E}_T\left[g_t^{2,z}\left(\theta\right)\right]=0$, where

$$g_t^{2,z}=\left(-p_t+\psi y_{St}+b^p\check{\eta}_t\left(m^y\right)+C_t^p m^p\right)z_t\left(m^y\right). \tag{68}$$

As part of our application of Newey and McFadden (1994)'s Theorem 6.1, we now show that $\mathbb{E}\left[g_{\theta_k}^{2,z}\right]$ has only one non-zero component k, the one corresponding to $\theta_k=\psi$.

We take all derivatives at the true value, θ_0. For notational simplicity, we drop the t when the meaning is clear, e.g. we write $g_\psi^{2,z}$ for $\frac{d}{d\psi}g_t^{2,z}$. As $g_\psi^{2,z}=y_{St}z_t$, it holds

$$\mathbb{E}\left[g_\psi^{2,z}\right]=\mathbb{E}\left[y_{St}z_t\right]=M\mathbb{E}\left[u_{St}z_t\right]. \tag{69}$$

[42]Moreover, the solution exists and is unique, by the same reasoning as for $\mathbb{E}\left[g_t\left(\theta\right)\right]=0$, replacing $\mathbb{E}$ by $\mathbb{E}_T$ when needed.

Next, $\check{\eta}_t(m^y) := R^{\check{\lambda}}(y_t - C_t^y m^y)$ and $z_t(m^y) := \Gamma'(y_t - C_t^y m^y)$ yield

$$\frac{d}{dm^y}\check{\eta}_t(m^y) = -R^{\check{\lambda}} C_t^y, \qquad \frac{d}{dm^y} z_t(m^y) = -\Gamma' C_t^y, \tag{70}$$

and thus

$$\mathbb{E}\left[\left(\frac{d}{dm^y}\check{\eta}_t(m^y)\right) z_t\right] = 0.$$

This allows us to verify that all the other derivatives of $\mathbb{E}\left[g_{\theta_k}^{2,z}\right]$ for the other components k of θ (except ψ) are 0. For instance, $\frac{d}{db^p} g_t^{2,z} = \check{\eta}_t z_t$, so $\mathbb{E}\left[g_{b^p}^{2,z}\right] = \mathbb{E}\left[\check{\eta}_t z_t\right] = 0$. Likewise, $\mathbb{E}\left[g_{\phi^d}^{2,z}\right] = 0$ and $\mathbb{E}\left[g_{b^y}^{2,z}\right] = 0$. More involved is:

$$\begin{aligned}
\mathbb{E}\left[g_{m_y}^{2,z}\right] &= \mathbb{E}\left[\left(\frac{d}{dm_y}\left(-p_t + \psi y_{St} + b^p \check{\eta}_t(m^y) + C_t^p m^p\right)\right) z_t + \left(-p_t + \psi y_{St} + b^p \check{\eta}_t(m^y) + C_t^p m^p\right)\left(\frac{d}{dm^y} z_t\right)\right] \\
&= b^p \mathbb{E}\left[\left(\frac{d}{dm^y}\check{\eta}_t(m^y)\right) z_t\right] + \mathbb{E}\left[(-\varepsilon_t)\left(-\Gamma' C_t^y\right)\right] = 0 + 0 = 0.
\end{aligned}$$

Hence, we just showed that $\mathbb{E}\left[g_\theta^{2,z}\right]$ only has one non-zero component, the $\mathbb{E}\left[g_{\theta_k}^{2,z}\right]$ corresponding to $\theta_k = \psi$. Then, applying Newey and McFadden (1994)'s Theorem 6.1 allows us to conclude that:[43]

$$\sigma_\psi^2 = \frac{var\left(g_t^{2,z}\right)}{\left(\mathbb{E}\left[g_\psi^{2,z}\right]\right)^2} \tag{71}$$

Given $g_t^{2,z} = -\varepsilon_t^\perp z_t$, and using the assumption that $\varepsilon_t^\perp$ is homoskedastic conditional on u_t,

$$var\left(g_t^{2,z}\right) = \mathbb{E}\left[\left(\varepsilon_t^\perp z_t\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\varepsilon_t^\perp\right)^2 | u_t\right] z_t^2\right] = \mathbb{E}\left[\left(\varepsilon_t^\perp\right)^2\right]\mathbb{E}\left[z_t^2\right],$$

so that, using (69),

$$\sigma_\psi^2 = \frac{var\left(\varepsilon_t^\perp\right)\mathbb{E}\left[z_t^2\right]}{\left(M\mathbb{E}\left[u_{St} z_t\right]\right)^2}.$$

This is the asymptotic variance σ_ψ^2 that a "naive" IV estimator would report, ignoring that m^y and $\check{\eta}_t$ had to be estimated. If we did not have to estimate m^y and $\check{\eta}_t$, we would obtain the same asymptotic error σ_ψ^2.

The optimal Γ remains as in Proposition 3. This comes straightforwardly from the fact that the variance of the estimator is proportional to $\frac{\mathbb{E}\left[u_{\Gamma t}^2\right]}{\mathbb{E}\left[u_{St} u_{\Gamma t}\right]^2}$, which is minimized at Γ^*.

Calculating the error on $\phi_T^{d,e}$. The argument is exactly the same as for ψ. It is spelled out in the Online Appendix, Section C.

[43] In the notations of Newey and McFadden (1994)'s Theorem 6.1, we have $G_\gamma = 0$, i.e. the gradient of the moment condition $g^{2,z}$ with respect to the nuisance parameters $\gamma := \theta \backslash \{\psi\}$ is 0.

Online Appendix for

"Granular Instrumental Variables"

Xavier Gabaix and Ralph S.J. Koijen

April 4, 2023

Table of contents

B Appendix: Notations

d, s : Indicates demand and supply. E.g., $\phi^d, \phi^s = \frac{1}{\psi}$ are the elasticities of demand and supply.
ε_t, η_t: Aggregate shocks.
C_t^x: Controls affecting variable x.
λ: Factor loadings.
u_t: Idiosyncratic shocks.
$\check{u}_t \coloneqq Qu_t$: Idiosyncratic shocks residualized by a projection matrix Q.
$z_t = \Gamma' u_t$: GIV.

C Additional Proofs

C.1 Complement to the proof of Proposition 2

We detail the derivation of $\sigma_{\phi^d}(\Gamma)$, which follows the same steps as that of $\sigma_\psi(\Gamma)$. We have $y_{\tilde{E}t} - \phi^d p_t = e_t^y$ with $e_t^y \coloneqq \lambda_{\tilde{E}} \eta_t + u_{\tilde{E}t}$. This implies

$$\phi_T^{d,e} - \phi^d = \frac{\mathbb{E}_T\left[y_{\tilde{E}t} z_t\right]}{\mathbb{E}_T\left[p_t z_t\right]} - \phi^d = \frac{\mathbb{E}_T\left[\left(y_{\tilde{E}t} - \phi^d p_t\right) z_t\right]}{\mathbb{E}_T\left[p_t z_t\right]} = \frac{\mathbb{E}_T\left[e_t^y u_{\Gamma t}\right]}{\mathbb{E}_T\left[p_t u_{\Gamma t}\right]} = \frac{a_T}{d_T},$$

where

$$d_T = \mathbb{E}_T\left[p_t u_{\Gamma t}\right] \xrightarrow{a.s.} d \coloneqq \mathbb{E}\left[p_t u_{\Gamma t}\right] = \psi M \mathbb{E}\left[u_{St} u_{\Gamma t}\right].$$

By the central limit theorem, $\sqrt{T} a_T \xrightarrow{d} \mathcal{N}\left(0, \sigma_a^2\right)$, where

$$\sigma_a^2 = \mathbb{E}\left[\left(e_t^y\right)^2 u_{\Gamma t}^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left(e_t^y\right)^2 | u_t\right] u_{\Gamma t}^2\right] = \mathbb{E}\left[\sigma_{e^y}^2 u_{\Gamma t}^2\right] = \sigma_{e^y}^2 \sigma_{u_\Gamma}^2.$$

This implies that $\sqrt{T}\left(\phi_T^{d,e} - \phi^d\right) \xrightarrow{d} N\left(0, \sigma_{\phi^d}^2\right)$, where

$$\sigma_{\phi^d}(\Gamma) = \frac{\sigma_a}{|d|} = \frac{\sigma_{e^y} \sigma_{u_\Gamma}}{\left|\psi M \mathbb{E}\left[u_{St} u_{\Gamma t}\right]\right|}.$$

C.2 Complement to the proof of Proposition 4

Verification that the solution of $\mathbb{E}\left[g_t(\theta)\right] = 0$ is unique. First, we show that the equation $\mathbb{E}\left[g_t^1(\theta)\right] = 0$ identifies m^y. We observe that $\check{y}_{it} \coloneqq y_{it} - y_{Et}$ can be written $\check{y}_t = Q^\iota y_t$, with

$Q^{\iota} = I - \iota E'$, and likewise $\check{C}_t^y = Q^{\iota} C_t^y$. In the following, we shall use

$$R^{\check{\lambda}} Q^{\iota} = R^{\check{\lambda}}, \qquad Q^{\check{\lambda}} Q^{\iota} = Q^{\lambda}, \qquad Q^{\lambda} Q^{\iota} = Q^{\lambda}, \tag{72}$$

which are easy to verify. We have

$$\check{\eta}_t (m^y) \coloneqq R^{\check{\lambda}} (y_t - C_t^y m^y) = R^{\check{\lambda}} Q^{\iota} (y_t - C_t^y m^y) = R^{\check{\lambda}} \left(\check{y}_t - \check{C}_t^y m^y \right)$$

so

$$\begin{aligned} \check{y}_t - \check{C}_t^y m^y - \check{\lambda} \check{\eta}_t (m^y) &= \left(I - \check{\lambda} R^{\check{\lambda}} \right) \left(\check{y}_t - \check{C}_t^y m^y \right) = Q^{\check{\lambda}} \left(\check{y}_t - \check{C}_t^y m^y \right) \\ &= Q^{\check{\lambda}} Q^{\iota} (y_t - C_t^y m^y) = Q^{\lambda} (y_t - C_t^y m^y) \\ &= Q^{\lambda} \left(\iota \phi^d p_t + \lambda \eta_t + u_t \right) = Q^{\lambda} u_t = \check{u}_t \end{aligned} \tag{73}$$

Next, as $Q^{\lambda} \check{C}_t^y = Q^{\lambda} Q^{\iota} C_t^y = Q^{\lambda} C_t^y$, and using $Q^{\lambda} = \left(Q^{\lambda} \right)' Q^{\lambda}$ per (53),

$$-g_t^1 (\theta)' = \check{C}_t^{y\prime} Q^{\lambda} (y_t - C_t^y m^y) = \check{C}_t^{y\prime} Q^{\lambda\prime} Q^{\lambda} (y_t - C_t^y m^y) = \left(Q^{\lambda} C_t^y \right)' \left(Q^{\lambda} y_t - Q^{\lambda} C_t^y m^y \right)$$

so that, using the full rank Assumption 5, there is a unique solution.

Given m^y, the other equations ($\mathbb{E}\left[g_t^k (\theta) \right] = 0$ for $k = 2, 3$) are linear in all the other parameters in θ, so that unicity of the solution is guaranteed if those linear equations have full rank. We thus verify that they have full rank. Take the equation $\mathbb{E}\left[g_t^2 (\theta) \right] = 0$. Calling $\beta \coloneqq (\psi, b^p, m^{p\prime})$, $A_t \coloneqq \left(y_{St}, \check{\eta}_t, C_t^{p\prime} \right)$ and $B_t \coloneqq (z_t, \check{\eta}_t', C_t^p)$, we have $g_t^2 (\theta) = (-p_t + \beta A_t) B_t$. Hence, if $\mathbb{E}[A_t B_t]$ is invertible, the solution in β is unique. To verify that $\mathbb{E}[A_t B_t]$ is invertible, we remark that it is a upper triangular block matrix with block diagonal elements $\mathbb{E}[y_{St} z_t]$, $\mathbb{E}[\check{\eta}_t \check{\eta}_t']$, $\mathbb{E}\left[C_t^{p\prime} C_t^p \right]$. So, by (4), which implies the relevance condition $\mathbb{E}[y_{St} z_t] \neq 0$, and the full rank Assumption 5, $\mathbb{E}[A_t B_t]$ is indeed invertible.

