Granular Instrumental Variables in Large Panels

Identification and Inference Across Strong, Nearly Weak, and Weak GIV

Gokul Gopalan Ramachandran

Department of Economics, University of Warwick

May 20, 2026

GIV: Aggregate Variables, No External Instruments

The empirical setting

Many questions in economics estimate structural relationships between aggregate time series: asset prices on portfolio demand, exchange rates on capital flows, commodity prices on aggregate consumption.

GIV: Aggregate Variables, No External Instruments

The empirical setting

Many questions in economics estimate structural relationships between aggregate time series: asset prices on portfolio demand, exchange rates on capital flows, commodity prices on aggregate consumption.

- Aggregate demand equation

$$d_t = \phi_d \, p_t + \varepsilon_t \quad t = 1, \ldots, T$$

- d_t is aggregate quantity, p_t is the market-clearing price, ϕ_d is the demand elasticity
- Endogeneity: $\mathbb{E}[p_t \varepsilon_t] \neq 0$. Aggregate demand shocks move price and quantity together
- External instruments not available

Model and Idiosyncratic Shocks

- Goal: estimate the demand elasticity ϕ_d in $d_t = \phi_d \, p_t + \varepsilon_t$

Model and Idiosyncratic Shocks

- Goal: estimate the demand elasticity ϕ_d in $d_t = \phi_d \, p_t + \varepsilon_t$
- Unit-level supply with interactive fixed effects

$$y_{it} = \phi_s \, p_t + \lambda_i' F_t + u_{it} \quad i = 1, \ldots, N; \;\; t = 1, \ldots, T$$

- Market clearing: $d_t = \sum_{i=1}^{N} S_i y_{it}$, where S_i is the individual market share
- Thus $p_t = f(F_t, \varepsilon_t)$ and $\mathrm{Cov}(F_t, \varepsilon_t) \neq 0$: common factors move both supply and the demand shock, inducing $\mathbb{E}[p_t \varepsilon_t] \neq 0$

Model and Idiosyncratic Shocks

- Goal: estimate the demand elasticity ϕ_d in $d_t = \phi_d \, p_t + \varepsilon_t$
- Unit-level supply with interactive fixed effects

$$y_{it} = \phi_s \, p_t + \lambda_i' F_t + u_{it} \quad i = 1, \ldots, N; \;\; t = 1, \ldots, T$$

- Market clearing: $d_t = \sum_{i=1}^{N} S_i y_{it}$, where S_i is the individual market share
- Thus $p_t = f(F_t, \varepsilon_t)$ and $\mathrm{Cov}(F_t, \varepsilon_t) \neq 0$: common factors move both supply and the demand shock, inducing $\mathbb{E}[p_t \varepsilon_t] \neq 0$

Assumption 1 (Idiosyncrasy)

$\mathbb{E}[F_t \, u_t] = 0 \quad$ and $\quad \mathbb{E}[\varepsilon_t \, u_t] = 0$

- u_t is uncorrelated with factors and demand shock. It is a natural candidate IV for p_t

Granular Instrumental Variables (GIV)

Key insight (Gabaix and Koijen, 2024)

When a few large units drive the aggregate, their idiosyncratic shocks instrument the aggregate.

Granular Instrumental Variables (GIV)

Key insight (Gabaix and Koijen, 2024)

When a few large units drive the aggregate, their idiosyncratic shocks instrument the aggregate.

- For $e_N = (1, \ldots, 1)$ define $D_N = I_N - \frac{e_N e_N'}{N}$. The demeaned-data has a pure interactive fixed effects structure

$$D_N y_t = D_N \Lambda F_t + D_N u_t$$

- Share-weighted sum of demeaned idiosyncratic shocks is the optimal GIV

$$z_t = S' D_N u_t; \quad S = (S_1, \ldots, S_N)$$

- Identification requires $S \neq e_N / N$. GK call this granularity

Empirical Applications Have Large Panels

- GK build the theory for fixed N and $T \to \infty$
- Empirical applications routinely have moderate to large N
  - Aldasoro et al. (2023): $N = 21$
  - Chodorow-Reich et al. (2021), Galaasen et al. (2020), Ma et al. (2021): $N > 100$
  - Exchange rates, monetary policy, sovereign bonds: similar magnitudes

Empirical Applications Have Large Panels

- GK build the theory for fixed N and $T \to \infty$
- Empirical applications routinely have moderate to large N
  - Aldasoro et al. (2023): $N = 21$
  - Chodorow-Reich et al. (2021), Galaasen et al. (2020), Ma et al. (2021): $N > 100$
  - Exchange rates, monetary policy, sovereign bonds: similar magnitudes
- Appropriate asymptotic framework: $N, T \to \infty$
- This requires extending GIV theory to large panels

The Challenge: As $N \to \infty$, Granularity Is Not Automatic

- Granularity is a property of the cross section: the behavior of $(S_i)_{1 \leq i \leq N}$ as $N \to \infty$
- When N is fixed, $S_i \neq 1/N$ for some i is a necessary condition
- In large panels, the *order* of S matters

$$S_i = O\left(\frac{1}{N}\right) \implies z_t = S' D_N u_t = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right)$$

- The instrument can vanish as $N \to \infty$

This Paper

Question. What is the right asymptotic theory for GIV in large panels, and how do we formalize granularity as $N \to \infty$?

