Online Appendix for

"Granular Instrumental Variables"

Xavier Gabaix and Ralph S.J. Koijen

October 19, 2023

Table of contents

B Appendix: Notations

d, s : Indicates demand and supply. E.g., $\phi^d, \phi^s = \frac{1}{\psi}$ are the elasticities of demand and supply.

ε_t, η_t: Aggregate shocks.

C_t^x: Controls affecting variable x.

λ: Factor loadings.

u_t: Idiosyncratic shocks.

$\check{u}_t \coloneqq Qu_t$: Idiosyncratic shocks residualized by a projection matrix Q.

$z_t = \Gamma' u_t$: GIV.

C Additional Proofs

C.1 Complement to the proof of Proposition 2

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

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

where

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

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

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

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

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

C.2 Complement to the proof of Proposition 4

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

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

which are easy to verify. We have

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

so

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

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

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

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

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

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

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

Derivation of the asymptotic variance $\sigma_{\phi^d}^2$. The argument is exactly the same as for ψ. We zoom on $g^{3,z}$:

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

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

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

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

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

Indeed, $\mathbb{E}\left[\varepsilon_t^y \Gamma' C_t^y\right] = 0$ as $\varepsilon_t^y = \eta_{1t}^{\perp} + u_{\tilde{E}t}$, with $\eta_{1t}^{\perp}$ and $u_{\tilde{E}t}$ uncorrelated with C_t^y.

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

$$\sigma_{\phi^d}^2 = \frac{var\left(g_t^{3,z}\right)}{\left(\mathbb{E}\left[g_{\phi^d}^{3,z}\right]\right)^2} \tag{76}$$

Given $g_t^{3,z} = -\varepsilon_t^y z_t$,

$$var\left(g_t^{3,z}\right) = \sigma_{\varepsilon^y}^2 \sigma_z^2,$$

and given (75), and that ε_t^y is homoskedastic conditional on u_t,

$$\sigma_{\phi^d}^2 = \frac{\mathbb{E}\left[z_t^2\right]}{\mu^2 \mathbb{E}\left[u_{St} z_t\right]^2} var\left(\varepsilon_t^y\right).$$

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

C.3 Proof of Proposition 5

We use the controls $c_t(m^y) = (C_t^p, C_{St}^y, \check{\eta}_t(m^y))$. We define the function $g_t = \left(g_t^k\right)_{k=1\ldots3}$ by:

$$g_t^1(\theta) \coloneqq \left(-\check{y}_t + \check{\lambda}\check{\eta}_t(m^y) + \check{C}_t^y m^y\right)' W \check{C}_t^y, \tag{77}$$

$$g_t^2(\theta) \coloneqq \left(-p_t + \mu z_t(m^y) + b^p c_t(m^y)\right)' \left(z_t(m^y), c_t(m^y)\right), \tag{78}$$

$$g_t^3(\theta) \coloneqq \left(-y_{St} + M z_t(m^y) + b^y c_t(m^y)\right)' \left(z_t(m^y), c_t(m^y)\right). \tag{79}$$

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

Asymptotic variance of M_T^e. We use the moment function

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

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

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

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

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

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

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

C.4 Proof of Corollary 2

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

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

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

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

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

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

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

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

where $h^2 = S'S - \frac{1}{N}$ and $x_S = S'X$. So $\frac{\sigma^2_{u\Gamma}}{\sigma^2_u} = S'QS = h^2 - \frac{x_S^2}{N\sigma_x^2}$.

C.5 Proof of Proposition 6

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

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

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

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

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

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

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

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

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

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

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

Proposition 9 *Take a vector Γ satisfying (4), so that $z_t := \Gamma'(y_t - C_t^y m^y) = \Gamma u_t$, and a $k \times N$*

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

matrix $\mathcal{E}$ satisfying (82). Then, we can identify $\dot{\phi}^d$ via the moment:

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

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

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

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

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

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

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

C.6 Proof of Proposition 7

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

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

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

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

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

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

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

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

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

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

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

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

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

The estimator is $\sqrt{T}$-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, $\sqrt{T}\left(\theta_T^e - \theta\right)$ converges in distribution to a normal distribution with mean 0 and positive definite variance covariance matrix V^θ.

C.7 Proof of Proposition 8

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

D Complements

D.1 Detailed links with previous literature

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

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

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

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

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

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

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

able to estimate a richer set of parameters.