The argument is the same for $\mathbb{E}\left[g_t^3 (\theta) \right] = 0$, using the relevance result $\mathbb{E}[p_t z_t] \neq 0$, which comes from $\mathbb{E}[p_t z_t] = \mu \mathbb{E}[u_{St} u_{\Gamma t}] \neq 0$ because of (4) and $\psi \neq 0$.

So, we proved that the solution of $\mathbb{E}[g_t (\theta)] = 0$ is indeed unique.

Derivation of the asymptotic variance of $\phi_T^{d,e}$. The argument is exactly the same as for ψ. We zoom on $g^{3,z}$:

$$g_t^{3,z} = \left(-y_{\tilde{E}t} + \phi^d p_t + b^y \check{\eta}_t (m^y) + C_{\tilde{E}t}^y m^y \right) z_t (m^y) \tag{74}$$

At the true value, $g_t^{3,z} = -\varepsilon_t^y z_t$. We thus have

$$\mathbb{E}\left[g_{\phi^d}^{3,z} \right] = \mathbb{E}[p_t z_t] = \mu \mathbb{E}[u_{St} u_{\Gamma t}], \tag{75}$$

which is nonzero because $\psi \neq 0$ implies $\mu \neq 0$, and (4) implies $\mathbb{E}\left[u_{St}u_{\Gamma t}\right] \neq 0$. But all the other derivatives $\mathbb{E}\left[g^{3,z}_{\theta_k}\right]$ are 0. We verify this in the same way, e.g.

$$\begin{aligned}
\mathbb{E}\left[g^{3,z}_{b^y}\right] &= \mathbb{E}\left[\check{\eta}_t z_t\right] = 0 \\
\mathbb{E}\left[g^{3,z}_{m_y}\right] &= \mathbb{E}\left[\left(\frac{d}{dm_y}\left(-y_{\tilde{E}t} + \phi^d p_t + b^y \check{\eta}_t\left(m^y\right) + C^y_{\tilde{E}t} m^y\right)\right) z_t - \varepsilon^y_t\left(\frac{d}{dm^y} z_t\right)\right] \\
&= \mathbb{E}\left[\left(\frac{d}{dm^y}\check{\eta}_t\left(m^y\right)\right) z_t\right] + \mathbb{E}\left[\varepsilon^y_t \Gamma' C^y_t\right] = 0 + 0 = 0
\end{aligned}$$

Indeed, $\mathbb{E}\left[\varepsilon^y_t \Gamma' C^y_t\right] = 0$ as $\varepsilon^y_t = \eta^{\perp}_{1t} + u_{\tilde{E}t}$, with $\eta^{\perp}_{1t}$ and $u_{\tilde{E}t}$ uncorrelated with C^y_t.

Hence, $\mathbb{E}\left[g^{3,z}_{\theta}\right]$ only has one non-zero component, the $\mathbb{E}\left[g^{3,z}_{\theta_k}\right]$ corresponding to $\theta_k = \psi$. Then, applying Newey and McFadden (1994)'s Theorem 6.1 allows us to conclude that:

$$\sigma^2_{\phi^d} = \frac{var\left(g^{3,z}_t\right)}{\left(\mathbb{E}\left[g^{3,z}_{\phi^d}\right]\right)^2} \tag{76}$$

Given $g^{3,z}_t = -\varepsilon^y_t z_t$,

$$var\left(g^{3,z}_t\right) = \sigma^2_{\varepsilon^y}\sigma^2_z,$$

and given (75), and that ε^y_t is homoskedastic conditional on u_t,

$$\sigma^2_{\phi^d} = \frac{\mathbb{E}\left[z^2_t\right]}{\mu^2 \mathbb{E}\left[u_{St} z_t\right]^2} var\left(\varepsilon^y_t\right).$$

The optimal Γ remains as in Proposition 3, even for $\sigma^2_{\phi^d}$. This is proven exactly as at the end of the proof of Proposition 4.

C.3 Proof of Proposition 5

We use the controls $c_t\left(m^y\right) = \left(C^p_t, C^y_{St}, \check{\eta}_t\left(m^y\right)\right)$. We define the function $g_t = \left(g^k_t\right)_{k=1...3}$ by:

$$\begin{aligned}
g^1_t\left(\theta\right) &\coloneqq \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y\right) + \check{C}^y_t m^y\right)' W \check{C}^y_t, & (77)\\
g^2_t\left(\theta\right) &\coloneqq \left(-p_t + \mu z_t\left(m^y\right) + b^p c_t\left(m^y\right)\right)'\left(z_t\left(m^y\right), c_t\left(m^y\right)\right), & (78)\\
g^3_t\left(\theta\right) &\coloneqq \left(-y_{St} + M z_t\left(m^y\right) + b^y c_t\left(m^y\right)\right)'\left(z_t\left(m^y\right), c_t\left(m^y\right)\right). & (79)
\end{aligned}$$

The reasoning is very similar to that in the proof of Proposition 4, so we are a bit more concise. Indeed, the unicity of the solution $\mathbb{E}\left[g_t\left(\theta\right)\right] = 0$, the consistency and $\sqrt{T}$-normality of the estimator are exactly like for the proof of proof of Proposition 4. The calculations of the asymptotic variance do change, however.

Variance of M_T^e. We use the moment function

$$g_t^{3,z} = \left(-y_{St} + Mz_t\left(m^y\right) + b^y c_t\left(m^y\right)\right) z_t\left(m^y\right).$$

At the true value θ_0, $g_t^{3,z} = -e_t^y z_t$. We take all derivatives at the true value θ_0. For notational simplicity, we again drop the t when the meaning is clear. We have $\mathbb{E}\left[g_M^{3,z}\right] = \mathbb{E}\left[z_t^2\right]$ and $\mathbb{E}\left[g_{b_y}^{3,z}\right] = \mathbb{E}\left[c_t z_t\right] = 0$. Less straightforwardly, but by the same reasoning as for the proof of Proposition 4, we have $\mathbb{E}\left[g_{m_y}^{3,z}\right] = 0$. Indeed, using (70),

$$\begin{aligned}\mathbb{E}\left[g_{m_y}^{3,z}\right] &= \mathbb{E}\left[\left(\frac{d}{dm_y}\left(-y_{St} + Mz_t\left(m^y\right) + b^y c_t\left(m^y\right)\right)\right) z_t - e_t^p\left(\frac{d}{dm^y} z_t\left(m^y\right)\right)\right] \\ &= \mathbb{E}\left[\left(-M\Gamma' C_t^y - b_{\check{\eta}}^y R^{\check{\lambda}} C_t^y\right) z_t\right] + \mathbb{E}\left[e_t^p \Gamma' C_t^y\right] = 0 + 0 = 0.\end{aligned}$$

Hence, $\mathbb{E}\left[g_\theta^{3,z}\right]$ only has one non-zero component, the $\mathbb{E}\left[g_{\theta_k}^{3,z}\right]$ corresponding to $\theta_k = M$. Then, applying Newey and McFadden (1994)'s Theorem 6.1 allows us to conclude that

$$\sigma_M^2 = \frac{var\left(g_t^{3,z}\right)}{\mathbb{E}\left[g_M^{3,z}\right]^2} = \frac{var\left(e_t^y z_t\right)}{\mathbb{E}\left[z_t^2\right]} = \frac{\mathbb{E}\left[\mathbb{E}\left[\left(e_t^y\right)^2 | u_t\right] z_t^2\right]}{\mathbb{E}\left[z_t^2\right]^2} = \frac{\mathbb{E}\left[\sigma_{e^y}^2 z_t^2\right]}{\mathbb{E}\left[z_t^2\right]^2} = \frac{\sigma_{e^y}^2}{\sigma_z^2}.$$

By the same reasoning (using $g^{2,z}$ rather than $g^{3,z}$), we have $\sigma_\mu^2 = \frac{\mathbb{E}\left[\left(e_t^p\right)^2\right]}{\mathbb{E}\left[z_t^2\right]}$.

The optimal Γ remains as in Proposition 3. This is proven exactly as at the end of the proof of Proposition 4.

C.4 Proof of Corollary 2

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

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

For a vector u, $\check{u} := Qu = u - \iota u_E$, which means $\check{u}_{it} = u_{it} - u_{Et}$. So, $z_t = \sum_i S_i \check{u}_{it} = \sum_i S_i\left(u_{it} - u_{Et}\right) = u_{St} - u_{Et}$. Finally, $\sigma_{u_\Gamma} = h\sigma_u$ comes from (57).

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

$$\sigma_{u_\Gamma}^2 = \sigma_u^2\left(h^2 - \frac{1}{N}\frac{x_S^2}{\sigma_x^2}\right). \tag{80}$$

where $\sigma_x^2 = \frac{\sum_k x_k^2}{N}$. This illustrates how controlling for more factors reduces the standard deviation of the GIV, resulting in a lower precision of the estimator (as in Proposition 2), especially if x_S^2 is large and N is small. An advantage of having lots of small firms (large N) is that they make the estimation of the aggregate shocks $\check{\eta}_t$ easier, and hence increase the precision of the GIV estimator by shrinking the last term in (80), $\frac{1}{N}\frac{x_S^2}{\sigma_{x_i}^2}$).

Derivation of (80). First note that $\frac{1}{N}\lambda'\lambda = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix}$. Define $\Omega := \lambda\left(\lambda'\lambda\right)^{-1}\lambda'$ so that $Q = I - \Omega$. With $X = (x_i)_{i=1\ldots N}$, $\Omega = \frac{1}{N}\iota\iota' + \frac{XX'}{N\sigma_x^2}$. This implies:

$$S'QS = S'S - S'\Omega S = S'S - \frac{1}{N} - \frac{\left(S'X\right)^2}{N\sigma_x^2} = h^2 - \frac{x_S^2}{N\sigma_x^2},$$

where $h^2 = S'S - \frac{1}{N}$ and $x_S = S'X$. So $\frac{\sigma_{u_\Gamma}^2}{\sigma_u^2} = S'QS = h^2 - \frac{x_S^2}{N\sigma_x^2}$.

C.5 Proof of Proposition 6

We provide the proof in both the homoskedastic and heteroskedastic cases, anticipating Proposition 8. For that, we will use two useful lemmata.

Lemma 2 *Suppose we have $x = \lambda b$ where x, λ and b are matrices with dimensions $N \times k$, $N \times r$ and $r \times k$ respectively. Then, for any $N \times N$ weight matrices W and $\tilde{W}$, using the definition (52),*

$$Q^{\lambda,\tilde{W}} = Q^{\lambda,\tilde{W}} Q^{x,W} \tag{81}$$

Proof of Lemma 2. As $Q^{\lambda,\tilde{W}}\lambda = 0$, we can write

$$Q^{\lambda,\tilde{W}} - Q^{\lambda,\tilde{W}}Q^{x,W} \;=\; Q^{\lambda,\tilde{W}}\left(I - Q^{x,W}\right) = Q^{\lambda,\tilde{W}}\left(xR^{x,W}\right) = Q^{\lambda,\tilde{W}}\lambda b R^{x,W} = 0,$$

so that $Q^{\lambda,\tilde{W}} = Q^{\lambda,\tilde{W}}Q^{x,W}$. $\square$

Lemma 3 *With $\mathcal{E} = R^{x,W}$ and $\Gamma' = S'Q^{\lambda,\tilde{W}}$, for $W = \left(V^u\right)^{-1}$ and $\tilde{W}$ any invertible $N \times N$ matrix:*

$$\mathbb{E}\left[\mathcal{E}u_t u_{\Gamma t}\right] = 0, \qquad \mathcal{E}x = I_k. \tag{82}$$

Proof of Lemma 3. We have $\mathcal{E}x = I_k$ by (53). Moreover, using (81),

$$\mathbb{E}\left[\mathcal{E}u_t u_{\Gamma t}\right] \;= \mathbb{E}\left[\mathcal{E}u_t u_t'\Gamma\right] = \; \mathcal{E}V^u\Gamma = R^{x,W}V^u\left(Q^{\lambda,\tilde{W}}\right)'S = R^{x,W}W^{-1}\left(Q^{x,W}\right)'\left(Q^{\lambda,\tilde{W}}\right)'S = 0,$$

as $R^{x,W}W^{-1}\left(Q^{x,W}\right)' = 0$, per the general properties in (53).[44] $\square$

[44]We note that $\mathbb{E}\left[u_{\Gamma t}u_{St}\right] = \Gamma'V^uS = S'Q^{\lambda,\tilde{W}}V^uS$, which we need to be non-zero. This is generically the case.