This Paper

Question. What is the right asymptotic theory for GIV in large panels, and how do we formalize granularity as $N \to \infty$?

Contributions

1. **Asymptotic theory.** Joint (N, T) theory for the GIV estimator under a power-law formalisation of granularity — consistency, rates, limiting distributions.
2. **Granularity.** The tail index μ is a sufficient statistic for instrument strength: three regimes (strong / nearly weak / weak); Wald for $\mu < 2$, Anderson–Rubin for $\mu > 2$.
3. **GIV under Estimated Factor Structure.** Same rates and inference rule when the factor structure is estimated (under $\sqrt{T}/N \to 0$). Variance changes.
4. **Empirics.** GIV applied to three commodity markets to estimate demand elasticities.

Topline: One Sufficient Statistic, Three Regimes

Model sizes as power-law

$\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}$, with shares $S_i = \mathcal{S}_i / \sum_j \mathcal{S}_j$. The tail index μ is the sufficient statistic for instrument strength.

Topline: One Sufficient Statistic, Three Regimes

Model sizes as power-law

$\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}$, with shares $S_i = \mathcal{S}_i / \sum_j \mathcal{S}_j$. The tail index μ is the sufficient statistic for instrument strength.

Identification	Condition	Rate	Inference
Strong	$\mu \in (0,1)$	$\sqrt{T}$	Wald
Nearly weak	$\mu \in (1,2)$, $N/T \to 0$	$\sqrt{T}/N^{1-1/\mu}$	Wald
Nearly weak	$\mu > 2$, $N/T \to 0$	$\sqrt{T/N}$	Anderson–Rubin
Weak	$\mu > 2$, $N/T \to c$	Inconsistent	Anderson–Rubin

- $N/T \to 0$ required when weakness is suspected. Time-series growth compensates for the weakness in the cross section
- Wald inference when $\mu < 2$

Topline: GIV Recovers Sensible Elasticities

- GIV estimates of demand elasticities: Copper (-0.135), Crude Oil (-0.109), Natural Gas (-0.056)
- All negative, $|\hat{\phi}_d| < 0.15$, consistent with the literature
- The factor structure is estimated by PCA in practice. Treating it as known gives incorrect SEs — the variance must account for first-stage estimation error
- Correction can shift SEs in either direction

Topline: All Three Markets Sit Below $\mu = 2$

	Copper	Crude Oil	Natural Gas
$\hat{\mu}$ (Gabaix–Ibragimov)	0.58	0.65	0.32
95% CI	$[0.28, 0.87]$	$[0.26, 1.05]$	$[0.15, 0.49]$
Identification	Strong	Strong / nearly weak	Strong

- Crude oil's CI reaches into $(1, 2)$.
- All three rule out $\mu > 2$. Wald inference is justified in each case

Where This Paper Sits

- **GIV theory.** Gabaix and Koijen (2024) introduce GIV for fixed N and $T \to \infty$. Banafti and Lee (2022) consider large panels under $\mu < 1$ only, with an additional condition I show is incompatible with the power law. Qian (2023) has heterogeneous slopes, Baumeister and Hamilton (2023) have joint MLE, both untenable as $N \to \infty$
- **Factor models for large panels.** Bai (2003) for first-stage contribution; Bai and Ng (2002); Ahn and Horenstein (2013) for number of factors
- **Weak instruments.** Weakness arises endogenously as in Armstrong (2016) without the local-to-zero assumption of Staiger and Stock (1997). $\mu > 1$ regime maps directly onto Antoine and Renault (2021) (nearly weak).
- **Power laws.** Axtell (2001); Gabaix (2009) give a survey of power-law settings. This regularity formalizes instrument strength

Outline

1. Model and Granularity
2. GIV with Known Factor Structure
3. GIV with Unknown Factor Structure
4. Empirical Application and Simulation
5. Conclusion

Model: Recap

Recall the structural model and market clearing:

$$d_t = \phi_d \, p_t + \varepsilon_t, \qquad y_t = e_N \phi_s \, p_t + \Lambda F_t + u_t, \qquad d_t = S' y_t.$$

- Wlog: $\Lambda = [e_N \;\; \tilde{\Lambda}]$ and $F_t = (F_{1t} \;\; \tilde{F}_t)'$
- Idiosyncrasy: $\mathbb{E}[F_t u_t] = 0$ and $\mathbb{E}[(y_{St} - \phi_d p_t) \, u_t] = 0$
- Demeaning kills the time fixed effect: $D_N y_t - \tilde{\Lambda}\tilde{F}_t = D_N u_t$
- GIV: $z_t = S' D_N u_t$

Granularity via the Power Law

Assumption (power-law sizes)

Individual sizes $\mathcal{S}_i$ follow the power-law distribution

$$\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}, \qquad \mu > 0$$

Sizes are independent of $(u_t, F_t, \varepsilon_t)$ and of $\tilde{\boldsymbol{\lambda}}_i$. The cross section $(\mathcal{S}_i, \tilde{\boldsymbol{\lambda}}_i)$ is i.i.d. Loadings have bounded fourth moments.