In a tangentially related recent paper, Sarto (2022) uses factor analysis to extract values of η_t (much as we do when we "recover" a factor η_t). Take the basic example in our paper. Then, Sarto does not identify $\phi^s = \frac{1}{\psi}$: even if η_t (the aggregate shock to demand) were perfectly identified, that would not allow to estimate p_t. In the supply and demand example, Sarto's approach would identify the demand elasticity ϕ^d, but not the supply elasticity ϕ^s.

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

Spatial econometrics In some applications of GIVs we have considered separately, growth in a region affects that of the other regions. So there is a similarity between our setup and that of spatial econometrics (e.g. Kelejian and Prucha (1999)). However, the estimators are quite different. The reason is that spatial econometrics studies the "local" influence (e.g. of neighboring cities on a city), while GIVs study the global influence. Hence, the sources of variation, identifiability conditions and methods are quite different. Certainly, the spatial literature has not identified, as we do, the GIVs as a simple way to estimate elasticities in contexts such as supply and demand problems, and models with general equilibrium effects as opposed to local effects. Still, some of the sophisticated techniques of the spatial literature might be used one day to enrich a GIV-type analysis.

Quasi-experimental instruments and identification by functional form A large literature explores identification by functional form, where consistency of the estimator depends on functional form or distributional assumptions. Classic examples include the Heckman (1978) selection model, identification via heteroskedasticity, as in Rigobon (2003) and Lewbel (2012), and

Arellano and Bond (1991) and Blundell and Bond (1998) in the context of dynamic panel data models. The typical concern with these approaches, compared to quasi-experimental instruments that are outside of the model, is that the estimators are inconsistent when the model is misspecified.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D.3 Estimating the variance of heteroskedastic idiosyncratic shocks

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

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

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

Proof. We have

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

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

Next, we have the following.

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

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

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

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

$$\check{y}_t \coloneqq q y_t = q\lambda\eta_t + q u_t$$

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

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

Hence,

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

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

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

Formally, defining

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

then we have, at the true value θ_0,

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

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

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

The estimator is $\sqrt{T}$-asymptotically normal. Under the regularity conditions we assumed, by

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

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

Newey and McFadden (1994), Theorem 3.4, $\sqrt{T}\left(\theta_T^e - \theta_0\right)$ converges in distribution to a normal distribution with mean 0 and positive variance covariance matrix V^θ. □

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

Stochastic volatility This procedure assumes constant volatility. It could also be extended, to stochastic volatility, along the lines of the GARCH literature. Recall that the generalization of $E_i = \frac{1}{N}$ was $\tilde{E}_i := \frac{1/\sigma_{u_i}^2}{\sum_j 1/\sigma_{u_j}^2}$, see (54). Then, the correct generalization of (54) is $\tilde{E}_{it} := \frac{1/\sigma_{u_{it|t-1}}^2}{\sum_j 1/\sigma_{u_{jt|t-1}}^2}$, where $\sigma_{u_{it|t-1}}^2 = \mathbb{E}_{t-1}\left[u_{it}^2\right]$ is the conditional expected variance.[50] Then, with $\Gamma_t = S - E_t$, we have the generalization of (10),[51]

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

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

D.4 Relation between Bartik instruments and GIVs

D.4.1 Relating the Bartik setup to the GIV setup

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

As a general matter, in a number of cases where a cross-section is used (e.g. Autor et al. (2013)),

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

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

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

Bartik applies, but GIV does not apply, for instance because there is no large idiosyncratic shock that one can use.

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

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

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

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

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

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

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

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

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

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

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

D.4.2 Defining and comparing the Bartik and GIV instruments

The Bartik instrument is defined as

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

The GIV is defined as

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

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

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

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

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

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

and

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

Hence, we have $\lim_{N\to\infty} z_t^{Bartik} = \eta_t$ and $\lim_{N\to\infty} z_t^{GIV} = u_{St}$. As before, we work with the large N version of the instruments to simplify the exposition. To provide a simple example where the Bartik instrument may be less appealing, we consider a trivial version of our baseline model in Section 2 (with $\phi^d = 0$)

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

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

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

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

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

Alternatively, we can use the Bartik instrument and assume

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

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

D.5 More general setup and multipliers

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

D.5.1 Framework

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

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

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

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

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

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

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

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

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

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

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

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

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

We rewrite model (99) in vector form:

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

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

D.5.2 Multipliers

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

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

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

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

Hence, an idiosyncratic shock has an impact on the aggregate action y_{St} that is M_t times bigger