Proposition 9 *Take a vector* Γ *satisfying (4), so that* $z_t := \Gamma' (y_t - C_t^y m^y) = \Gamma u_t$*, and a* $k \times N$ *matrix* $\mathcal{E}$ *satisfying (82). Then, we can identify* $\dot{\phi}^d$ *via the moment:*

$$\mathbb{E}\left[\left(\mathcal{E}\left(y_t - C_t^y m^y\right) - \dot{\phi}^d p_t\right) z_t\right] = 0, \tag{83}$$

i.e. $\dot{\phi}^d = \frac{\mathbb{E}[\mathcal{E} y_t z]}{\mathbb{E}[p_t z_t]}$.

Note that (82) and (83), for the parametric-heterogeneous elasticity case, are the natural generalizations of the homogeneous elasticity case, in particular of (10), which said that $\mathbb{E}\left[u_{\tilde{E}t} u_{\Gamma t}\right] = 0$, and (11), which said that $\mathbb{E}\left[\left(y_{\tilde{E}t} - \phi^d p_t\right) z_t\right] = 0$: we replace $\tilde{E}'$ by $\mathcal{E}$.

Proof of Proposition 9. We have $y_t - C_t^y m^y - \phi^d p_t = \lambda \eta_t + u_t$ and $\mathcal{E}\phi^d = \mathcal{E} x \dot{\phi}^d = \dot{\phi}^d$, so

$$\mathcal{E}(y_t - C_t^y m^y) - \dot{\phi}^d = \mathcal{E}\left(y_t - C_t^y m^y - \phi^d p_t\right) = \mathcal{E}\left(\lambda \eta_t + u_t\right) = \mathcal{E}\lambda\eta_t + \mathcal{E} u_t$$

As $\mathbb{E}\left[\eta_t z_t\right] = 0$ and $\mathbb{E}\left[\mathcal{E} u_t u_{\Gamma t}\right] = 0$, we have $\mathbb{E}\left[\left(\mathcal{E}\left(y_t - C_t^y m^y\right) - \dot{\phi}^d p_t\right) z_t\right] = 0$. $\square$

Proof of Proposition 6. Together, Proposition 9 and Lemma 3 imply Proposition 6.

C.6 Proof of Proposition 7

With $Q^\iota = I - \iota E'$ the demeaning operator, we define the demeaned values $\check{y}_t = Q^\iota y_t$ and $\check{C}_t^y = Q^\iota C_t^y$, and the function $g_t(\theta) = \left(g_t^k(\theta)\right)_{k=0,\dots,3}$ as

$$\begin{aligned}
g_t^0(\theta) &:= \left(-\check{y}_t + \check{C}_t^y m^y\right)' \check{C}_t^y, && (84)\\
g_t^1(\theta) &:= \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y, \check{\lambda}\right) + \check{C}_t^y m^y\right) \check{\eta}_t\left(m^y, \check{\lambda}\right)' && (85)\\
g_t^2(\theta) &:= \left(-p_t + \psi y_{St} + b^p \check{\eta}_t\left(m^y, \check{\lambda}\right) + C_t^p m^p\right)\left(z_t\left(m^y, \check{\lambda}\right), \check{\eta}_t\left(m^y, \check{\lambda}\right), C_t^p\right) && (86)\\
g_t^3(\theta) &:= \left(-y_{\tilde{E}t} + \phi^d p_t + b^y \check{\eta}_t\left(m^y, \check{\lambda}\right) + C_{\tilde{E}t}^y m^y\right)\left(z_t\left(m^y, \check{\lambda}\right), \check{\eta}_t\left(m^y, \check{\lambda}\right)'\right) && (87)
\end{aligned}$$

The reasoning is very similar to that in the proof of Proposition 4, so we are a bit more concise.

At the true value, $\mathbb{E}\left[g_t(\theta_0)\right] = 0$**.** $\mathbb{E}\left[g_t^0(\theta_0)\right] = 0$ is equivalent to $\mathbb{E}\left[Q^\iota\left(\lambda\eta_t + u_t\right)' C_t^y Q^\iota\right] = 0$, which is true from our maintained assumption that $C_t^y \perp \eta_t, u_t$. $\mathbb{E}\left[g_t^1(\theta)\right] = 0$ is equivalent to $\mathbb{E}\left[Q^\iota \check{u}_t \check{\eta}_t'\right] = 0$, which is true as Lemma 1 implies $\check{u}_t \perp \check{\eta}_t$. $\mathbb{E}\left[g_t^k(\theta)\right] = 0$ with $k = 2, 3$ were verified in the proof of Proposition 4

The solution of $\mathbb{E}\left[g_t(\theta)\right] = 0$ **is unique.** We first show that the equation $\mathbb{E}\left[g_t^0(\theta)\right] = 0$ uniquely identifies m^y. Indeed it is equivalent to $m^y = \mathbb{E}\left[\left(\check{C}^y \check{C}_t^y\right)\right]^{-1} \mathbb{E}\left[\check{C}_t^{y\prime} Q^{\check{\lambda}} \check{y}_t\right]$, since $\mathbb{E}\left[\left(\check{C}^y \check{C}_t^y\right)\right]^{-1}$ has full rank by Assumption 5.

Next, we show that the equation $\mathbb{E}\left[g_t^1(\theta)\right]=0$ uniquely identifies $\check{\lambda}$. To do so, we are going to show that $\mathbb{E}\left[g_t^0(\theta)\right]=0$ is equivalent to saying that $\check{\lambda}^f$ is an eigenvector of $V^{\check{y}_t - C_t^y m^y}$. The proof goes as follows. For notational simplicity, let us replace $y_t - C_t^y m^y$ by y_t. Given our normalization $\frac{\lambda'\lambda}{N}=I_r$, we have $\frac{\check{\lambda}'\check{\lambda}}{N}=I_{r-1}$ so that $R^{\check{\lambda}}=\frac{1}{N}\check{\lambda}'$. Hence $\check{\eta}_t = R^{\check{\lambda}}Q^{\iota}y_t = \frac{1}{N}\check{\lambda}'y_t$. The moment $\mathbb{E}\left[g_t^1(\theta)\right]=0$ is then $\mathbb{E}\left[\left(-\check{y}_t+\check{\lambda}\check{\eta}_t\right)\check{\eta}_t'\right]=0$ i.e.

$$\check{\lambda}\mathbb{E}\left[\check{\eta}_t\check{\eta}_t'\right]=\mathbb{E}\left[\check{y}_t\check{\eta}_t'\right]=\mathbb{E}\left[\check{y}_t\check{y}_t\right]\frac{1}{N}\check{\lambda}=V^{\check{y}}\frac{1}{N}\check{\lambda}.$$

As part of our normalizations, $\mathbb{E}\left[\check{\eta}_t\check{\eta}_t'\right]$ is diagonal i.e. $\mathbb{E}\left[\check{\eta}_t\check{\eta}_t'\right]=Diag\left(d_f\right)_{f=1\dots r-1}$ for some d_f.[45] Post-multiplying by e_f, the basis vector with entry 1 at position f and 0 elsewhere,

$$\check{\lambda}^f N d_f = V^{\check{y}}\check{\lambda}^f$$

Therefore, $\check{\lambda}^f$ is an eigenvector of $V^{\check{y}}$, with eigenvalue Nd_f. Given our normalizations stated before Proposition 4.2, this ensures that $\check{\lambda}^f$ is indeed uniquely identified: it is the f-th largest eigenvector of matrix $V^{\check{y}_t - C_t^y m^y}$.

Finally, given m^y and λ, the other parameters of θ satisfy a linear equation, like in the proof of Proposition 4. So, by the rank conditions of Assumption 5, the solution is unique.

The estimator is consistent. As in the proof of Proposition 4, we proceed to apply general GMM results (Hansen (1982); Newey and McFadden (1994)). In particular under the regularity conditions and i.i.d. sampling, by Newey and McFadden (1994), Theorem 2.6, the estimator is consistent.

The estimator is $\sqrt{T}$-normally distributed. Under the regularity conditions we assumed (and in particular that $\psi\phi^d \neq 1$, which ensures bounded gradients), by Newey and McFadden (1994), Theorem 3.4, $\sqrt{T}\left(\theta_T^e-\theta\right)$ converges in distribution to a normal distribution with mean 0 and positive definite variance covariance matrix V^θ.

C.7 Proof of Proposition 8

The earlier proofs were written in such a way that they readily generalize, where one replaces Q^λ, R^λ by $Q^{\lambda,W}$ and $R^{\lambda,W}$ as in (52), with $W=(V^u)^{-1}$ as spelled out in Proposition 8. Similarly, cross-sectional moments in the proofs become weighted by W as stated in Proposition 8, for instance $g_t^1(\theta)$ in (65) becomes $g_t^1(\theta) := \left(-\check{y}_t+\check{\lambda}\check{\eta}_t\left(m^y\right)+\check{C}_t^y m^y\right)' W\check{C}_t^y$. Then all the arguments above go through, with heteroskedasticity. To generalize the optimality of Γ^* in Proposition 3, the arguments just repeat the Cauchy-Schwarz arguments of the proof of Proposition 3 for the optimality of Γ^*.

[45] Indeed, as is well-known, $\check{\lambda}\check{\eta}_t$ can be expressed as $\check{\lambda}U'U\check{\eta}_t$ for any unitary matrix U (with $U'U=I_r$). We can choose U so that $U\check{\eta}_t$ has diagonal variance-covariance matrix.

D Complements

D.1 Detailed links with previous literature

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

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

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

There are two main differences compared to GIV. First, both papers use idiosyncratic shocks to another variable than y_t, say g_t, to construct an instrument for y_{wt}. Leary and Roberts (2014) 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 H.3).

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

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

In a tangentially related recent paper, Sarto (2022) uses factor analysis to extract values of η_t (much as we do when we "recover" a factor η_t). Take the basic example in our paper. Then, Sarto

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

does not identify $\phi^s = \frac{1}{\psi}$: even if η_t (the aggregate shock to demand) were perfectly identified, that would not allow to estimate p_t. In the supply and demand example, Sarto's approach would identify the demand elasticity ϕ^d, but not the supply elasticity ϕ^s.

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

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 general equilibrium effects as opposed to local effects. Still, some of the sophisticated techniques of the spatial literature might be used one day to enrich a GIV-type analysis.

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

In the case of GIVs, we generally start from a structural model that motivates the estimating equation, as in our empirical example. This prescribes the definition of the size vector S and, in some cases, the characteristics that determine the exposures x_{it}. To extract idiosyncratic shocks, we rely on statistical factor models.[47]

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

[47]We discuss the robustness of GIVs to various forms of misspecification in Section 3.3.

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.