Granularity via the Power Law

Assumption (power-law sizes)

Individual sizes $\mathcal{S}_i$ follow the power-law distribution

$$\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}, \qquad \mu > 0$$

Sizes are independent of $(u_t, F_t, \varepsilon_t)$ and of $\tilde{\boldsymbol{\lambda}}_i$. The cross section $(\mathcal{S}_i, \tilde{\boldsymbol{\lambda}}_i)$ is i.i.d. Loadings have bounded fourth moments.

- Long-run sizes are determined separately from short-run shocks. Independence is natural
- Shares are constructed as $S_i = \mathcal{S}_i / \sum_{j=1}^{N} \mathcal{S}_j$
- Eigenvalues of $\Omega = \mathbb{E}[u_t u_t']$ are bounded above and away from zero
- The tail index μ determines instrument strength

The Three Regimes: Order of the Instrument

Proposition 1 (Order of the GIV)

Under power-law sizes and bounded eigenvalues of Ω,

$$z_t = S' D_N u_t \;=\; \begin{cases} O_{\mathbb{P}}(1) & \mu \in (0,1), \\ O_{\mathbb{P}}\left(N^{-(1-1/\mu)}\right) & \mu \in (1,2), \\ O_{\mathbb{P}}\left(N^{-1/2}\right) & \mu > 2. \end{cases}$$

Extends Banafti and Lee (2022) ($\mu < 1$) to all μ, following Gabaix (2011).

The Three Regimes: Order of the Instrument

Proposition 1 (Order of the GIV)

Under power-law sizes and bounded eigenvalues of Ω,

$$z_t = S' D_N u_t \; = \begin{cases} O_{\mathbb{P}}(1) & \mu \in (0,1), \\ O_{\mathbb{P}}\left(N^{-(1-1/\mu)}\right) & \mu \in (1,2), \\ O_{\mathbb{P}}\left(N^{-1/2}\right) & \mu > 2. \end{cases}$$

Extends Banafti and Lee (2022) ($\mu < 1$) to all μ, following Gabaix (2011).

- $\mu \in (0,1)$: Heavy tails preserve the instrument
- $\mu > 1$: the instrument dilutes at rate $N^{-(1-1/\mu)}$
- $\mu > 2$: finite second moment, $z_t = O_{\mathbb{P}}(1/\sqrt{N})$ by the LLN

proof

Table of Contents

Estimator and Moment Condition

- **Known Factor Structure:** Suppose Λ is known. For example:
  - Only a time fixed effect: $\Lambda = e_N$
  - Loadings given by observed unit characteristics: $\lambda_i = X_i \dot{\lambda}$
- Then we recover $D_N u_t$. The GIV is

$$z_t = S' D_N u_t$$

- Moment condition $\mathbb{E}[(y_{St} - \phi_d p_t) z_t] = 0$ delivers the estimator

$$\hat{\phi}_d = \phi_d + \frac{T^{-1} \sum_t z_t \varepsilon_t}{T^{-1} \sum_t z_t p_t}$$

Why Large Panels Are Non-Trivial

Even with Λ known, four complications arise as $N \to \infty$:

1. The Herfindahl $S'S$ is random and possibly vanishing (Proposition 1)
2. Shares are random. Random shares interact with random loadings: $S'\tilde{\Lambda}$
3. Share-weighted loading $S'\tilde{\Lambda}$ is a growing sum as $N \to \infty$
4. CLT for $T^{-1/2} \sum_t z_t \varepsilon_t$ is non-standard. It combines a share-weighted cross-sectional sum with time-series dependence

Why Large Panels Are Non-Trivial

Even with Λ known, four complications arise as $N \to \infty$:

1. The Herfindahl $S'S$ is random and possibly vanishing (Proposition 1)
2. Shares are random. Random shares interact with random loadings: $S'\tilde{\Lambda}$
3. Share-weighted loading $S'\tilde{\Lambda}$ is a growing sum as $N \to \infty$
4. CLT for $T^{-1/2}\sum_t z_t \varepsilon_t$ is non-standard. It combines a share-weighted cross-sectional sum with time-series dependence

How each of these plays out depends on the tail index μ. We have three regimes

Strong Identification: Standard $\sqrt{T}$, Wald Works

Theorem 1 (Strong identification)