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

than its direct effect. Also, the total impact of an idiosyncratic shock on factor f is:

$$F_t^f = M_t \alpha^f u_{St} + \varepsilon_t^f, \tag{106}$$

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

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

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

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

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

$$\begin{aligned}
y^a_{S^a t} &= \sum_f \lambda^a_{S^a,f} F^f + u^a_{S^a t}, \\
F_t^f &= \eta_t^f + \sum_a \alpha_a^f y^a_{S^a,t},
\end{aligned}$$

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

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

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

where

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

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

$q \times q$, and

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

This implies

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

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

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

We will form a GIV:

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

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

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

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

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

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

We have:

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

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

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

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

Then,

$$\Omega^{ab} = \sum_i \frac{k}{\sigma_i^2} \Gamma_i^b \sigma_i^2 v^{ab} = k v^{ab} \sum_i \Gamma_i^b = 0,$$

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

Given (107), we have

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

so

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

hence our estimator is

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

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

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

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

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

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

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

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

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

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

Hence, we can identify μ_i, by regression

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

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

D.7 Nonlinear GIV

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

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

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

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

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

D.8 When the researcher assumes too much homogeneity

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

The model becomes, for the demand:

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

and for the supply

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

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

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

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

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

so

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

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

Hence, when we run the first stage

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

we will gather

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

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

If we run

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

we will estimate

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

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

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

D.9 Heterogeneous demand elasticities: Non-parametric extension

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

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

We still assume parametric loading of unobserved factors η.

We propose two procedures to estimate ϕ^d.

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

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

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

We now need not assume parametric knowledge of λ.

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

We propose the following procedure.

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

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

2. We estimate $\check{\lambda} := Q^\phi \lambda$, $\eta_t^e = R^{\check{\lambda}} Q^\phi y_t$, V^u, $\check{u}_t = Q^{\check{\lambda},\phi} u_t$. We form $z_t := S' \check{u}_t$, and $\Gamma := \left(Q^{\check{\lambda},\phi}\right)' S$.

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

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

   $$a_j^i := \frac{V_{ij}^{\check{u}}}{V_{ii}^{\check{u}}} \tag{120}$$

   (b) We define the instrument for entity i,

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

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

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

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

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

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

This may be useful as a starting point numerically.

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

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

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

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

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

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

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

so that

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

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

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

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

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

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

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

We examine in more detail how to estimate $\psi := \phi^d_\Gamma$. Calling the true value $z_t(\psi) = u_{\Gamma t}$, we have $\mathbb{E}\left[z_t^2\right] = \sigma^2_{u_\Gamma}$, where $\sigma^2_{u_\Gamma} = \Gamma' V^u \Gamma$ is the theoretical variance of z_t given in (130). So, we solve for ψ^c (a *c*andidate answer for ψ) so that the empirical variance of the GIV is equal to its theoretical variance:

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

i.e. $\mathbb{E}\left[p_t^2\right]\left(\psi^c\right)^2 - 2\mathbb{E}\left[y_{\Gamma t} p_t\right]\psi^c + \mathbb{E}\left[y^2_{\Gamma t}\right] - \sigma^2_{u_\Gamma} = 0$. This is a quadratic equation in ψ^c, which yields two roots:[59] a good (i.e. correct) root, $\psi^G = \psi$, and a bad root, $\psi^B = \psi + 2\frac{\mathbb{E}\left[z_t^* p_t\right]}{\mathbb{E}\left[p_t^2\right]}$. Fortunately, there is an economic way to determine which is the correct root. Calling G (resp. B) the estimation with the good (resp. bad) root, one can show that:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

We wish to estimate ϕ and αM.

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

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

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

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

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

$$V^Y = \sigma_\eta^2 \lambda \lambda' + \sigma_\varepsilon^2 \Psi \Psi' + \sigma_{\eta\varepsilon} (\lambda \Psi' + \Psi \lambda') + (I + \Psi S') V^U (I + S \Psi') \tag{136}$$

$$= \sigma_\eta^2 \lambda \lambda' + \gamma \Psi \Psi' + \Psi b' + b \Psi' + V^U \tag{137}$$

$$b = \sigma_{\eta\varepsilon} \lambda + V^U S \tag{138}$$

$$\gamma = \sigma_\varepsilon^2 + S' V^U S \tag{139}$$

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

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

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

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

[60] This idea of stacking together then y_t and p_t, with a "size 0" for the innovations to the price, could be fruitfully used more generally.

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

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

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

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

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

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

If we truncate the residuals, i.e. use

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

for the hard thresholding function

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

for some $b > 0$, then everything works too. Indeed, we have that $\check{u}_{it} \coloneqq u_{it} - u_{Et}$ is orthogonal to u_{Et}. Let us assume that it is independent. Also, this could be narratively checked. In our basic example of Section 2.1, we still have $\mathbb{E}\left[(p_t - \psi y_{St}) z_t\right] = 0$, so that the IV procedure still works.

