Granular Instrumental Variables*

Xavier Gabaix[1] and Ralph S.J. Koijen[2]

[1]Harvard University

[2]University of Chicago

October 19, 2023

Abstract

We develop 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. Idiosyncratic shocks to these large players significantly affect aggregate outcomes and are valid instruments. We provide a methodology to extract these idiosyncratic shocks to create "granular instrumental variables" (GIVs), which are size-weighted sums of idiosyncratic shocks. We show how GIVs enable us to estimate causal parameters, such as elasticities and multipliers, and are applicable in a broad range of empirically relevant settings.

* 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, Adam Guren, 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). Edited by Harald Uhlig.

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.

[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).

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 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. Baumeister and Hamilton (2023) use a GIV to estimate supply and demand elasticities and dynamics in the world oil market, using idiosyncratic supply shocks (for instance, the production drop of Kuwait after it was invaded by Iraq in 1990) as granular instruments. We developed the GIV while working on Gabaix and Koijen (2022), where we use it to measure the elasticity of demand for 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 $X = (X_i)_{i=1\dots N}$ 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, $\iota = (1,\dots,1)'$ 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 random variables a_t and b_t are uncorrelated with random variables 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

[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.

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 $N \times r$ 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 $V^u = \mathbb{E}[u_t u_t']$ 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 $\Gamma' S \neq 0$:

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

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

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

[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, $C_t^y = 0$.

[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.

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).

[7] See (57) for a proof. The heteroskedastic case will lead to (54).

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

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

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

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

We provide further economic intuition for the asymptotic variances of the GIV estimators in Section 3.1. The condition $\mathbb{E}[u_{St}u_{\Gamma t}] \neq 0$ is generically true. For instance, with homoskedastic u_{it}, $\mathbb{E}[u_{St}u_{\Gamma t}] = \sigma_u^2 \Gamma' S$, 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_ψ^2 and $\sigma_{\phi^d}^2$ 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_ψ^2 *and* $\sigma_{\phi^d}^2$ *of the GIV estimators in Proposition 2, with* $\frac{\mathbb{E}\left[u_{\Gamma t}^2\right]}{\mathbb{E}\left[u_{St}u_{\Gamma t}\right]^2} = \frac{1}{\mathbb{E}\left[u_{\Gamma^* t}^2\right]}$. *For any other* Γ *that is not proportional to* Γ^*, *the asymptotic variances* $\sigma_\psi^2(\Gamma)$ *and* $\sigma_{\phi^d}^2(\Gamma)$ *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

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

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 $\check{u}_{it}$.

[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.

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:*

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

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

[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 $y_t - C^y_t m^y$.

[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).

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:

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

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

where $c_t = (C_t^p, C_{St}^y, \check{\eta}_t)$ is a vector of controls, and $(\varepsilon_t^p, \varepsilon_t^y)$ 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.

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

$$p_t = \mu z_t + b^p c_t + e_t^p, \tag{36}$$

$$y_{St} = M z_t + b^y c_t + e_t^y, \tag{37}$$

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 $\check{\eta}_t(m^y) := R^{\check{\lambda}}(y_t - C_t^y m^y)$*

[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 $\check{\eta}_t$.

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 $c_t = \left(C_t^p, C_{St}^y, \check{\eta}_t\left(m^y\right)\right)$ 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 $\check{\eta}_t\left(m^y\right)$ 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).

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

[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 $o_p\left(T^{-1/2}\right)$.

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, $y_{Et} = \frac{1}{N}\sum_{i=1}^N y_{it}$:

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

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 $\mathbb{E}\left[u_t \varepsilon_t\right] = 0$, and hence $\mathbb{E}\left[z_t \varepsilon_t\right] = 0$. This gives (using (38)): $\mathbb{E}\left[\left(p_t - \psi y_{St}\right) z_t\right] = 0$ 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 $\mathbb{E}\left[z_t \eta_t\right] = 0$. In addition, under the Assumption 3 that the u_{it} are homoskedastic we have $\mathbb{E}\left[u_{Et} u_{\Gamma t}\right] = 0$.[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

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

elasticity of demand of the other countries, and the elasticity of supply, which is equal (as supply 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 "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 $\psi \neq 0$)*

[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.

$\sqrt{T}\left(\phi_T^{d,e}-\phi^d\right)\xrightarrow{d}N\left(0,\sigma^2_{\phi^d}\right)$, *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):*

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

where h is the excess Herfindahl:

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

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.