Suppose idiosyncrasy, power-law sizes with $\mu \in (0,1)$, bounded eigenvalues of Ω, time-series mixing, and standard moment conditions hold. Then, conditional on S, for almost every realization,

$$\frac{\Gamma_{zp}}{\sqrt{V_{z\varepsilon}(S)}} \cdot \sqrt{T}\,[\hat{\phi}_d - \phi_d] \xrightarrow{d} \mathcal{N}(0,1)$$

where $\Gamma_{zp} = \frac{S'\Sigma_u S}{\phi_d - \phi_s} \neq 0$ *a.s.*

- Standard $\sqrt{T}$ rate
- Wald inference is reliable
- This is the classical strong-instrument regime

proof

Nearly Weak: Slower Rate, Wald Still Works

Theorem 2 (Nearly weak identification)

Under the same setting with $\mu > 1$ and $N/T \to 0$, let $\delta = \min(1 - 1/\mu, 1/2)$. Then, conditional on S,

$$\frac{\Gamma_{zp}}{\sqrt{V_{z\varepsilon}(S)}} \cdot \sqrt{T}\,[\hat{\phi}_d - \phi_d] \xrightarrow{d} \mathcal{N}(0,1), \qquad \frac{\Gamma_{zp}}{\sqrt{V_{z\varepsilon}(S)}} = O_{\mathbb{P}}(N^{-\delta})$$

so $\hat{\phi}_d$ concentrates at rate $\sqrt{T}/N^{\delta}$.

- Maps onto the nearly weak framework of Antoine and Renault (2021) and Hahn and Kuersteiner (2002)
- Studentization absorbs the rate. δ does not enter the Wald formula
- Practitioner does not need to know δ to do inference

proof

Weak Identification: $\mu > 2$ and $N/T \to c$

Corollary

If $\mu > 2$ and $N/T \to c > 0$, then $\hat{\phi}_d$ is inconsistent, and identification is weak in the sense of Staiger and Stock (1997).

- From Proposition 1, $z_t = O_{\mathbb{P}}(1/\sqrt{N})$ when $\mu > 2$
- As $N/T \to c > 0$, time-series information no longer compensates for cross-sectional dilution
- The Staiger–Stock weak-IV arise endogenously in the setting

Inference: Wald for $\mu < 2$, AR for $\mu > 2$

- Identification and asymptotic normality settled (Theorems 1–2). Now turn to confidence intervals

Recommendation

- $\mu < 2$: standard Wald confidence interval
- $\mu > 2$: invert the Anderson–Rubin test

Inference: Wald for $\mu < 2$, AR for $\mu > 2$

- Identification and asymptotic normality settled (Theorems 1–2). Now turn to confidence intervals

Recommendation

- $\mu < 2$: standard Wald confidence interval
- $\mu > 2$: invert the Anderson–Rubin test

Wald. t-based CI with

$$\widehat{\mathrm{Avar}}(\hat{\phi}_d) = \frac{1}{T} \cdot \frac{\hat{V}_{z\varepsilon}(S)}{\hat{\Gamma}^2_{zp}(S)}$$

Inference: Wald for $\mu < 2$, AR for $\mu > 2$

- Identification and asymptotic normality settled (Theorems 1–2). Now turn to confidence intervals

Recommendation

- $\mu < 2$: standard Wald confidence interval
- $\mu > 2$: invert the Anderson–Rubin test

Wald. t-based CI with

$$\widehat{\mathrm{Avar}}(\hat{\phi}_d) \;=\; \frac{1}{T} \cdot \frac{\hat{V}_{z\varepsilon}(S)}{\hat{\Gamma}^2_{zp}(S)}$$

Anderson–Rubin. At $H_0 : \phi_d = \phi_0$, invert

$$AR_T(\phi_0) = \frac{T\, g_T(\phi_0)^2}{\hat{V}_{z\varepsilon}(S)} \xrightarrow{d} \chi^2_1, \quad g_T(\phi_0) = \frac{1}{T}\sum_t z_t(y_{St} - \phi_0 p_t)$$

Why AR in $\mu > 2$ Even Under $N/T \to 0$

- Theorem 2 gives asymptotic normality for any $\mu > 1$ under $N/T \to 0$. Wald is fine *asymptotically*
- Finite-sample reliability hinges on the concentration parameter, the signal-to-noise of the first stage

$$\kappa^2_{\text{conc}} = \frac{T\,\Gamma^2_{zp}}{V_{z\varepsilon}(S)}$$

Why AR in $\mu > 2$ Even Under $N/T \to 0$

- Theorem 2 gives asymptotic normality for any $\mu > 1$ under $N/T \to 0$. Wald is fine *asymptotically*
- Finite-sample reliability hinges on the concentration parameter, the signal-to-noise of the first stage

$$\kappa^2_{\text{conc}} = \frac{T\,\Gamma^2_{zp}}{V_{z\varepsilon}(S)}$$

Polynomial floor in N

- $\mu < 2$: κ^2_{conc} has polynomial rate. Wald is safe even at moderate N
- $\mu > 2$: no guarantee of polynomial rate. κ^2_{conc} can grow arbitrarily slowly. Gaussian approximation fails in finite samples