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

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

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

D.11 Sporadic factors

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

One solution, besides the narrative check, would be to filter out days with a high "sporadicity statistic" $\mathcal{S}_t$ that we now propose. Suppose that for each date we filter out the idiosyncratic shocks $\check{u}_{it}$. For each date and actor i we form $b_{it} = \frac{\check{u}_{it}^2}{\sigma_{u_i,t-1}^2}$, where a high b_{it} is an indicator of extra activity, and $\sigma_{u_i,t-1}^2$ is a predictor of the volatility of u_{it}. We may allow that one entity has a large idiosyncratic shock, but if two (or more) do, this is suspicious, and possibly the sign of a sporadic factor. So, calling $b_{(2)t}$ the activity of the second more active actor, we form $\mathcal{S}_t = b_{(2)t}$.[61] Over the entire sample, we might remove the days with anomalously high sporadicity statistics, e.g. in the top 5% by that metric.

D.12 PCA and IPCA

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

We can consider the case where we impose no structure on the loadings and we use PCA to estimate the loadings and the factor realizations. One can use the GMM procedure of Proposition

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

7, which works with finite N. The asymptotic theory for PCA (with $N \to \infty$) has been developed in Bai (2003) and in the context of GIV by Banafti and Lee (2022).[62] To compute standard errors, we can use the GMM values from Proposition 7; or a bootstrap. The NBER Working Paper of this version contains detailed guidance of this.

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

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

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

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

Here we take the basic set up as in Section D.5.1, in the simplified case where $\lambda_i^f = \lambda^f$ for all "endogenous" factors, i.e. for the factors f such that $\alpha^f \neq 0$, the other exogenous factors η all have an impact of 1:

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

This implies

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

We posit that demand and supply disturbances follow:

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

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

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

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

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

so that the supply shock is

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

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

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

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

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

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

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

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

Then we have

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

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

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

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

Proposition 12 (Identification with disaggregated supply and demand data). *Suppose that we have disaggregated supply and demand data following* (150). *Suppose that the shock are idiosyncratic*

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

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

in the sense of (152)*, and that we have i.i.d.* $\left(u_{it}^d, u_{it}^s\right)$ *across* i*,* t*. Then, the GIVs* z_t^d*,* z_t^s *in* (153) *identify* ϕ^d *and* ϕ^s*, via moments* (154)*.*

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

D.15 GIV for differentiated product demand systems

We develop the basic ideas for the logit demand model and extend these ideas to the random-coefficients logit model as in Berry et al. (1995) in the next subsection.[65]

D.15.1 Logit demand

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

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

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

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

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

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

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

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

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

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

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

It is convenient to rewrite the model as

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

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

$$z_{it}^{BLP} = \frac{1}{N-1}\sum_{j,j\neq i} x_{jt},$$

which results in valid instruments under some assumptions (see Nevo (2000) and the references therein).[67] The resulting moment is $\mathbb{E}\left[z_{it}^{BLP}\xi_{it}\right] = 0$.[68] Third, one can use panel data for the same firm that operates in different locations. Under the assumption that demand shocks are uncorrelated across locations, prices in other locations of the same firm will be valid instruments. The intuition is that prices across locations share the same marginal cost but the demand shocks are, by assumption, uncorrelated, see Nevo (2001).

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

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

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

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

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

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

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

where $\bar{s}_{j,t-1}$ is the average market share for product j up to time $t-1$. This allows us to add a moment condition

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

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

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

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

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

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

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

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

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

D.15.2 Random coefficients logit as in BLP

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

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

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

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

where

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

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

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

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

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

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

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

This suggests forming

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

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

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

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

D.16 Dealing with fat tails

Here we propose an enrichment of the GIV when there are outliers.

D.16.1 OLS with winsorized residuals

For clarity, we consider the following problem first — our main problem is just a more complex variant. Suppose that we want to estimate β in a regression:

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

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

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

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

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

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

with

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

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

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

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

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

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

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

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

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

2. Run the OLS of

$$y_i^{w,(n)} = \beta x_i + \varepsilon_i. \tag{167}$$

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

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

$$\min_\beta \sum_i L(y_i - \beta x_i) \tag{168}$$

whose first order condition is

$$\sum_i L'(y_i - \beta x_i)\, x_i = 0 \tag{169}$$

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

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

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

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

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

Then, the FOC (169) becomes

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

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

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

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

We next state a simple proposition.