[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.

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 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 (35) 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 $\mathbb{E}\left[\left(p_t - \psi y_{St}\right) z_t\right] = 0$, 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 the equal-weighted averages of coefficients when N is large. 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 (see Section D.8).

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, we will not completely eliminate the error term $\check{\lambda}_\Gamma \check{\eta}_t - \check{\lambda}_\Gamma^e \check{\eta}_t^e$.[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 $\check{\lambda}_\Gamma = 0$

[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 (for large T) $\mu^e = \mu\xi$ 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.

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

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

[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_t^e that are usable when running regressions, which increases the precision of estimators.

$z_t^{(2)} = S^{(2)\prime}\check{u}_t$ 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).

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 (η_t', ε_t) 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.

[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.

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* $R^x := (x'x)^{-1} x'$, $\dot{y}_t := R^x (y_t - C_t^y m^y)$ *(which has dimension k),* $\check{u}_t := Q^\lambda (y_t - C_t m^y)$, *and* $z_t := \Gamma' \check{u}_t$, *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.*

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

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 $y_{it} - C_{it}^y m^y$ on x_i and $\check{\lambda}_i$, for every date t (and given m^y):

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

[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.

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

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:*

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

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

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 $\sqrt{T}\left(\theta^e_T - \theta_0\right) \to N\left(0, V^\theta\right)$ for a matrix V^θ.

[27] Banafti and Lee (2022) extend the results in this paper by studying 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.

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,

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

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 $W = (V^u)^{-1}$. 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 $\sum_t \left(-\check{y}_t + \check{\lambda}\check{\eta}_t\left(m^y, \check{\lambda}\right) + \check{C}_t^y m^y\right) W \check{\eta}_t\left(m^y, \check{\lambda}\right)' = 0$.*

[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):

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

When idiosyncratic shocks are homoskedastic, $\tilde{E}_i = \frac{1}{N}$, while if they are uncorrelated but heteroskedastic (i.e. $W = Diag\left(1/\sigma_{u_i}^2\right)$), we have[32]

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

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

[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 $\tilde{E} \coloneqq S - \Gamma$.

[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.

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 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

[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.

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

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.

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.

[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).

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 $(u_i)_{i=1\ldots N}$ 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

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

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

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

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

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, $\mathbb{E}_T[Y_t] := \frac{1}{T}\sum_{t=1}^T Y_t$. 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:

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

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

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

The problem is invariant to changing Γ into $k\Gamma$ 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

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

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 $Q = I - \lambda\left(\lambda'\lambda\right)^{-1}\lambda'$ (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), $R \coloneqq (\lambda' W \lambda)^{-1} \lambda' W$ and $Q \coloneqq I - \lambda R$, 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* $\mathbb{E}[\check{\eta}_t \check{\eta}_t']$, $\mathbb{E}\left[\left(Q^\lambda C_t^y\right)' \left(Q^\lambda C_t^y\right)\right]$, $\mathbb{E}\left[C_{St}^{y\prime} C_{St}'\right]$ *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

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

We again use $\mathbb{E}_T$ for the sample temporal mean, $\mathbb{E}_T[Y_t] \coloneqq \frac{1}{T} \sum_{t=1}^T Y_t$. 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, $\mathbb{E}[g_t(\theta_0)] = 0$**.** Indeed, $g_t^1(\theta_0) = -\check{u}_t' \check{C}_t^y$, as spelled out in the Online Appendix, equation (73). This implies that $\mathbb{E}[g_t^1(\theta_0)] = 0$, by (3), and the fact that $\check{u}_t = Q^\lambda u_t$. Likewise, $g_t^2(\theta_0) = -\varepsilon_t^\perp \left(z_t, \check{\eta}_t', C_t^{p\prime}\right)$, so $\mathbb{E}[g_t^2(\theta_0)] = 0$, 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 $\check{\eta}_t \coloneqq \eta_t + R^\lambda u_t \perp C_t^{p\prime}, \check{u}_t, z_t$ 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 $\mathbb{E}[g_t^3(\theta_0)] = 0$.