D.2 A benchmark model: A simple supply and demand model

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

$$y_{it} = \phi^d p_t + \lambda_i \eta_t + u_{it}, \tag{88}$$

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

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

$$s_t = \phi^s p_t + \varepsilon_t^s, \tag{89}$$

where $\phi^s = \frac{1}{\psi}$ is the elasticity of demand and ε_t^s is a supply shock. This can also be written as in (2)

$$p_t = \psi y_{St} + \varepsilon_t, \qquad \psi = \frac{1}{\phi^s} \tag{90}$$

with $\varepsilon_t := -\frac{\varepsilon_t^s}{\phi^s}$

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

$$p_t = \frac{u_{St} + \lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi^d} = \mu u_{St} + \varepsilon_t^p, \tag{91}$$

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

$$s_t = y_{St} = \frac{\phi^s u_{St} + \phi^s \lambda_S \eta_t - \phi^d \varepsilon_t}{\phi^s - \phi^d} = M u_{St} + \varepsilon_t^s, \tag{92}$$

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

D.3 Estimating the variance of heteroskedastic idiosyncratic shocks

When the u_{it} have a potentially different variance $\sigma_{u_i}^2$, we need to estimate $\sigma = diag\left(\sigma_{u_i}\right)$. This is not a difficult problem, but it requires some care as the u_i are estimated as residuals from a factor model. Accordingly, we proceed in the following way. First, we make an observation.

Lemma 4 *Suppose that we have recovered residuals $\check{u}_t = Q^{\Lambda,W} u_t$ for some arbitrary Λ, with $W = \left(V^u\right)^{-1} = diag\left(\sigma_{u_i}^{-2}\right)$ and $Q^{\Lambda,W}$ defined as in (52). Then,*

$$var\left(\check{u}_{it}\right) = Q_{ii}^{\Lambda,W} \sigma_{u_i}^2 \tag{93}$$

Proof. We have

$$V^{\check{u}} = \mathbb{E}\left[\check{u}_t \check{u}_t'\right] = Q\mathbb{E}\left[u_t u_t'\right] Q' = Q V^u Q' = Q W^{-1} Q'.$$

Using $(I - Q) W^{-1} Q' = 0$ (see (53)), $V^{\check{u}} = W^{-1} Q' = V^u Q'$, so, using the fact that V^u is diagonal, we get (93). □

Next, we have the following.

Proposition 10 (GIV estimation when the variances $\sigma_{u_i}^2$ are estimated). *Let the assumptions in Proposition 8 hold, except that we now we need to estimate $W = diag\left(\sigma_{u_i}^{-2}\right)$. Take the system of Proposition 8, and append $\left(\sigma_{u_i}^2\right)_{i=1\ldots N}$ to the vector θ of parameters to estimate; and append the collection of N moments:*

$$\sum_{t=1}^{T} \check{u}_{it}^2 = T Q_{ii}^{\lambda,W} \sigma_{u_i}^2, \qquad \check{u}_t := Q^{\lambda,W}\left(y_t - C_t^y m^y\right) \tag{94}$$

to the moments of Proposition 8. Then, under the true value θ_0, all the moments, including (94), hold in expectation. Furthermore, assuming $\psi \neq 0$ and that θ_0 is the unique solution of the aforementioned moment conditions, then the vector θ_0 is consistently estimated (with $T \to \infty$ and fixed N), and we have $\sqrt{T}\left(\theta_T^e - \theta_0\right) \to N\left(0, V^\theta\right)$ for a matrix V^θ.

Proof. Using the matrix $q := I - \iota\tilde{E}'$, we demeaned $\check{y}_t = y_t - \iota y_{\tilde{E}t}$ is also

$$\check{y}_t := qy_t = q\lambda\eta_t + qu_t$$

Next, with $\lambda = (\iota, \Lambda)$, we apply $Q^{\lambda,W}$. We remark that:[48]

$$Q^{\lambda,W} = Q^{\lambda,W}q \tag{95}$$

Hence,

$$Q^{\lambda,W}\check{y}_t := Q^{\lambda,W}qy_t = Q^{\lambda,W}\left(\check{\lambda}\eta_t + qu_t\right) = Q^{\lambda,W}u_t =: \check{u}_t$$

Hence we obtain $\check{u}_t$ as the residual. We then use the moment (93), with $W = (V^u)^{-1}$ i.e. (94). This shows that,

$$\mathbb{E}\left[\check{u}_{it}^2\right] = Q_{ii}^{\lambda,W}\sigma_{u_i}^2$$

Formally, defining

$$g_t^u(\theta) := \left(\check{u}_{it}^2(\theta) - Q_{ii}^{\lambda,W}(\theta)\,\sigma_{u_i}^2(\theta)\right)_{i=1\ldots N}, \qquad \check{u}_t(\theta) := Q^{\lambda,W}(\theta)\left(y_t - C_t^y m^y(\theta)\right) \tag{96}$$

then we have, at the true value θ_0,

$$\mathbb{E}\left[g_t^u(\theta_0)\right] = 0_N.$$

We append $g_t^u(\theta)$ to the vector of conditions $g_t(\theta)$. We just proved that at the true value, $\mathbb{E}\left[g_t(\theta_0)\right] = 0$. We also assumed that this was the only solution.[49]

We thus proceed to apply general GMM results (Hansen, 1982; Newey and McFadden, 1994). In particular under the regularity conditions and i.i.d. sampling, by Newey and McFadden (1994), Theorem 2.6, the estimator is consistent.

The estimator is $\sqrt{T}$-normally distributed. Under the regularity conditions we assumed, by Newey and McFadden (1994), Theorem 3.4, $\sqrt{T}\left(\theta_T^e - \theta_0\right)$ converges in distribution to a normal distribution with mean 0 and positive variance covariance matrix V^θ. □

There is a fixed point in (94): given the $\sigma_{u_i}^2$'s, we form $W = diag\left(\sigma_{u_i}^{-2}\right)$, which in turns gives an estimate of $\sigma_{u_i}^2$ via (94).

Stochastic volatility This procedure assumes constant volatility. It could also be extended, to stochastic volatility, along the lines of the GARCH literature. Recall that the generalization of $E_i = \frac{1}{N}$ was $\tilde{E}_i := \frac{1/\sigma_{u_i}^2}{\sum_j 1/\sigma_{u_j}^2}$, see (54). Then, the correct generalization of (54) is $\tilde{E}_{it} := \frac{1/\sigma_{u_{it|t-1}}^2}{\sum_j 1/\sigma_{u_{jt|t-1}}^2}$, where $\sigma_{u_{it|t-1}}^2 = \mathbb{E}_{t-1}\left[u_{it}^2\right]$ is the conditional expected variance.[50] Then, with $\Gamma_t = S - E_t$, we have

[48] Indeed, we have $Q^{\lambda,W} - Q^{\lambda,W}q = Q^{\lambda,W}(I - q) = Q^{\lambda,W}\iota\tilde{E}' = 0$.

[49] While so far we dealt with block-linear equations in θ, now we have a non-linear equation. We suspect that in most cases of interest, the solution is unique, but we assume this rather than attempt to prove it.

[50] The reason is that all we need is to control expressions of the type $\mathbb{E}_{t-1}\left[u_t u_t'\right]$, i.e. $\mathbb{E}_{t-1}\left[V_t^u\right]$.

the generalization of (10),[51]

$$\mathbb{E}_{t-1}\left[u_{\tilde{E}_t t} u_{\Gamma_t t}\right] = 0 \tag{97}$$

Likewise, when constructing the weighing matrices (52) with $W = (V^u)^{-1}$, we replace W by $W_{t-1} = (\mathbb{E}_{t-1}[V_t^u])^{-1}$. Those remarks might be useful to a full econometric analysis of GIV with stochastic volatility, but we leave such an analysis to future research.

D.4 Relation between Bartik instruments and GIVs

D.4.1 Relating the Bartik setup to the GIV setup

Bartik instruments are widely used to estimate parameters of interest. In this appendix, we compare the assumptions under which Bartik instruments are valid to those under which GIVs are valid. This comparison is useful also to highlight settings in which GIVs can and cannot be used (and vice versa for Bartik).

As a general matter, in a number of cases where a cross-section is used (e.g. Autor et al. (2013)), Bartik applies, but GIV does not apply, for instance because there is no large idiosyncratic shock that one can use.

Next, to study the difference between GIV and Bartik more analytically, we start from the setup in Borusyak et al. (2022), and then map it to our model.[52] Their model can be summarized as

$$y_l = \beta x_l + \epsilon_l,$$

where we omit observable controls, $w_l'\gamma$. In this specification, l corresponds to locations. The endogeneity concern is that $\mathbb{E}[x_l \epsilon_l] \neq 0$. The endogenous variable can be written in terms of industry-location shares, where industries are indexed by n,

$$x_l = \sum_n s_{ln} g_{ln},$$

and $\sum_n s_{ln} = 1$. To connect Bartik instruments to GIV, we assume a simple factor model in g_{ln},

$$g_{ln} = g_n + \tilde{g}_{ln},$$

[51]Derivation: with $\tilde{E}_{it} := k_t/\sigma^2_{u_{it|t-1}}$, with $\sum_i \tilde{E}_{it} = 1$,

$$\begin{aligned}
\mathbb{E}_{t-1}\left[u_{\tilde{E}_t t} u_{\Gamma_t t}\right] &= \sum_i \tilde{E}_{it}\left(S_i - \tilde{E}_{it}\right)\mathbb{E}_{t-1}\left[u_{it}^2\right] = \sum_i \tilde{E}_{it}\left(S_i - \tilde{E}_{it}\right)\sigma^2_{u_{it|t-1}} = \sum_i k_t\left(S_i - \tilde{E}_{it}\right) \\
&= k_t\left(\sum_i S_i - \sum_i \tilde{E}_{it}\right) = k_t(1-1) = 0.
\end{aligned}$$

[52]We are grateful to a referee for suggesting this connection.

that is, the loadings on the common factor, g_n, are equal to one. In Bartik applications, a concern is typically that $\mathbb{E}[\tilde{g}_{ln}\epsilon_l] \neq 0$, for instance, when local economic conditions in location l are correlated with the idiosyncratic growth rate of industry n in location l.

To express the identifying assumption in Borusyak et al. (2022), they write the model at the industry level

$$\bar{y}_n = \alpha + \beta \bar{x}_n + \bar{\epsilon}_n, \tag{98}$$

where $\bar{b}_n = \frac{\sum_l s_{ln} b_l}{\sum_l s_{ln}}$, for some variable b_l. The shares, s_{ln}, are assumed to be non-stochastic, and the main identifying assumption is that $\mathbb{E}[g_n \bar{\epsilon}_n] = 0$.

D.4.2 Defining and comparing the Bartik and GIV instruments

The Bartik instrument is defined as

$$z_n^{Bartik} = \frac{1}{L} \sum_l g_{ln}.$$

The GIV is defined as

$$z_n^{GIV} = \sum_l \tilde{s}_{ln} g_{ln} - \frac{1}{L} \sum_l g_{ln},$$

where $\tilde{s}_{ln}$ is the location share of industry n so that $\sum_l \tilde{s}_{ln} = 1$. Hence, we have $\lim_{L\to\infty} z_n^{Bartik} = g_n$ and $\lim_{L\to\infty} z_n^{GIV} = \lim_{L\to\infty} \sum_l \tilde{s}_{ln}\tilde{g}_{ln} = \tilde{g}_{\tilde{S}n}$. In the remainder of this section, we work with the large L version of the instruments to simplify the exposition.

For Bartik instruments to be valid in (98), we need $\mathbb{E}[g_n \bar{\epsilon}_n] = 0$. For the GIV to be valid, we need $\mathbb{E}[\sum_l \tilde{s}_{ln}\tilde{g}_{ln}\bar{\epsilon}_n] = \mathbb{E}[\sum_l \tilde{s}_{ln}\tilde{g}_{ln}\frac{\sum_l s_{ln}\epsilon_l}{\sum_l s_{ln}}] = 0$. As $\mathbb{E}[\tilde{g}_{ln}\epsilon_l]$ may not be zero in cross-sectional settings, as discussed before, GIV is not the most natural instrument in those circumstances.

By the same logic, the identifying assumption of the Bartik instrument may be less appealing in settings where the identifying assumption of the GIV is more plausible. To connect the Borusyak et al. (2022) setup to the one we consider in this paper, we relabel l to $i = 1, ..., N$ and n to $t = 1, ..., T$. In addition, we set $\lambda_i = 1$, g_{ln} to y_{it}, g_n to η_t, $\tilde{g}_{ln}$ to u_{it}, which implies $y_{it} = \eta_t + u_{it}$. For simplicity, we assume that the shares do not vary across time (and thus across n in the Bartik setup). We use S_i, with $S_{it} = S_i$, to denote the relative size such that $\sum_i S_i = 1$.