Proposition 13 *Suppose that $\mathbb{E}[r(u_i) x_i] = 0$ for all i.. Then, at the true value β we have*

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

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

Proof. At the correct value of β, we have

$$\mathbb{E}[(y_i^w - \beta x_i) x_i] = \mathbb{E}[r(y_i - \beta x_i) x_i] = \mathbb{E}[r(u_i) x_i] = 0.$$

□

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

D.16.2 Factor models with winsorized residuals

We state a proposition summarizing the situation. We define for $x \in \mathbb{R}^N$, we define $r(x) = (r(x_i))_{i=1...N}$.

Proposition 14 (Factor model with winsorized residuals) *Using the tail-extraction function r in (164), consider the value $\mu_t \in \mathbb{R}^{\dim \eta_t}$ such that, defining: $u_t^\tau = -r(u_t - \lambda \mu_t)$ the tail residual, and*

$v_t := u_t - u_t^\tau$ *the winsorized residual, we have*

$$\mu_t = R^{\lambda, W^v} v_t \tag{175}$$

Assume that μ_t *exists. Then, with* $y_t^w := y_t - u_t^\tau$ *the outcome modified to have winsorized residuals, and*

$$\check{\eta}_t := \eta_t + R v_t = \eta_t + \mu_t, \qquad \check{v}_t := v_t - \lambda R v_t = Q^{\lambda, W^v} v_t \tag{176}$$

we have the decomposition of the factor model with winsorized residuals:

$$y_t^w = \lambda \eta_t + v_t = \lambda \check{\eta}_t + \check{v}_t, \tag{177}$$

and $\check{v}_t \perp \check{\eta}_t$.

In addition,

$$u_t^\tau = -r\left(y_t - \lambda \check{\eta}_t\right) \tag{178}$$

Finally, we have:

$$\lambda' W^v \left(y_t^w - \lambda \check{\eta}_t\right) = 0 \tag{179}$$

$$\mathbb{E}\left[W^v \left(y_t^w - \lambda \check{\eta}_t\right) \check{\eta}_t'\right] = 0 \tag{180}$$

so the factor model (177) can be estimated in the usual way, via $\min_{\lambda, \check{\eta}_t} \sum_t \left(y_t^w - \lambda \check{\eta}_t\right)' W^v \left(y_t^w - \lambda \check{\eta}_t\right)$.

Equation (178) says that the tails terms u_t^τ are tails of the estimated factor model; so that y_t^w is the value of y_t, but winsorizing the residuals, as in Section D.16.1, but for a factor model. The term v_t can be thought of as the "winsorized" version of $u_t - \lambda \mu_t$; so v_t can be thought of a winsorized version of u_t, while u_t^τ is its tail. The term $\lambda \mu_t$ is there because we keep finite N, as in Lemma 1 and the rest of the paper; in practice is is small.

Pragmatically, given the tails u_t^τ, we estimate a factor model by (179), which gives the value of $\check{\eta}_t$, and (180), which gives the value of λ; and given the factor model, we estimate the tails by (178).

Hence this is a fixed point problem given by equations (178)-(180).

Practical algorithm One can use the following algorithm.[69]

1. Start from some $(\lambda, \check{\eta}_t)$ and δ.[70]

2. Compute u_t^τ using (178).

3. Compute $y_t^w := y_t - u_t^\tau$

4. Run PCA on y_t^w to update $(\lambda, \check{\eta}_t)$. Note that we use a weighted PCA, weighing by W^{v_t}.

5. Get the new residuals and update δ in (164) so that the fraction winsorized remains 5% in each step.

6. Iterate until convergence.

Proof of Proposition 14 Given this μ_t, we have $y_t - u_t^\tau = \lambda\eta_t + v_t$, and (177) comes from Lemma 1, extended for heteroskedastic v_t, as in the proof of Lemma 1. Then, $\mu_t = R\left(u_t - u_t^\tau\right) = R\left(y_t - \lambda\eta_t - u_t^\tau\right) = Rv_t$ and

$$y_t^w - \lambda\check{\eta}_t = \lambda\check{\eta}_t + \check{v}_t - \lambda\check{\eta}_t = \check{v}_t = v_t - \lambda R v_t = v_t - \lambda\mu_t = u_t - u_t^\tau - \lambda\mu_t$$

so that

$$u_t - \lambda\mu_t = y_t^w - \lambda\check{\eta}_t + u_t^\tau = y_t - \lambda\check{\eta}_t$$

which proves (178) by:

$$u_t^\tau := -r\left(u_t - \lambda\mu_t\right) = -r\left(y_t - \lambda\check{\eta}_t\right)$$

Finally, (179) and (180) hold by the properties of factor models, as in the main paper.□

D.16.3 Implication for the GIV

We now see how the procedure in Section D.16.2 allows to form a GIV.