The solution of $\mathbb{E}[g_t(\theta)] = 0$ **is unique.** The high-level idea is that the equation $\mathbb{E}[g_t^1(\theta)] = 0$ is linear equation in m^y, and identifies m^y given our full-rank Assumption 5. Then, given m^y, the other equations ($\mathbb{E}\left[g_t^k(\theta)\right] = 0$ 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 $\mathbb{E}_T[g_t(\theta)] = 0$.[42] Note that θ_T^e is equivalent to the GMM estimator that minimizes $\mathbb{E}_T[g_t(\theta)]' \mathbb{E}_T[g_t(\theta)]$, as indeed is achieves $\mathbb{E}_T[g_t(\theta)] = 0$. 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 $\sqrt{T}$–asymptotically normal. 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, $\sqrt{T}(\theta_T^e - \theta)$ converges in distribution to a normal distribution with mean 0 and positive variance covariance matrix V^θ.

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

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

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

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

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

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

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

and thus

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

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:

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

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]

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

[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.

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.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Adao, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics 134*(4), 1949–2010.

Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022). The term structure of growth-at-risk. *American Economic Journal: Macroeconomics 14*(3), 283–323.

Amiti, M., O. Itskhoki, and J. Konings (2019). International shocks, variable markups, and domestic prices. *The Review of Economic Studies 86*(6), 2356–2402.

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

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

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

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*(6), 2121–68.

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

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

Banafti, S. and T.-H. Lee (2022). Inferential theory for granular instrumental variables in high dimensions. *arXiv preprint arXiv:2201.06605*.

Baqaee, D. R. and E. Farhi (2019). The macroeconomic impact of microeconomic shocks: beyond hulten's theorem. *Econometrica 87*(4), 1155–1203.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

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

Baumeister, C. and J. D. Hamilton (2023). A full-information approach to granular instrumental variables. *Working Paper, UCSD*.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies 89*(1), 181–213.

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

Caldara, D., M. Cavallo, and M. Iacoviello (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics 103*, 1–20.

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

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

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

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

Chodorow-Reich, G., A. Ghent, and V. Haddad (2021). Asset insulators. *The Review of Financial Studies 34*(3), 1509–1539.

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

di Giovanni, J., A. A. Levchenko, and I. Méjean (2014). Firms, destinations, and aggregate fluctuations. *Econometrica 82*(4), 1303–1340.

Dong, X., N. Kang, and J. Peress (2022). Fast and slow arbitrage: The predictive power of capital flows for factor returns. *Available at SSRN 3675163*.

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

Flynn, J. P. and K. Sastry (2022). The macroeconomics of narratives. *Available at SSRN 4140751*.

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

Gabaix, X. and R. S. Koijen (2022). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *NBER Working Paper No. 28967*.

Galaasen, S., R. Jamilov, R. Juelsrud, H. Rey, et al. (2021). Granular credit risk. *NBER Working Paper No. 27994*.

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

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–54.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–59.

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

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics 35*(1), 73–101.

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

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 509–533.

Kelly, B. T., S. Pruitt, and Y. Su (2020). Instrumented principal component analysis. *Available at SSRN 2983919*.

Kundu, S. and N. Vats (2021). Banking networks and economic growth: From idiosyncratic shocks to aggregate fluctuations. *Available at SSRN 3556299*.

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

Lee, Y., L. A. N. Amaral, D. Canning, M. Meyer, and H. E. Stanley (1998). Universal features in the growth dynamics of complex organizations. *Physical Review Letters 81*(15), 3275.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1), 67–80.

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

Ludvigson, S. C., S. Ma, and S. Ng (2020). Shock restricted structural vector-autoregressions. *NBER Working Paper No. 23225*.

Ma, Y., T. Paligorova, and J.-L. Peydro (2022). Expectations and bank lending. *Working Paper*.

MacKay, A. and N. Miller (2019). Estimating models of supply and demand: Instruments and covariance restrictions. *Available at SSRN 3025845*.

Montiel Olea, J. L. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Nakamura, E. and J. Steinsson (2018, August). Identification in macroeconomics. *Journal of Economic Perspectives 32*(3), 59–86.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

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

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

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics 2*, 71–162.

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

Sarto, A. (2022). Recovering macro elasticities from regional data. *Working Paper*.

Schubert, G., A. Stansbury, and B. Taska (2022). Employer concentration and outside options. *Available at SSRN 3599454*.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics 2*, 415–525.

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

Sun, Q., W.-X. Zhou, and J. Fan (2020). Adaptive huber regression. *Journal of the American Statistical Association 115*(529), 254–265.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effects of fees for musical performance rights on the commercial radio industry. *Econometrica 81*(5), 1763–1803.

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