We redefine the Bartik instrument and the GIV using these definitions:

$$z_t^{Bartik} = \frac{1}{N} \sum_i g_{it} = g_{Et},$$

and

$$z_t^{GIV} = \sum_i S_i g_{it} - \frac{1}{N} \sum_i g_{it}.$$

Hence, we have $\lim_{N\to\infty} z_t^{Bartik} = \eta_t$ and $\lim_{N\to\infty} z_t^{GIV} = u_{St}$. As before, we work with the large N

version of the instruments to simplify the exposition. To provide a simple example where the Bartik instrument may be less appealing, we consider a trivial version of our baseline model in Section 2 (with $\phi^d = 0$)

$$\begin{aligned} y_{it} &= \eta_t + u_{it}, \\ s_t &= \phi^s p_t + \varepsilon_t, \end{aligned}$$

where $p_{St} = (\eta_t + u_{St} - \varepsilon_t)/\phi^s$. If we average the model in the cross-section (again using the limit when $N \to \infty$)

$$\begin{aligned} y_{Et} &= \eta_t, \\ s_t &= \phi^s p_t + \varepsilon_t, \end{aligned}$$

To estimate ϕ^S, we can use two instruments. First, we can use GIV, which requires

$$\mathbb{E}\left[\varepsilon_t u_{St}\right] = 0.$$

Alternatively, we can use the Bartik instrument and assume

$$\mathbb{E}\left[\varepsilon_t \eta_t\right] = 0.$$

In this example, the Bartik instrument requires assumptions that are too strong in models in which η_t and ε_t are correlated. These are the settings that we focus on in this paper, and GIV is well suited to estimate the parameters of interest.

D.5 More general setup and multipliers

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

D.5.1 Framework

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

$$y_{it} = \sum_f \lambda_{it}^f F_t^f + u_{it} + C_{it}^y m, \tag{99}$$

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

$$F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f. \tag{100}$$

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

We use the structure (99)-(100) because many economic models of interest follow this structure, at least after linearization, so that the GIV allows to estimate some of their parameters.

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

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

$$\lambda_{it}^f = \dot{\lambda}_0^f + \dot{\lambda}_1^f x_{it}^f, \tag{101}$$

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:

$$\mathbb{E}\left[u_{it}\left(\eta_t^f, C_t^y, C_t^f, x_t^f\right)\right] = 0, \tag{102}$$

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

We rewrite model (99) in vector form:

$$y_t = \lambda_t F_t + u_t + C_t^y m, \qquad F_t^f = \alpha^f y_{St} + \eta_t^f + C_t^f m^f, \tag{103}$$

with λ_t a $N \times r$ matrix, F_t a $r \times 1$ vector, C_t^y an $N \times c$ matrix, m is $c \times 1$, where c is the dimension of the controls.[53]

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

D.5.2 Multipliers

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

$$y_{St} = M_t \left(u_{St} + \varepsilon^y_t \right), \tag{104}$$

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

$$M_t = \frac{1}{1 - \lambda_{St}\alpha} = \frac{1}{1 - \sum_f \lambda^f_{St} \alpha^f}. \tag{105}$$

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:

$$F^f_t = M_t \alpha^f u_{St} + \varepsilon^f_t, \tag{106}$$

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

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

D.6 The GIV with multi-dimensional outcomes for each entity

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

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

$$\begin{aligned} y^a_{S^a t} &= \sum_f \lambda^a_{S^a, f} F^f + u^a_{S^a t}, \\ F^f_t &= \eta^f_t + \sum_a \alpha^f_a y^a_{S^a, t}, \end{aligned}$$

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

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

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

where

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

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

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

This implies

$$y_{St} = M\left(u_{St} + \varepsilon_t\right), \tag{107}$$

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

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

We will form a GIV:

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

which is $q-$dimensional: $u_\Gamma = \left(u^a_{\Gamma^a}\right)_{a=1,\ldots,q}$. 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, $\Omega^{ab} = 0$, where

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

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

We have:

$$\Omega^{ab} = \sum_i E^a_i \Gamma^b_i v^{ab}_i, \qquad v^{ab}_i := \mathbb{E}\left[u^a_{it} u^b_{it}\right]. \tag{108}$$

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

$$v^{ab}_i = \sigma^2_i v^{ab}. \tag{109}$$

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

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

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

Given (107), 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

$$M = \mathbb{E}\left[y_{St} z_t'\right] \mathbb{E}\left[z_t z_t'\right]^{-1}. \tag{110}$$

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

$$u_{\Gamma t,-i} := u_{\Gamma t} - S_i^u u_{it}, \tag{111}$$

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

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

Hence, we can identify μ_i, by regression

$$y_{it} = \mu_i M u_{\Gamma t,-i} + \phi^i \mathcal{C}_t + \varepsilon_{it}^y, \tag{112}$$

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

D.7 Nonlinear GIV

We imagine a nonlinear GIV. Suppose that instead of the simple $p_t = \psi y_{St} + \varepsilon_t$ (equation (1)) we have a more complex

$$p_t = \Phi\left(y_{St}, \psi\right) + \varepsilon_t \tag{113}$$

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

$$\mathbb{E}\left[\left(p_t - \Phi\left(y_{St}, \psi\right)\right) z_t\right] = 0 \tag{114}$$

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

D.8 When the researcher assumes too much homogeneity

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

The model becomes, for the demand:

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

and for the supply

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

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

$$p_t = \frac{u_{St} + \lambda_S \eta_t - \varepsilon_t}{\phi^s - \phi_S^d}. \tag{115}$$

In this thought experiment, the econometrician assumes identical elasticities of demand across countries, $\phi_i^d = \phi^d$. He runs a panel model for $y_{it} - y_{Et}$, and we assume that it's large enough that he can extract η_t successfully.[54] The GIV (we use the notation Z_t rather than z_t to denote the GIV before controls by η_t) is then

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

so

$$Z_t = \frac{1}{\psi} u_{\Gamma t} + \lambda^Z \tilde{\eta}_t, \qquad \frac{1}{\psi} = \frac{\phi^s - \phi_E^d}{\phi^s - \phi_S^d}, \tag{116}$$

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

Hence, when we run the first stage

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

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

we will gather

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

If we run

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

we will estimate

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

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

In the polar opposite case where η_t cannot be estimated or controlled for, then the simple procedure becomes biased, however, as (116) shows. To fix it, one can estimate the model with non-parametric coefficients (Section D.9).

D.9 Heterogeneous demand elasticities: Non-parametric extension

Non-parametric version for ϕ^d We present a variant of the procedure in Section 4.1, but now with non-parametric heterogeneous demand elasticities ϕ^d_i. The model is

$$y_t = \phi^d p_t + \lambda \eta_t + u_t, \tag{117}$$

We still assume parametric loading of unobserved factors η.

We propose two procedures to estimate ϕ^d.

D.9.1 First procedure for the nonparametric estimation of heterogeneous demand elasticities

Recall the model (117). We replace ϕ^d by $\phi = (\phi_1, \ldots, \phi_N)'$ for notational simplicity:

$$y_t = \phi p_t + \lambda \eta_t + u_t, \tag{118}$$

We now need not assume parametric knowledge of λ.

We define $\Gamma^* := \left(Q^{\langle \lambda, \phi \rangle}\right)' S$, where Q is the usual projection operatoer $Q^\Lambda = I - \Lambda \left(\Lambda' \Lambda\right)^{-1} \Lambda$, with Λ formed by the span of λ and ϕ. We assume that $\Gamma^* \neq 0$: this is saying that S is not spanned by the factor loadings λ and the vector of demand elasticities ϕ.

We propose the following procedure.

1. Guess a candidate for ϕ, called ϕ^c, and $W = (V^u)^{-1}$ (initially, as it's enough to know all those up to a multiplicative factor, we might take $\phi^c = \iota$, and $W = I$, or $W = Diag\left(1/var\left(y_{it}\right)\right)$).

We define $Q^{\phi} := Q^{\phi^c,W}$, keeping W implicit in this step and the next. If $\phi^c = \phi$, then

$$Q^{\phi} y_t = \left(Q^{\phi} \lambda\right) \eta_t + Q^{\phi} u_t \tag{119}$$

2. We estimate $\check{\lambda} := Q^{\phi}\lambda$, $\eta_t^e = R^{\check{\lambda}} Q^{\phi} y_t$, 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 the vector of sensitivities ϕ. We use a specific instrument z_{it} for each entity i. We proceed as follows:
   (a) We define the debiasing vector a^i. As $\check{u}_t = Q u_t$, we have $V^{\check{u}} = Q V^u Q'$, and we define

   $$a^i_j := \frac{V^{\check{u}}_{ij}}{V^{\check{u}}_{ii}} \tag{120}$$

   (b) We define the instrument for entity i,

   $$z_{it} := S' \left(\check{u}^e_t - a^i \check{u}^e_{it}\right) \tag{121}$$

   Morally, it's the size weighted sum idiosyncratic shocks of the entities different from i.[55]

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

   $$\mathbb{E}\left[\left(y_{it} - \phi_i p_t - \lambda_i \eta^e_t\right) z_{it}\right] = 0 \tag{124}$$

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

This algorithm also applies to the parametric case where we know that $\phi_{it} = X_i \dot{\phi}$ (Section 4.1), but keep the loadings λ non-parametric. Then, in steps 1-2 we replace ϕ by X, and in the last step we replace ϕ by $X\dot{\phi}$ and estimate $\dot{\phi}$.

Proposition 11 (Moment conditions to identify non-parametric elasticities). *Define $\check{u}_t = Q^{\check{\lambda},\phi} u_t$ in the notations above, and the entity-i specific GIV z_{it} defined in* (121). *Then the moment condition* (124) *holds.*

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

$$a^i_j = 1_{i=j} \tag{122}$$

so that $z_{it} = \sum_{j \neq i} S_j \check{u}^e_{jt}$ is a "leave one out" estimator. In the general case, z_{it} is in some sense a refined quasi-leave one out estimator, refined so that (126) holds. In the simple case of Section 2.1 with an additive shock ($\lambda_i \equiv 1$), we just have $\check{u}_{it} = u_{it} - u_{Et}$, i.e. $Q_{ij} = 1_{i=j} - \frac{1}{N}$ and

$$a^i_j := \frac{1_{i=j} - \frac{1}{N}}{1 - \frac{1}{N}}. \tag{123}$$

This may be useful as a starting point numerically.

Proof of Proposition 11. Definition (120), together with $\check{u}_t = Qu_t$ and $V^{\check{u}} = QV^uQ'$, implies

$$a_j^i = \frac{\mathbb{E}\left[\check{u}_{jt}\check{u}_{it}\right]}{\mathbb{E}\left[\check{u}_{it}^2\right]} \tag{125}$$

so that

$$\mathbb{E}\left[z_{it}\check{u}_{it}\right] = \mathbb{E}\left[S'\left(\check{u}_t - a^i\check{u}_{it}\right)\check{u}_{it}\right] = 0 \tag{126}$$

As z_{it} is also uncorrelated with η_t^e, moment (124) holds. □

D.9.2 Second procedure for the nonparametric estimation of heterogeneous demand elasticities

We premultiply (117) by $Q = Q^\lambda$ and set $\check{x}_t := Qx_t$. So $\check{y}_t = \check{\phi}^d p_t + \check{u}_t$. With $\Gamma = Q'S$, we have $y_{\Gamma t} = \phi_\Gamma^d p_t + u_{\Gamma t}$. To ease on notations, we call $\psi := \phi_\Gamma^d$. Given a candidate estimate ψ^c of ψ we form the associated GIV: $z_t\left(\psi^c\right) := y_{\Gamma t} - \psi^c p_t$.