[69] It does not require the calculation of μ_t (which is simply useful to state the existence conditions).
[70] A good starting point can be obtained by winsorizing y_{it} and use this to obtain these initial estimates.

Elasticity of supply For the GIV, we form

$$\check{u}_t := \check{v}_t + u_t^{\tau}, \qquad z_t := S'\check{u}_t$$

so we do include the tails (recall that $\check{v}_t$ is the winsorized main term, and u_t^{τ} is the tails) in the GIV z_t. Then, we have:

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

so we can identify ψ with that z_t. We can also control for $\check{\eta}_t$, in the first and second stage.

Elasticity of demand We have $\check{v}_t \perp \check{\eta}_t$ by Proposition 14. But there is no guaranty that $u_t^{\tau} \perp \check{\eta}_t$, as the term μ_t might introduces correlations between the two. In practice, those correlations can be expected to be small ($O\left(\frac{1}{N}\right)$), so that pragmatically one might be tempted to ignore this. But to be strict, it is useful to have a rigorously grounded proven procedure, so we present one here.

So we propose to proceed as follows. We estimate λ globally. Then, for each i, we use the subsample made of all observations except for those of entity i, and from this subsample we estimate $\check{\eta}_t^{(i)}$ (using the winsorized residuals, as in $\check{\eta}_t^{(i)} = R^{\lambda^{(i)}} y_t^{w,(i)}$,where $\lambda^{(i)}$ and $y_t^{w,(i)}$ are the vectors of dimensions $N-1$ that exclude the i-th entry) and $\check{u}_t^{(i)} := \check{v}_t^{(i)} + u_t^{\tau,(i)}$ (using both winsorized residuals, *and the tails*). We set $\check{u}_{it}^{(i)} := 0$. We form the GIV for entity i, $z_t^{(i)} := S'\check{u}_t^{(i)}$. This guaranties

$$z_t^{(i)}, \check{\eta}_t^{(i)} \perp u_{it} \tag{181}$$

Then, the following moment holds

$$\mathbb{E}\left[\left(y_{it} - \phi^d p_t\right) z_t^{(i)}\right] = 0 \tag{182}$$

and we can estimate ϕ^d for instance by $\sum_{i,t}\left(y_{it} - \phi^d p_t\right) z_t^{(i)} = 0$. We can also control for $\check{\eta}_t^{(i)}$, as in

$$\mathbb{E}\left[\left(y_{it} - \phi^d p_t - b_i \check{\eta}_t^{(i)}\right)\left(z_t^{(i)}, \check{\eta}_t^{(i)}\right)\right] = 0 \tag{183}$$

for some b_i also to be estimated.

The same can be used for the network GIV. For each entity i, we estimate the $\eta_t^{(i)}$ with winsorization of the residuals of entities j, but use also the tail events from the GIV, with $\check{u}_t^{(i)} \coloneqq \check{v}_t^{(i)} + u_t^{\tau,(i)}$. The cost is that it requires N^2T estimations rather than NT estimations (as we need to estimate $u_{jt}^{(i)}$ rather than just u_{jt}). The benefit is that this is robust to outliers, and to misspecification in heteroskedasticity.

D.17 When aggregate shocks are made of idiosyncratic shocks

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

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

$$y_{it}^k = \phi^k p_t + \lambda_i^k \eta_t^k + u_{it}^k, \tag{184}$$

where $k = s, d$ indicates supply or demand, respectively. We allow $\mathbb{E}\left[u_{it}^s u_{it}^d\right]$ to be nonzero: for instance, if the US has a positive "fracking shock" that affects both supply and demand, it will be captured by a positive u_{it}^s and u_{it}^d for $i =$ USA. This is a concrete case in which supply and demand shocks are correlated: this happens via the correlations in country-level shocks. At the same time, we impose that the u_{it}^k are uncorrelated with the aggregate shocks $\eta_t^{k'}$ for $k, k' \in \{s, d\}$. Then, Section D.14 shows how to identify the elasticities of supply and demand.

One can also consider an economy as a network (Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019)). Under some assumptions, one can obviate the network structure, for instance via aggregation theorems such as Hulten's theorem. This is developed in Section D.18. It shows that we can identify important multipliers even if we have only crude proxies for the primitive shocks such as TFP. The GIV for a general network is a rich topic, which are developing in another paper.