Why AR in $\mu > 2$ Even Under $N/T \to 0$

- Theorem 2 gives asymptotic normality for any $\mu > 1$ under $N/T \to 0$. Wald is fine *asymptotically*
- Finite-sample reliability hinges on the concentration parameter, the signal-to-noise of the first stage

$$\kappa^2_{\text{conc}} = \frac{T\,\Gamma^2_{zp}}{V_{z\varepsilon}(S)}$$

Polynomial floor in N

- $\mu < 2$: κ^2_{conc} has polynomial rate. Wald is safe even at moderate N
- $\mu > 2$: no guarantee of polynomial rate. κ^2_{conc} can grow arbitrarily slowly. Gaussian approximation fails in finite samples

- Simulations support this

Table of Contents

Why We Must Estimate Factor Structure in Practice

- In most applications the factor structure is not observed and must be estimated.
- Demean the data to get a pure factor model

$$D_N y_t = \tilde{\Lambda} \tilde{F}_t + D_N u_t$$

- Large panel PCA estimates the factors and loadings, $\hat{C}_t = \hat{\tilde{\Lambda}} \hat{\tilde{F}}_t$
- GIV with estimated factor structure

$$\hat{z}_t = S'(D_N y_t - \hat{C}_t)$$

Estimating Factor Structure Changes the Variance and impose $\sqrt{T}/N \to 0$

Estimating Λ via PCA preserves the convergence rate and the inference rule.

1. **Rate preserved.** Same $\sqrt{T}$ or $\sqrt{T}/N^{\delta}$ as the known-loadings case (Theorems 3, 4)
2. **Inference rule preserved.** Wald for $\mu < 2$, AR for $\mu > 2$
3. **Growth restriction**. $\sqrt{T}/N \to 0$. Time and cross-section both grow, but the cross-section grows fast enough
4. **Variance picks up a first-stage term:** $V_{z\varepsilon} \to V_{z\varepsilon}^{F}$. Standard errors must use $\widehat{V}_{z\varepsilon}^{F}$

Strong Regime with Estimated Factor Structure

Theorem 3 (Strong regime, estimated loadings)

Suppose the assumptions of Theorem 1 hold, together with $\sqrt{T}/N \to 0$. Then, conditional on S, for almost every realization,

$$\frac{\Gamma_{zp}}{\sqrt{V^F_{z\varepsilon}(S)}} \cdot \sqrt{T}\,[\hat{\phi}^F_d - \phi_d] \xrightarrow{d} \mathcal{N}(0,1)$$

at the same $\sqrt{T}$ rate as the known-loadings GIV.

- Rate unchanged. Variance is now $V^F_{z\varepsilon}(S) \neq V_{z\varepsilon}(S)$

proof

Banafti and Lee (2022) Does Not Apply Under Power Laws

- Banafti and Lee (2022) consider GIV with estimated loadings under the same power-law assumption, restricted to $\mu < 1$
- They conclude the first-stage estimation *does not* affect inference. The argument rests on Assumption 4(iii): an *ex ante* finite k such that $S_{(i),N} = O_{\mathbb{P}}(1/N)$ for all $i > k$: a finite dominant set and a large fringe

Banafti and Lee (2022) Does Not Apply Under Power Laws

- Banafti and Lee (2022) consider GIV with estimated loadings under the same power-law assumption, restricted to $\mu < 1$
- They conclude the first-stage estimation *does not* affect inference. The argument rests on Assumption 4(iii): an *ex ante* finite k such that $S_{(i),N} = O_{\mathbb{P}}(1/N)$ for all $i > k$: a finite dominant set and a large fringe
- Under the power law, for every finite k, $N\, S_{(k),N} \xrightarrow{\mathbb{P}} \infty$ — no such cutoff exists
- Their conclusion does not transfer. Theorem 3 settles the strong regime; nearly weak and weak are new

proof

Nearly Weak with Estimated Factor Structure

Theorem 4 (Nearly weak, estimated loadings)

Suppose the assumptions of Theorem 2 hold, together with $\sqrt{T}/N \to 0$*. Let* $\delta = \min(1 - 1/\mu, 1/2)$*. Then, conditional on* S*,*

$$\frac{\Gamma_{zp}}{\sqrt{V^F_{z\varepsilon}(S)}} \cdot \sqrt{T}\,[\hat{\phi}^F_d - \phi_d] \xrightarrow{d} \mathcal{N}(0,1),$$

so $\hat{\phi}^F_d$ *concentrates at rate* $\sqrt{T}/N^{\delta}$*. If instead* $\mu > 2$ *and* $N/T \to c > 0$*, then* $\hat{\phi}^F_d$ *is inconsistent.*