If we have the correct z_t, the following moments hold[56,57], with $b^p = \frac{1}{\phi^s - \phi_S^d}$ the coefficient of the first stage regression (205), $p_t = b^p z_t + \varepsilon_t^p$,

$$\begin{aligned} \mathbb{E}\left[\left(y_t - \phi^d p_t\right) z_t\right] &= V^u\Gamma, \qquad \mathbb{E}\left[\left(p_t - b^p z_t\right) z_t\right] = 0 \qquad (129) \\ \mathbb{E}\left[\left(\check{y}_{it} - \check{\phi}_i^d p_t\right)^2\right] &= V_{ii}^{\check{u}}, \qquad \mathbb{E}\left[z_t^2\right] = \Gamma' V^u \Gamma \qquad (130) \end{aligned}$$

which potentially allow to estimate, respectively, ϕ^d, b^p (hence ϕ^s), V_{ii}^u and ϕ_Γ^d. Indeed, if we know z_t, we know ϕ^d and b^p.[58]

We examine in more detail how to estimate $\psi := \phi_\Gamma^d$. Calling the true value $z_t\left(\psi\right) = u_{\Gamma t}$, 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 (130). So, we solve for ψ^c (a *c*andidate answer for ψ) so that the empirical variance of the GIV is equal to its theoretical variance:

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

i.e. $\mathbb{E}\left[p_t^2\right]\left(\psi^c\right)^2 - 2\mathbb{E}\left[y_{\Gamma t}p_t\right]\psi^c + \mathbb{E}\left[y_{\Gamma t}^2\right] - \sigma_{u_\Gamma}^2 = 0$. This is a quadratic equation in ψ^c, which yields

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

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

$$\mathbb{E}\left[\left(y_{\tilde{E}t} - \phi_{\tilde{E}}p_t\right) z_t\right] = 0 \tag{127}$$

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

$$\mathbb{E}\left[\left(\check{y}_{it} - \check{\phi}_i p_t\right) z_t\right] = \left(QV^u\Gamma\right)_i \tag{128}$$

[58] We recommend starting from the parametric estimates of Section 4.1, which gives potentially good starting values for ϕ^d, z_t and V^u.

two roots:[59] 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:

$$b^{p,B} = -b^{p,G}, \tag{131}$$

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

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

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

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

$$\mathbb{E}\left[z_t^B p_t\right] = -\mathbb{E}[z_t p_t], \qquad \mathbb{E}\left[\left(z_t^B\right)^2\right] = \mathbb{E}\left[z_t^2\right] \tag{132}$$

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

$$b^{p,B} = \frac{\mathbb{E}\left[p_t z_t^B\right]}{\mathbb{E}\left[\left(z_t^B\right)^2\right]} = \frac{-\mathbb{E}[p_t z_t]}{\mathbb{E}[z_t^2]} = -b^{p,eG} \tag{133}$$

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

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

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

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

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

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

linear combination of $\dot{\varepsilon}_t$ and η_t), $p_t = \alpha M u_{St} + \ddot{\varepsilon}_t$, hence:

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

We wish to estimate ϕ and αM.

We consider the vector $Y_t = (y_t', p_t)'$ stacking together y_t and p_t. Then, with $U_t = (u_t', 0)'$, $\Phi = (\phi', 1)'$, $\lambda = (\lambda', 0)'$, and adding a weight "0" to the last component of the vector S (extended here to have 1 more component, with a mild abuse of notations) we have[60]

$$Y_t = \lambda\eta_t + \Phi\ddot{\varepsilon}_t + \alpha M\Phi u_{St} + U_t \tag{134}$$

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

$$Y_t = \lambda\eta_t + \Psi\varepsilon_t + \Psi u_{St} + U_t = \lambda\eta_t + \Psi\varepsilon_t + (I + \Psi S')\, U_t \tag{135}$$

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

$$\begin{aligned}
V^Y &= \sigma_\eta^2\lambda\lambda' + \sigma_\varepsilon^2\Psi\Psi' + \sigma_{\eta\varepsilon}\left(\lambda\Psi' + \Psi\lambda'\right) + (I + \Psi S')\, V^U\, (I + S\Psi') && (136)\\
&= \sigma_\eta^2\lambda\lambda' + \gamma\Psi\Psi' + \Psi b' + b\Psi' + V^U && (137)\\
b &= \sigma_{\eta\varepsilon}\lambda + V^U S && (138)\\
\gamma &= \sigma_\varepsilon^2 + S'V^U S && (139)
\end{aligned}$$

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

$$(\Psi - a)\,(\Psi - a)' = C := \frac{1}{\gamma}\left(V^Y - V^U - \sigma_\eta^2\lambda\lambda'\right) + aa' \tag{140}$$

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

$$(\Psi^c - a)\,(\Psi^c - a)' = C$$

We know that this identity holds under the correct root, so that $C = (\Psi - a)\,(\Psi - a)'$. Now, there are two solutions to the equation $XX' = DD'$, with X the unknown vector and D a known vector: $X = D$ and $X = -D$. Hence, the two solutions are $\Psi^c - a = \Psi - a$ and $\Psi^c - a = -(\Psi - a)$. The

[60]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.

first one is the good root, $\Psi^G = \Psi$, and the second one is the bad root:

$$\Psi^B = 2a - \Psi \tag{141}$$

Now, because λ_p and S_p (i.e., the component of those vectors on the last coordinate, corresponding to p) are both 0, we have $b_p = 0$ (see 138) and thus $a_p = 0$. So the component of the bad root on the price is $\Psi_p^B = 2a_p - \Psi_p = -\Psi_p$:

$$\Psi_p^B = -\Psi_p \tag{142}$$

This allows to distinguish between the two roots, as the right one has $\Psi_p = \alpha M$ and the other one has $\Psi_p = -\alpha M$. Hence, if economic reasoning tells us the sign of αM (e.g., it is positive in a supply and demand context), we can pick the good root by inspecting the sign of Ψ_p.

D.10 When only some shocks are kept in the GIV

If we truncate the residuals, i.e. use

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

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. Also, this could be narratively checked. In our basic example of Section 2.1, we still have $\mathbb{E}\left[\left(p_t - \psi y_{St}\right) z_t\right] = 0$, so that the IV procedure still works.

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

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

where $z_t^< = \sum_i \tau^< \left(S_i \check{u}_{it}\right)$, using $\tau^< \left(x\right) = x 1_{|x| < b}$, so that $z_t \perp z_t^<$. Hence, regressing $u_{\Gamma t}$ on this truncated z_t gives a coefficient of 1, and all the analysis goes through.

D.11 Sporadic factors

A potential issue is that of a "sporadic factor", i.e. a factor η_t that affects a few actors special ways, but is not recurrent. An example would be a one-off policy announcement by the European Central Bank that they will buy both Italian and Spanish bonds, so that the truth is not that Italy is affecting Spain or vice-versa, but rather the ECB affecting both.

One solution, besides the narrative check, 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 $\mathcal{S}_t = b_{(2)t}$.[61] Over the entire sample, we might remove the days with anomalously high sporadicity statistics, e.g. in the top 5% by that metric.

D.12 PCA and IPCA

We can have time-varying factors. For instance, suppose that we have a vector of characteristics X_{it} (with 1 as a first component). Then, we could have $\lambda_{it} = X_{it}\dot{\lambda}$ for some $\dot{\lambda}$ to be estimated. We could also have a mixture, as in $\lambda_{it} = \left[X_t\dot{\lambda} : \check{\lambda}_i^{\text{Fixed}}\right]$, where $\check{\lambda}_i^{\text{Fixed}}$ is fixed while X_{it} is allowed to vary. We have $\dot{\lambda}$ has dimensions $k \times s$ with $k > s$, implying that we can model the loadings more flexibly as a function of a larger number of characteristics. Kelly et al. (2020) refer to this model as Instrumented Principal Components Analysis (IPCA) and they develop the asymptotic theory.

We can consider the case where we impose no structure on the loadings and we use PCA to estimate the loadings and the factor realizations. One can use the GMM procedure of Proposition 7, which works with finite N. The asymptotic theory for PCA (with $N \to \infty$) has been developed in Bai (2003) and in the context of GIV by Banafti and Lee (2022).[62] To compute standard errors, we can use the GMM values from Proposition 7; or a bootstrap. The NBER Working Paper of this version contains detailed guidance of this.

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

D.13 Full recovery when different factors have different "size" weights

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

We give some conditions under which we can actually also identify the λ^f (in addition to α^f and M). We show here that this is the case if we assume that the size S^f differs across all factors f, and this knowledge is given to us (from a model).

Here we take the basic set up as in Section D.5.1, in the simplified case where $\lambda_i^f = \lambda^f$ for all "endogenous" factors, i.e. for the factors f such that $\alpha^f \neq 0$, the other exogenous factors η all have

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

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

an impact of 1:

$$\begin{aligned} y_{it} &= u_{it} + \sum_f \lambda^f F_t^f + \eta_t^y, && (143) \\ F_t^f &= \alpha^f y_{S^f,t} + \eta_t^f. && (144) \end{aligned}$$

This implies

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

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

$$\begin{aligned} y_t &= \left(I - \iota \sum_f \lambda^f \alpha^f S^{f\prime} \right)^{-1} \left(u_t + \iota \varepsilon_t^1 \right) \\ &= \left(I + M\iota \sum_f \lambda^f \alpha^f S^{f\prime} \right) \left(u_t + \iota \varepsilon_t^1 \right) \\ y_t &= u_t + M\iota \sum_f \lambda^f \alpha^f u_{S^f,t} + \iota \varepsilon_t^y, && (145) \end{aligned}$$

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

$$F_t^f = \alpha^f \left(u_{S^f,t} + M \sum_g \lambda^g \alpha^g u_{S^g,t} \right) + \varepsilon_t^f. \tag{146}$$

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

$$z_{\Gamma^f t} := S^{f\prime} \check{u}_t = u_{S^f t} - u_{Et}. \tag{147}$$

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

$$F_t^f = \sum_g b_g^f z_{\Gamma^g t} + \varepsilon_t^{f,1} \tag{148}$$

(for ε^{f1} some residual noise) yields a regression coefficient:

$$b_g^f = \alpha^f \left(1_{f=g} + M \lambda^g \alpha^g \right). \tag{149}$$

This allows to recover everything, and with several overidentifying restrictions. Indeed,

$$b^f := \sum_g b_g^f = \alpha^f \left(1 + M \sum_g \lambda^g \alpha^g \right) = \alpha^f M,$$

which identifies $\alpha^f M$. Next, for $f \neq g$,

$$\frac{b_g^f}{b^f} = \lambda^g \alpha^g,$$

which gives $\lambda^g \alpha^g$ (and should be equal for all f), thus M. Hence, we obtained $\alpha^f M$, M and $\lambda^g \alpha^g$ — therefore all quantities: α^f, λ^f, M.

D.14 When we have disaggregated data for both the demand and the supply side

To estimate supply and demand elasticities, it is enough to have idiosyncratic shocks to one side of the market — demand in our basic example (Proposition 5). We complete our examination of supply and demand, with disaggregated data for both the demand and the supply side.

We posit that demand and supply disturbances follow:

$$y_{it}^k = \phi^k p_t + \lambda_i^k \eta_t^k + u_{it}^k, \tag{150}$$

for type $k = s, d$ for supply and demand. Total quantity demanded or supplied in side k of the market is (as a disturbance from the average), $y_{S^k t}^k := \sum_i S_i^k y_{it}^k$, where S_i^d (resp. S_i^s) is the average fraction of demand (resp. supply) accounted by country i) The price p_t adjusts so that supply equals demand, $y_{S^s t}^s = y_{S^d t}^d$, i.e.

$$p_t = \frac{u_{S^d}^d - u_{S^s}^s + \lambda_{S^d}^d \eta_t^d - \lambda_{S^s}^s \eta_t^s}{\phi_{S^s}^s - \phi_{S^d}^d} \tag{151}$$

So, the aggregate supply that was $s_t = \phi^s p_t + \varepsilon_t$ (see (200)) in the aggregated model is now

$$s_t = y_{S^s t}^s = \phi^d p_t + \lambda_{S^s}^s \eta_t^s + u_{S^s t}^s$$

so that the supply shock is

$$\varepsilon_t = \lambda_{S^s}^s \eta_t^s + u_{S^s t}^s.$$

We allow $\mathbb{E}\left[u_{it}^s u_{it}^d\right]$ to be nonzero: for instance, if the US has a "fracking shock" that affects both supply and demand, it will be captured by both u_{it}^s and u_{it}^d for $i =$ USA. Then, the initial exclusion restriction $\mathbb{E}\left[u_{it}\varepsilon_t\right] = 0$ (see (3)) fails. A fracking shock in the US both increases idiosyncratic US demand (as the US is richer) and also world supplies (as the US supplies more oil via its fracking technology).