In conclusion, one can often handle cases where aggregate shocks are made of idiosyncratic shocks: then, some more disaggregated data and economic reasoning allows to use a GIV to estimate macro parameters of interest.

D.18 Identification of the TFP to GDP multiplier in a production network economy

Suppose a two-period model with a production network, as in Long and Plosser (1983); Gabaix (2011); Acemoglu et al. (2012); Carvalho and Gabaix (2013); Carvalho and Grassi (2019). There are both idiosyncratic TFP shocks $\hat{\lambda}_{it}$ and a government reform that creates correlated shocks η_t to TFP and change in labor supply $\hat{L}_t$. Utility is $C_t - e^{\eta_t^L} L_t^{1+1/\phi}$, so that ϕ is the Frisch elasticity of labor supply. We call λ_t total TFP, which depends on the industry TFPs λ_{it}. So, as $C_t = \lambda_t L_t$, labor supply is $\hat{L}_t = \phi\left(\hat{\lambda}_t - \eta_t^L\right)$,[71] and GDP is $\hat{Y}_t = \hat{L}_t + \hat{\lambda}_t$, i.e.

$$\hat{Y}_t = m\hat{\lambda}_t - \phi\eta_t^L, \qquad m = 1 + \phi \tag{185}$$

We seek to find the "GDP multiplier" $m = 1 + \phi$, so that a TFP increase of 1 percent translates into a GDP increase of m percent.[72]

This is potentially a complicated problem, as for instance, in the Long and Plosser (1983) case with input-output matrix A, output changes are $\hat{Y}_t = (I - A)^{-1}\hat{\lambda}_t + \hat{L}_t$, so that output changes are correlated in complicated ways. However, one can sidestep using this disaggregated production

[71] The problem is $\max_{L_t} \lambda_t L_t - e^{\eta_t^L} L_t^{1+1/\phi}$, which leads to $\left(1+\frac{1}{\phi}\right) L_t^{1/\phi} = \lambda_t e^{-\eta_t^L}$, hence the announced expression.

[72] If more than one factor changes, m has the broader interpretation of a multiplier between TFP and GDP.

data. We assume that TFP change in industry i is:

$$\hat{\lambda}_{it} = \lambda_i \eta_t^{\lambda} + u_{it}. \tag{186}$$

In the neoclassical equilibrium, TFP follows Hulten's theorem, so is $\hat{\lambda}_t = \sum_i s_i \hat{\lambda}_{it}$ where s_i is the Domar weight (sales of industry i over GDP).

We can identify the multiplier m if we have disaggregated TFP data. In the simplest case, we assume that industry-level productivities are available, and we get the residuals u_{it}^e. Then, we can identify the multiplier m by GIV.

We can identify the multiplier m if we have even crude proxies for disaggregated TFP. The same procedure works (with less efficiency) if our data is made of proxies for productivity growth $\tilde{\hat{\lambda}}_{it}$ (where the tilde indicates that we deal with a proxy). An example could be growth of sales per employee, or even the growth rate of sales. We assume a factor model

$$\tilde{\hat{\lambda}}_{it} = \tilde{\lambda}_i \tilde{\eta}_t^{\lambda} + \tilde{u}_{it}. \tag{187}$$

The proxy is of better quality when the proxy's idiosyncratic shock $\tilde{u}_{it}$ has a high correlation with the true idiosyncratic shock u_{it}. Then, we extract the $\tilde{u}_{it}^e$ from a factor model, form $z_t = \tilde{u}_{St}^e - \tilde{u}_{Et}^e$ (with $S_i = \frac{s_i}{\sum_j s_j}$), and use the moment $\mathbb{E}\left[\left(\hat{Y}_t - m\hat{\lambda}_t\right) z_t\right] = 0$, which identifies the TFP to GDP multiplier m.

Using more general models (e.g. taking into account imperfections as in Baqaee and Farhi (2020)) would be very interesting, but would be a new paper by itself. Indeed, even in that case z_t is likely to be a useful instrument, even though it won't be the optimal one. In any case, those examples show how GIV, with some economic reasoning, translate to more complex economies where aggregate shocks can be made of idiosyncratic shocks.

D.19 Identification of the elasticity of substitution between capital and labor / Elasticity of demand in partially segmented labor markets

Here we show how GIVs can estimate the elasticity of substitution between capital and labor; and how to estimate the elasticity of demand in partially segmented markets. The first problem uses the second one.