- Same rate as Theorem 2 (known loadings). Variance is different

proof

Standard Errors Must Absorb the First Stage

- Expand the GIV with unknown loadings

$$\sqrt{T}[\hat{\phi}_d^F - \phi_d] = \Delta^{-1}\left[\underbrace{\frac{1}{\sqrt{T}}\sum_t z_t \varepsilon_t}_{\text{known-loadings term}} + \underbrace{\Delta_\Lambda \sqrt{T}\,\mathrm{vec}(\hat{\Lambda} - \tilde{\Lambda})}_{\text{first-stage term}}\right] + o_{\mathbb{P}}(1)$$

Standard Errors Must Absorb the First Stage

- Expand the GIV with unknown loadings

$$\sqrt{T}[\hat{\phi}_d^F - \phi_d] = \Delta^{-1}\left[\underbrace{\frac{1}{\sqrt{T}}\sum_t z_t \varepsilon_t}_{\text{known-loadings term}} + \underbrace{\Delta_\Lambda \sqrt{T}\,\mathrm{vec}(\hat{\Lambda} - \tilde{\Lambda})}_{\text{first-stage term}}\right] + o_{\mathbb{P}}(1)$$

- Under $\sqrt{T}/N \to 0$, both terms are the same order
- Asymptotic variance picks up an extra component

$$V_{z\varepsilon}^F(S) \;=\; V_{z\varepsilon}(S) \;+\; \Delta_\Lambda\,\mathbb{V}\left[\sqrt{T}\,\mathrm{vec}(\hat{\Lambda} - \tilde{\Lambda})\right]\Delta_\Lambda'$$

- Inference must use SEs that account for this term
- Direction is not signed: the first-stage term can be negatively correlated with the known-loadings term (we will see this empirically)

Table of Contents

Three Commodity Markets

- Country-level supply panels: copper (Bloomberg), crude oil (EIA), natural gas (JODI)
- Demand equation:

$$d_t = \phi_d p_t + X_t'\beta + \varepsilon_t$$

	Copper	Crude Oil	Natural Gas
N (countries)	29	21	27
T (months, usable)	192	611	156
Factors r (ER criterion)	1	2	2

Markets are Concentrated

Method	Copper		Crude Oil		Natural Gas	
	$\hat{\mu}$	95% CI	$\hat{\mu}$	95% CI	$\hat{\mu}$	95% CI
Hill (1975) MLE	0.22	$[0.21, 0.76]$	0.38	$[0.34, 0.77]$	0.20	$[0.17, 0.31]$
Log-rank OLS	0.51	$[0.39, 0.64]$	0.58	$[0.44, 0.71]$	0.29	$[0.24, 0.34]$
Gabaix–Ibragimov (2011)	0.58	$[0.28, 0.87]$	0.65	$[0.26, 1.05]$	0.32	$[0.15, 0.49]$

- All three methods agree: every market sits in the strong regime ($\hat{\mu} < 1$)
- 95% CIs rule out $\mu > 2$ in every case. Crude oil's GI CI reaches into $(1, 2)$ — nearly weak is admissible
- Theory still says Wald

pareto

GIV Flips OLS Signs and Lifts Magnitudes

Estimator	Copper (N=29)	Crude Oil (N=21)	Natural Gas (N=27)
OLS	−0.0512 (0.0226)	0.0041 (0.0033)	0.0010 (0.0046)
GIV, naive SE	−0.1351 (0.0502)	−0.1092 (0.0538)	−0.0556 (0.0338)
GIV, corrected SE	−0.1351 (0.0513)	−0.1092 (0.0548)	−0.0556 (0.0255)

- Same GIV point estimate in both rows: both use the PCA-recovered $\hat{\Lambda}$
- Naive SE treats $\hat{\Lambda}$ as known; corrected SE absorbs first-stage estimation error
- Natural gas: corrected SE is smaller — first-stage correction is negatively correlated with the moment
- The first-stage correction *materially affects inference*

Simulation: Sensitivity of GIV to Granularity

- Calibrate the factor structure to copper (N=29, T=192) and crude oil (N=21, T=611)
- Draw shares from $\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}$, sweep $\mu \in \{0.3, 0.5, 0.8, 1.2, 1.4, 1.8, 2.5, 3.5, 6.0\}$
- True ϕ_d at the GIV point estimates (-0.135 for copper, -0.110 for crude oil). $B = 5{,}000$ replications

Simulation: Sensitivity of GIV to Granularity

- Calibrate the factor structure to copper (N=29, T=192) and crude oil (N=21, T=611)
- Draw shares from $\mathbb{P}(\mathcal{S}_i > s) = c\, s^{-\mu}$, sweep $\mu \in \{0.3, 0.5, 0.8, 1.2, 1.4, 1.8, 2.5, 3.5, 6.0\}$
- True ϕ_d at the GIV point estimates (-0.135 for copper, -0.110 for crude oil). $B = 5{,}000$ replications

What theory predicts

Strong ($\mu < 1$): nominal coverage, large F. Nearly weak ($1 < \mu < 2$): coverage holds but F falls. Weak ($\mu > 2$): F slowly divergent. Wald breaks down.