But the situation is not so bleak. We make the following assumption

$$\mathbb{E}\left[u_{it}^k \eta_t^{k'}\right] = 0 \text{ for all } k, k' \in \{s, d\}. \tag{152}$$

For instance, when $k = d$ and $k' = s$, (152) means that idiosyncratic demand shocks are uncorrelated with the aggregate supply shocks η_t — once we control for idiosyncratic supply shocks (i.e. they may be correlated with ε_t but not with η_t^s). For simplicity, we discuss the homoskedastic case (where the $\left(u_{it}^d, u_{it}^s\right)$ are i.i.d. across i, t).

Then, we can still identify the elasticity of supply and demand. Indeed, we can form two GIVs, based on supply and demand respectively:

$$z_t^k := \Gamma^{k\prime} y_t^k = u_{\Gamma^k t}^k, \tag{153}$$

for $k = s, d$ (with $\Gamma^k = Q^{\lambda^k \prime} S^k$ in the general case and $\Gamma^k = S^k - E^k$ in the simple case $\lambda^k = \iota$, as in Corollary 2).

Then we have

$$\mathbb{E}\left[z_t^k \eta_t^{k'}\right] = 0 \text{ for all } k, k' \in \{s, d\}.$$

and one can easily see (as in the main paper) that the following identification moments hold, for $k, k' \in \{s, d\}$

$$\mathbb{E}\left[\left(y_{Et}^k - \phi^k p_t\right) z_t^{k'}\right] = 0. \tag{154}$$

So, we can estimate the demand and supply elasticities.[63'64].

Proposition 12 (Identification with disaggregated supply and demand data). *Suppose that we have disaggregated supply and demand data following* (150)*. Suppose that the shock are idiosyncratic in the sense of* (152)*, and that we have i.i.d.* $\left(u_{it}^d, u_{it}^s\right)$ *across* i, t*. Then, the GIVs* z_t^d*,* z_t^s *in* (153) *identify* ϕ^d *and* ϕ^s*, via moments* (154)*.*

In summary, in that example, the GIV fails if aggregate shocks (ε_t) are importantly made of idiosyncratic shocks. However, in the same example, having more disaggregated data (on both the demand and supply side), together with a slightly different exclusion restriction, allow estimation of both elasticities by GIV.

D.15 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. (1995) in the next subsection.[65]

[63]The optimal instrument is $z_t = z_t^d - z_t^s$, as this is the most correlated with the price (151) (this generalizes the reasoning of Proposition 3).

[64]One could imagine variants. For instance, if we assume only that $\mathbb{E}\left[z_t^\ell \eta_t^k\right] = 0$ for a given (k, ℓ), we can identify ϕ^k via $\mathbb{E}\left[\left(y_{Et}^k - \phi^k p_t\right) z_t^\ell\right] = 0$.

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

D.15.1 Logit demand

The utility that household h derives from product i, for $i = 0, ..., N$, is given by[66]

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

where e_{hit} follows a Type-1 extreme-value distribution, p_{it} denotes the log price, x_{it} observable characteristics, and $\mathbb{E}\left[\xi_{it}\right] = 0$. We refer to $i = 0$ as the outside option and normalize $\delta_{0t} = 0$. This model implies that the market share s_{it} is the probability that a given household selects product i, meaning that $s_{it} = \mathbb{P}\left(U_{hit} > \max_{j \neq i} U_{hjt}\right)$, and can be expressed as

$$s_{it} = \frac{\exp(\delta_{it})}{\sum_{j=0}^{N} \exp(\delta_{jt})}.$$

Firms set prices to maximize profits and we assume that each product is produced by a single firm, which solves

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

where C_{it} equals marginal cost and $Q_{it} = s_{it} Q_t$ with Q_t the total size of the market. The firm optimally sets the price to

$$P_{it} = \left(1 - \frac{1}{\epsilon_{it}}\right)^{-1} C_{it},$$

where $\epsilon_{it} = -\frac{\partial \ln s_{it}}{\partial p_{it}}$, that is, the negative of the price elasticity of demand. The goal is to estimate $\theta = (\beta, \gamma)$.

It is convenient to rewrite the model as

$$\log\left(\frac{s_{it}}{s_{0t}}\right) = -\gamma p_{it} + \beta' x_{it} + \alpha_i + \xi_{it}.$$

To identify β, it is commonly assumed that $\mathbb{E}\left[x_{it}\xi_{it}\right] = 0$ and we maintain this assumption. However, as prices respond to demand shocks, ξ_{it}, we cannot assume $\mathbb{E}\left[p_{it}\xi_{it}\right] = 0$. There are three common approaches to create instrumental variables in the demand estimation literature. First, variables that capture variation in marginal cost, C_{it}, that is unrelated to demand shocks. Second, Berry et al. (1995) 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

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

therein).[67] The resulting moment is $\mathbb{E}\left[z_{it}^{BLP}\xi_{it}\right] = 0$.[68] Third, one can use panel data for the same firm that operates in different locations. Under the assumption that demand shocks are uncorrelated across locations, prices in other locations of the same firm will be valid instruments. The intuition is that prices across locations share the same marginal cost but the demand shocks are, by assumption, uncorrelated, see Nevo (2001).

GIV provides an alternative by exploiting exogenous variation in markups due to idiosyncratic demand shocks to large firms. We assume that demand shocks follow a factor model,

$$\xi_{it} = \eta_t + u_{it}, \tag{155}$$

which can be extended to allow for heterogeneous exposures, i.e. replacing η_t by $\lambda_i \eta_t = \sum_k \lambda_i^k \eta_t^k$. Also, we assume for simplicity that η_t and u_{it} are i.i.d. over time, but the logic in this section can be extended to persistent demand shocks (see also Sweeting (2013)).

We propose to use the GIV instrument as the weighted sum of idiosyncratic demand shocks of the competitors:

$$z_{it} = \sum_{j:j\neq i} \bar{s}_{j,t-1} u_{jt}, \tag{156}$$

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

$$\mathbb{E}\left[z_{it}\xi_{it}\right] = 0, \tag{157}$$

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

The intuition for why z_{it} is a meaningful instrument is the following: if there is a high idiosyncratic shock for Tesla cars (high u_{jt}, with j being Tesla), this leads Ford (firm i) to reduce the price of its cars (in this particular model, this is because the positive shock for Tesla cars reduces the demand for Ford, which sees its market share s_{it} fall, so that it wants to lower its price p_{it}).

Generalizing this intuition, we sum over all the demand shocks of the competitors, $z_{it} = \sum_{j:j\neq i} \bar{s}_{j,t-1} u_{jt}$, weighing them by size, i.e. market share. As in our general GIV, even a single shock u_{jt} is a valid instrument (for $j \neq i$). The size-weighted sum is simply a typically useful way to pool those idiosyncratic shocks. It is optimal in our basic GIV, and is likely to be reasonably close to optimal in this IO context. The same idea generalizes: e.g. using a weighted sum of the idiosyncratic cost shocks, rather than demand shocks, of the competitors would also be a valid GIV instrument.

A motivation for the weighting in (156) is as follows. Recall that in this simple model the demand elasticity is

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

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

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

and also that $\frac{\partial \log s_{it}}{\partial \delta_{jt}} = -s_{jt}$, so that $\frac{\partial \log s_{it}}{\partial u_{jt}} = -s_{jt}$ (controlling for the price p_{jt}). This implies that the direct impact of all idiosyncratic demand shocks to other companies on s_{it}, and hence ϵ_{it}, is

$$\sum_{j:j\neq i} \frac{\partial \log s_{it}}{\partial u_{jt}} u_{jt} = -\sum_{j:j\neq i} s_{jt} u_{jt}. \tag{158}$$

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

D.15.2 Random coefficients logit as in BLP

Berry, Levinsohn and Pakes (1995) extend the standard logit model by allowing for random variation in the preference parameters

$$\theta_h = \theta + \nu_h,$$

where $\nu_h = \left(\nu_h^{\beta}, \nu_h^{\gamma}\right)$ and $\nu_h \sim F_\nu(\nu; \Theta)$, for some vector of parameters Θ. The market share equation modifies to

$$s_{it} = \int_\nu s_{hit} dF_\nu(\nu; \Theta),$$

where

$$s_{hit} = \frac{\exp\left(\delta_{it} - \nu_h^{\gamma} p_{it} + \nu_h^{\beta\prime} x_{it}\right)}{\sum_{j=0}^{N} \exp\left(\delta_{jt} - \nu_h^{\gamma} p_{jt} + \nu_h^{\beta\prime} x_{jt}\right)}.$$

To estimate the model, Berry (1994) suggests to recover δ_{it} from the market shares using a contraction mapping (see Nevo (2000) for an introduction). With δ_{it} in hand, we form moment conditions as before to estimate (θ, Θ).

To construct a GIV instrument in this model, one can also use (156) as an instrument.

One can also refine it. For instance, we can recompute the total impact of idiosyncratic shocks to other firms on the demand elasticity, which is now slightly more involved. The negative of the demand elasticity, which enters into the pricing equation via the markup, is given by

$$\epsilon_{it} = \int_\nu \nu_h^{\gamma} \frac{s_{hit}}{s_{it}} (1 - s_{hit})\, dF_\nu(\nu; \Theta).$$

An approximation of the model around $\theta_h = \theta$ yields the same weights as before, although it is feasible to numerically calculate the optimal weights by computing

$$\sum_{j,j\neq i} \frac{\partial \epsilon_{it}}{\partial u_{jt}} u_{jt}.$$

This suggests forming

$$z_{it} := \sum_{j:j\neq i} s_{j,t-1}^{i} u_{jt}, \tag{159}$$

where $s^i_{j,t}$ is

$$s^i_{j,t} := -\frac{\partial \log s_{it}}{\partial u_{jt}}. \tag{160}$$

Indeed, in the homogeneous elasticity case, $s^i_{j,t} = s_{jt}$. This generalization to heterogeneous elasticity allows to capture that if firms i and j tend to serve the same consumers (e.g., both sell family cars), then the $s^i_{j,t}$ will be high, and u_{jt} receives a high weight in the firm-i specific GIV z_{it}.

D.16 Dealing with fat tails

Here we provide a justification of the procedure in Appendix H.2 to dampen the influence of outliers. For clarity, we consider the following problem first — our main problem is just a more complex variant. Suppose that we want to estimate β in a regression:

$$y_i = \beta x_i + u_i \tag{161}$$

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

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

$$y_i^W := sign(y_i) \min(|y_i|, \delta), \tag{162}$$

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

$$y_i^W := y_i + r(y_i) \tag{163}$$

with

$$r(u) := r_\delta(u) = -\max(|u| - \delta, 0)\,\mathrm{sign}(u). \tag{164}$$

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

$$\mathbb{E}\left[\left(y_i^W - \beta x_i\right) x_i\right] \neq 0 \tag{165}$$

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

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

$$y_i^w := y_i + r(y_i - \beta x_i) \tag{166}$$

instead of the traditional (163), and then to run OLS of $y_i^w = \beta x_i + \varepsilon_i$.