As a motivation, imagine that industry i has the CES production function[73]

$$Q_{it} = B_{it}\left(K_{it}^{\frac{\phi_i-1}{\phi_i}} + A_{it}^{\frac{1}{\phi_i}} L_{it}^{\frac{\phi_i-1}{\phi_i}}\right)^{\frac{\phi_i}{\phi_i-1}} \tag{188}$$

The first order condition of the problem $\max_{K_{it},L_{it}} Q_{it} - R_t K_t - W_{it} L_{it}$ is $A_{it}^{\frac{1}{\phi_i}}\left(\frac{L_{it}}{K_{it}}\right)^{-\frac{1}{\phi_i}} = \frac{W_{it}}{R_t}$,i.e. a demanded labor / capital ratio:

$$\frac{L_{it}}{K_{it}} = A_{it}\left(\frac{W_{it}}{R_t}\right)^{-\phi_i} \tag{189}$$

We'd like to estimate the elasticity of substitution ϕ_i between capital and labor. This is the wage elasticity of demand. GIVs allow to estimate that, as we shall see.

Let us use our general notations, and define $y_{it}^d = \ln L_{it}$, $p_{it} = \ln W_{it}$, $C_{it} = \ln K_{it}$, and $\phi_i^d = -\phi_i$ (as this is the elasticity of labor demand). Then, we can write (189) as:

$$y_{it}^d = \phi_i^d p_{it} + C_{it} + \lambda_i^d \eta_t + u_{it}^d \tag{190}$$

where C_{it} is a control, and as usual vector η_t is a common shock, and u_{it}^d is a demand shock (those in turn come from the productivity A_{it}). For notational simplicity we will drop C_{it}, but this is not important.

Now, log labor supply is modeled as:

$$y_{it}^s = \phi_i^s p_{it} - \psi_i p_{St} + \lambda_i^s \eta_t + u_{it}^s \tag{191}$$

[73] We thank Julieta Caunedo for prompting us to think about this identification problem.

It is increasing in wage p_{it} in industry i, and decreasing in the wage in the other industries (p_{St}). One could imagine replacing $\psi_i p_{St}$ by a different average for each industry, and we will examine that in an extension. But for now we keep the simple structure.

As supply equals demand in each market ($y_{it}^d = y_{it}^s$), we obtain the price of labor in each market i:

$$p_{it} = \frac{\psi_i p_{St} + u_{it}^d - u_{it}^s + \left(\lambda_i^d - \lambda_i^s\right)\eta_t}{\phi_i^s - \phi_i^d} \tag{192}$$

i.e.

$$p_{it} = \gamma_i p_{St} + v_{it} + \lambda_i^p \eta_{it} \tag{193}$$

when we define $\gamma_i = \frac{\psi_i}{\phi_i^s - \phi_i^d}$, $v_{it} = \frac{u_{it}^d - u_{it}^s}{\phi_i^s - \phi_i^d}$, $\lambda_i^p = \frac{\lambda_i^d - \lambda_i^s}{\phi_i^s - \phi_i^d}$.

Problem (193) is a standard GIV. In the general case, we can estimate γ_i as in Section (D.9).[74]

So, we obtain γ_i and v_{it}^e (the proxy for v_{it}) in (193). We also form $z_{it} = S'\left(v_t^e - a^i v_{it}^e\right)$ as in (121). This is the GIV formed of the idiosyncratic shock of all industries but industry i. We will use the shock to those other industries, and their impact on the outside wage, as an instrument to estimate labor demand. Indeed, we go back to the labor demand equation (190), and instrument for p_{it} using the z_{it}

$$p_{it} = b_i z_{it} + \varepsilon_{it}^p \tag{194}$$

we estimate b_i, and define $p_{it}^e = b_i z_{it}$ as the price in industry i instrumented by the changes in other industries. We use the estimated η_t^e as controls, and run

$$y_{it}^d = \phi_i^d p_{it}^e + C_{it} + \lambda_i^d \eta_t^e + u_{it}^d \tag{195}$$

which yields a consistent estimate ϕ_i^d of the labor demand.

[74]This procedure is much simplified if the γ_i and λ_i^p are assumed to be constant. Then, we can just define $z_t := p_{\Gamma t} = p_{St} - p_{Et}$, so that $z_t = v_{\Gamma t}$ and as $p_{St} = \frac{v_{St} + \lambda^p \eta_t^s}{1-\gamma}$, regressing $p_{St} = b z_t + \varepsilon_t^p$ yields $b = \frac{1}{1-\gamma}$.

References for Online Appendix

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.