First-Stage F Collapses as μ Crosses 2

Crude oil calibration: $N=21$, $T=611$, *true* $\phi_d = -0.110$

μ	Median $\hat{\phi}_d$	Bias	RMSE	Coverage	CI Length	Median F
Strong						
0.3	−0.1100	−0.0008	0.0094	0.947	0.0244	128.3
0.8	−0.1100	−0.0002	0.0059	0.947	0.0127	460.2
Nearly weak						
1.2	−0.1100	−0.0010	0.0106	0.949	0.0360	59.9
1.4	−0.1098	−0.0019	0.0157	0.950	0.0522	29.7
1.8	−0.1098	−0.0066	0.1718	0.942	0.0797	13.0
Weak						
2.5	−0.1082	−0.0036	0.6001	0.932	0.1213	5.8
3.5	−0.1036	−0.0060	1.9241	0.930	0.1815	2.6
6.0	−0.0851	0.2616	17.6126	0.936	0.3133	0.9

- Strong and nearly weak: median at the truth, coverage at nominal 0.95
- Weak ($\mu > 2$): bias drifts toward zero, RMSE dominated by tails, $F < 1$ at $\mu = 6$.

Table of Contents

Conclusion

Main message

The power-law tail index μ is the single sufficient statistic for GIV strength in large panels. Three regimes determine consistency, rate, and inference.

- **Theory.** Strong ($\sqrt{T}$), nearly weak ($\sqrt{T}/N^{\delta}$), weak (inconsistent). Wald valid for $\mu < 2$. AR valid for any $\mu > 2$
- **Estimating Factor Structure.** Same rate as the known-factor-structure GIV across all regimes, under $\sqrt{T}/N \to 0$. Inference must account for the first stage
- **Empirics.** Three commodity panels in the strong regime (crude oil borders nearly weak). GIV recovers -0.135, -0.109, -0.056

Thank You

Questions

Appendix: Proof of Proposition 1

- Step 1. $z_t = S' D_N u_t$ has $\mathbb{V}(z_t \mid S) = S' D_N \Omega D_N S$. Eigenvalues of Ω bounded so

$$\mathbb{V}(z_t \mid S) \asymp S' D_N S \asymp S' S$$

- Step 2. Order of $S'S$ under the power law
  - $\mu \in (0,1)$: $\mathcal{S}_i$ has no finite mean. Stable-law limits give $\sum_j \mathcal{S}_j^2 / (\sum_j \mathcal{S}_j)^2 = O_{\mathbb{P}}(1)$
  - $\mu \in (1,2)$: finite mean, infinite variance. $S'S = O_{\mathbb{P}}(N^{-2(1-1/\mu)})$ by Karamata
  - $\mu > 2$: finite variance. LLN: $\sum_j \mathcal{S}_j^2 = O_{\mathbb{P}}(N)$, $(\sum_j \mathcal{S}_j)^2 = O_{\mathbb{P}}(N^2)$. Hence $S'S = O_{\mathbb{P}}(1/N)$
- Step 3. Substitute into Step 1 and take square roots

back

Appendix: Proof of Theorem 1

- Numerator. $T^{-1/2}\sum_t z_t\varepsilon_t$ is a sum of weakly dependent increments. Under strong mixing of size $-(2+\pi)/\pi$ and bounded $8+2\pi$ moments, a Rosenthal-type bound delivers the CLT uniformly in N
- Denominator. $T^{-1}\sum_t z_t p_t \xrightarrow{\mathbb{P}} \Gamma_{zp} = (\phi_d - \phi_s)^{-1} S'\Sigma_u S$, which is $O_{\mathbb{P}}(1)$ a.s. for $\mu < 1$ and bounded away from zero
- Share-weighted loading. By independence of $\mathcal{S}$ and $\boldsymbol{\lambda}$, and bounded fourth moments, $S'\tilde{\Lambda} = O_{\mathbb{P}}(1)$ even under heavy tails
- Combine via the continuous mapping theorem and Slutsky

back

Appendix: Proof of Theorem 2

- Same numerator CLT, but now $V_{z\varepsilon}(S) = O_{\mathbb{P}}(N^{-2\delta})$ by Proposition 1
- Denominator $\Gamma_{zp} = O_{\mathbb{P}}(N^{-2\delta})$ as well, by the same calculation
- Combining

$$\hat{\phi}_d - \phi_d = O_{\mathbb{P}}\left(\frac{N^{\delta}}{\sqrt{T}}\right)$$