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

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

2. Run the OLS of
$$y_i^{w,(n)} = \beta x_i + \varepsilon_i. \tag{167}$$

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

We next justify this. Define $L(u) = -\ln p(u)$, the log likelihood of y is $\sum_i L(y_i - \beta x_i)$, so the maximum likelihood estimator is
$$\min_\beta \sum_i L(y_i - \beta x_i) \tag{168}$$
whose first order condition is
$$\sum_i L'(y_i - \beta x_i)\, x_i = 0 \tag{169}$$

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

$$L'(u) = k\,(u + r(u)) \tag{170}$$

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

$$L^{\text{Huber}}(u) = \frac{u^2}{2} 1_{|x| \le \delta} + \left(|u| - \frac{\delta}{2}\right) \delta 1_{|u| > \delta} \tag{171}$$

then we have $L^{\text{Huber}\prime}(u) = u + r(u)$ with $r(u)$ exactly as in (164). This is why we take this value of $r(u)$ in practice. But we continue the discussion for a general $r(u)$.

Then, the FOC (169) becomes

$$\sum_i \left(y_i - \beta x_i + r(y_i - \beta x_i)\right) x_i = 0 \tag{172}$$

To get more intuition, we define $y_i^w := y_i + r(y_i - \beta x_i)$ as in (166), which has the interpretation of sort of "winsorized" y_i, hence the w superscript. Then FOC is

$$\sum_i (y_i^w - \beta x_i)\, x_i = 0 \tag{173}$$

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

We next state a simple proposition.

Proposition 13 *Suppose that $\mathbb{E}\left[r\left(u_i\right)x_i\right]=0$, for instance, because u_i and x_i are independent, and $\mathbb{E}\left[r\left(u_i\right)\right]=0$. Then, at the true value β we have*

$$\mathbb{E}\left[\left(y_i^w-\beta x_i\right)x_i\right]=0 \tag{174}$$

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

The proof is almost a tautology: the statement is equivalent to saying that $\mathbb{E}\left[r\left(u_i\right)x_i\right]=0$, which is which exactly the main assumption of the proposition. But the advantage is that it lays out a simple procedure to "winsorize" outliers: run the OLS (167), $y_i^w=\beta x_i+\varepsilon_i$. If u_i is e.g. non-symmetric, it shows a simple criterion for other residual functions, $\mathbb{E}\left[r\left(u\right)\right]=0$, e.g. by choosing a function $r\left(u\right)$ that is non-symmetric.

Link with the procedure in Appendix H.2. With the more complex factor model of Appendix H.2, the arguments are exactly the same. We can state the following proposition, and prove it exactly the same way. Hence moment conditions used in Appendix H.2 are valid.

Proposition 14 *Suppose that $\mathbb{E}\left[r\left(u_i\right)\left(\lambda_i,a_i,\eta_t,b_t\right)\right]=0$. Then at the true values, the following moments hold*

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

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

D.17 When aggregate shocks are made of idiosyncratic shocks

GIVs extend to economies where aggregate shocks η_t are themselves made of idiosyncratic shocks u_{it}. We summarize the situation here.

Take the basic supply and demand model of Section 2.1. We achieved identification provided that $u_{\Gamma t}\perp\varepsilon_t$; we did not need $u_{\Gamma t}\perp\eta_t$, so aggregate demand shocks can be influenced by idiosyncratic shocks, but not aggregate supply shocks. If aggregate supply shocks are affected by idiosyncratic shocks, the elementary strategy does not work, but a variant does work, with a slightly different identification assumption. We suppose disaggregated supply and demand data (for the commodity in question, e.g. oil) is available, at least for large countries. We model country i's supply and demand with the following factor model:

$$y_{it}^k=\phi^k p_t+\lambda_i^k\eta_t^k+u_{it}^k, \tag{175}$$

where $k=s,d$ indicates supply or demand, respectively. We allow $\mathbb{E}\left[u_{it}^s u_{it}^d\right]$ to be nonzero: for instance, if the US has a positive "fracking shock" that affects both supply and demand, it will be captured by a positive u_{it}^s and u_{it}^d for $i=$ USA. This is a concrete case in which supply and demand shocks are correlated: this happens via the correlations in country-level shocks. At the same time, we impose that the u_{it}^k are uncorrelated with the aggregate shocks $\eta_t^{k'}$ for $k, k' \in \{s, d\}$. Then, Section D.14 shows how to identify the elasticities of supply and demand.

One can also consider an economy as a network (Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019)). Under some assumptions, one can obviate the network structure, for instance via aggregation theorems such as Hulten's theorem. This is developed in Section D.18. 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, which are developing in another paper.

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.

D.18 Identification of the TFP to GDP multiplier in a production network economy

Suppose a two-period model with a production network, as in Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019). There are both idiosyncratic TFP shocks $\hat{\lambda}_{it}$ and a government reform that creates correlated shocks η_t to TFP and change in labor supply $\hat{L}_t$. Utility is $C_t - e^{\eta_t^L} L_t^{1+1/\phi}$, so that ϕ is the Frisch elasticity of labor supply. We call λ_t total TFP, which depends on the industry TFPs λ_{it}. So, as $C_t = \lambda_t L_t$, labor supply is $\hat{L}_t = \phi \left(\hat{\lambda}_t - \eta_t^L \right)$,[69] and GDP is $\hat{Y}_t = \hat{L}_t + \hat{\lambda}_t$, i.e.

$$\hat{Y}_t = m\hat{\lambda}_t - \phi \eta_t^L, \qquad m = 1 + \phi \tag{176}$$

We seek to find the "GDP multiplier" $m = 1 + \phi$, so that a TFP increase of 1 percent translates into a GDP increase of m percent.[70]

This is potentially a complicated problem, as for instance, in the Long and Plosser (1983) case with input-output matrix A, output changes are $\hat{Y}_t = (I - A)^{-1} \hat{\lambda}_t + \hat{L}_t$, so that output changes are correlated in complicated ways. However, one can sidestep using this disaggregated production data. We assume that TFP change in industry i is:

$$\hat{\lambda}_{it} = \lambda_i \eta_t^\lambda + u_{it}. \tag{177}$$

In the neoclassical equilibrium, TFP follows Hulten's theorem, so is $\hat{\lambda}_t = \sum_i s_i \hat{\lambda}_{it}$ where s_i is the Domar weight (sales of industry i over GDP).

We can identify the multiplier m if we have disaggregated TFP data. In the simplest case, we

[69] 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.

[70] If more than one factor changes, m has the broader interpretation of a multiplier between TFP and GDP.

assume that industry-level productivities are available, and we get the residuals u_{it}^e. Then, we can identify the multiplier m by GIV.

We can identify the multiplier m if we have even crude proxies for disaggregated TFP. The same procedure works (with less efficiency) if our data is made of proxies for productivity growth $\tilde{\hat{\lambda}}_{it}$ (where the tilde indicates that we deal with a proxy). An example could be growth of sales per employee, or even the growth rate of sales. We assume a factor model

$$\tilde{\hat{\lambda}}_{it} = \tilde{\lambda}_i \tilde{\eta}_t^\lambda + \tilde{u}_{it}. \tag{178}$$

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.19 Identification of the elasticity of substitution between capital and labor / Elasticity of demand in partially segmented labor markets

Here we show how GIVs can estimate the elasticity of substitution between capital and labor; and how to estimate the elasticity of demand in partially segmented markets. The first problem uses the second one.

As a motivation, imagine that industry i has the CES production function[71]

$$Q_{it} = B_{it} \left(K_{it}^{\frac{\phi_i - 1}{\phi_i}} + A_{it}^{\frac{1}{\phi_i}} L_{it}^{\frac{\phi_i - 1}{\phi_i}} \right)^{\frac{\phi_i}{\phi_i - 1}} \tag{179}$$

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:

$$\frac{L_{it}}{K_{it}} = A_{it} \left(\frac{W_{it}}{R_t}\right)^{-\phi_i} \tag{180}$$

We'd like to estimate the elasticity of substitution ϕ_i between capital and labor. This is the wage elasticity of demand. GIVs allow to estimate that, as we shall see.

Let us use our general notations, and define $y_{it}^d = \ln L_{it}$, $p_{it} = \ln W_{it}$, $C_{it} = \ln K_{it}$, and $\phi_i^d = -\phi_i$

[71]We thank Julieta Caunedo for prompting us to think about this identification problem.

(as this is the elasticity of labor demand). Then, we can write (180) as:

$$y_{it}^{d}=\phi_{i}^{d}p_{it}+C_{it}+\lambda_{i}^{d}\eta_{t}+u_{it}^{d} \tag{181}$$

where C_{it} is a control, and as usual vector η_t is a common shock, and u_{it}^{d} is a demand shock (those in turn come from the productivity A_{it}). For notational simplicity we will drop C_{it}, but this is not important.

Now, log labor supply is modeled as:

$$y_{it}^{s}=\phi_{i}^{s}p_{it}-\psi_{i}p_{St}+\lambda_{i}^{s}\eta_{t}+u_{it}^{s} \tag{182}$$

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:

$$p_{it}=\frac{\psi_{i}p_{St}+u_{it}^{d}-u_{it}^{s}+\left(\lambda_{i}^{d}-\lambda_{i}^{s}\right)\eta_{t}}{\phi_{i}^{s}-\phi_{i}^{d}} \tag{183}$$

i.e.

$$p_{it}=\gamma_{i}p_{St}+v_{it}+\lambda_{i}^{p}\eta_{it} \tag{184}$$

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 (184) is a standard GIV. In the general case, we can estimate γ_i as in Section (D.9).[72]

So, we obtain γ_i and v_{it}^{e} (the proxy for v_{it}) in (184). We also form $z_{it}=S'\left(v_t^e-a^i v_{it}^e\right)$ as in (121). This is the GIV formed of the idiosyncratic shock of all industries but industry i. We will use the shock to those other industries, and their impact on the outside wage, as an instrument to estimate labor demand. Indeed, we go back to the labor demand equation (181), and instrument for p_{it} using the z_{it}

$$p_{it}=b_{i}z_{it}+\varepsilon_{it}^{p} \tag{185}$$

we estimate b_i, and define $p_{it}^{e}=b_i z_{it}$ as the price in industry i instrumented by the changes in other industries. We use the estimated η_t^e as controls, and run

$$y_{it}^{d}=\phi_{i}^{d}p_{it}^{e}+C_{it}+\lambda_{i}^{d}\eta_{t}^{e}+u_{it}^{d} \tag{186}$$

which yields a consistent estimate ϕ_i^d of the labor demand.

[72] This procedure is much simplified if the γ_i and λ_i^p are assumed to be constant. Then, we can just define $z_t:=p_{\Gamma t}=p_{St}-p_{Et}$, so that $z_t=v_{\Gamma t}$ and as $p_{St}=\frac{v_{St}+\lambda^p\eta_t^s}{1-\gamma}$, regressing $p_{St}=bz_t+\varepsilon_t^p$ yields $b=\frac{1}{1-\gamma}$.

References for Online Appendix

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

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

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

Arellano, Manuel and Stephen Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 1991, *58* (2), 277–297.

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

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

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

Banafti, Saman and Tae-Hwy Lee, "Inferential theory for granular instrumental variables in high dimensions," *arXiv preprint arXiv:2201.06605*, 2022.

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, *63* (4), 841–890.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 1998, *87* (1), 115–143.

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

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

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

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

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

Hansen, Lars, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–54.

Heckman, James, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–59.

Huber, Peter J., "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 1964, *35* (1), 73–101.

Kelejian, Harry H and Ingmar R Prucha, "A generalized moments estimator for the autoregressive parameter in a spatial model," *International Economic Review*, 1999, *40* (2), 509–533.

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

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

Lewbel, Arthur, "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models," *Journal of Business & Economic Statistics*, 2012, *30* (1), 67–80.

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

MacKay, Alexander and Nathan Miller, "Estimating models of supply and demand: Instruments and covariance restrictions," *Available at SSRN 3025845*, 2019.

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

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

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

Sarto, Andres, "Recovering macro elasticities from regional data," *Working Paper*, 2022.

Sun, Qiang, Wen-Xin Zhou, and Jianqing Fan, "Adaptive huber regression," *Journal of the American Statistical Association*, 2020, *115* (529), 254–265.

Sweeting, Andrew, "Dynamic Product Positioning in Differentiated Product Markets: The Effects of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, 2013, *81* (5), 1763–1803.