- Consistency requires $N^{2\delta}/T \to 0$. Since $2\delta \leq 1$, this is implied by $N/T \to 0$
- Studentized statistic is $\mathcal{N}(0,1)$. The rate is absorbed into the studentization
- Failure mode under $\mu > 2$ and $N/T \to c$

$$\hat{\phi}_d - \phi_d = O_{\mathbb{P}}(\sqrt{N/T}) = O_{\mathbb{P}}(1)$$

back

Appendix: Proof of Theorems 3 and 4

- PCA convergence rates from Bai (2003)

$$\hat{\Lambda} - \tilde{\Lambda} H = O_{\mathbb{P}}\left(\frac{1}{\delta_{NT}^2}\right), \quad \delta_{NT} = \min(\sqrt{N}, \sqrt{T})$$

- Decompose $\hat{z}_t - z_t = S'(\tilde{C}_t - \hat{C}_t)$. Apply Cauchy–Schwarz and the share weighting to bound the order in terms of $S'S$ and the factor estimation error
- Under $\sqrt{T}/N \to 0$, the estimation contribution is the same order as the known-loadings IV term in the numerator
- Joint CLT delivers the limiting distribution. The variance has two pieces

$$V_{z\varepsilon}^F(S) = V_{z\varepsilon}(S) + \Delta_\Lambda \, \mathbb{V}(\sqrt{T}\,\mathrm{vec}(\hat{\Lambda} - \tilde{\Lambda}))\, \Delta_\Lambda'$$

- Rate is unchanged. Variance must be corrected

back

Appendix: Why BL's Assumption 4(iii) Fails Under the Power Law

BL Assumption 4(iii)

Order $\mathcal{S}_{(1)} \geq \cdots \geq \mathcal{S}_{(N)}$. There exists finite k such that $S_{(i),N} = O_{\mathbb{P}}(1/N)$ for all $i > k$.

Proposition (this paper). Under the power law with $\mu < 1$, $N\, S_{(k),N} \xrightarrow{\mathbb{P}} \infty$ for every finite k — so no such partition exists.

Sketch.

- Pitman and Yor (1997): $\big(S_{(1),N}, S_{(2),N}, \ldots\big) \xrightarrow{d} (p_1, p_2, \ldots) \sim \mathrm{PD}(\mu, 0)$, with $p_k > 0$ a.s. for every finite k
- Pick η with $\mathbb{P}(p_k > \eta) > 1 - \varepsilon$. Portmanteau: $\liminf_N \mathbb{P}\big(S_{(k),N} > \eta\big) > 1 - \varepsilon$
- For $N\eta > M$, $\{S_{(k),N} > \eta\} \subseteq \{NS_{(k),N} > M\}$; take $M \to \infty$, $\varepsilon \downarrow 0$. $\square$

Consequence: BL's dropped term (Lemma 2, Term IV) needs this assumption to be $o_{\mathbb{P}}(1)$

back

Appendix: Anderson–Rubin in Full

- Sample moment at $H_0 : \phi_d = \phi_0$

$$g_T(\phi_0) = \frac{1}{T}\sum_{t=1}^{T} \hat{z}_t(y_{St} - \phi_0 p_t)$$

- Under H_0, $g_T(\phi_0) = T^{-1}\sum_t \hat{z}_t\varepsilon_t$. CLT delivers

$$\sqrt{T}\, g_T(\phi_0)/\sqrt{\hat{V}^F_{z\varepsilon}(S)} \xrightarrow{d} \mathcal{N}(0,1)$$

- AR statistic and $1-\alpha$ confidence set

$$AR_T(\phi_0) = \frac{T\, g_T(\phi_0)^2}{\hat{V}^F_{z\varepsilon}(S)} \xrightarrow{d} \chi^2_1, \qquad \mathcal{C}^{AR}_{1-\alpha} = \{\phi_0 : AR_T(\phi_0) \le \chi^2_{1,1-\alpha}\}$$

- Pivotal limit. No condition on κ^2_{conc}. Valid uniformly under $\mu > 2$ and any admissible (N, T)
- Use $\hat{V}^F_{z\varepsilon}(S)$ (with the first-stage correction)

Appendix: A Recipe for the Practitioner

1. **Estimate** $\hat{\mu}$ from cross-sectional shares (Gabaix–Ibragimov or Hill)
2. **Estimate** Λ **and** F by PCA on demeaned y_t. Pick r by the ER criterion (Ahn and Horenstein, 2013)
3. **Construct** $\hat{z}_t = S'(D_N y_t - \hat{C}_t)$
4. **Run 2SLS** of y_{St} on p_t with $\hat{z}_t$ as instrument
5. **Compute SE with the first-stage correction**
6. **Inference**
    - If $\hat{\mu} < 2$ (CI too): Wald
    - If $\hat{\mu} > 2$ (or CI reaches above 2): Anderson–Rubin

Appendix: Estimating the Tail Index

- **Hill MLE (1975).**

$$\hat{\mu}_{\text{Hill}} = \frac{N}{\sum_{i=1}^{N} \log(S_i / S_{\min})}$$

  Bootstrap CI from 5,000 replications

- **Log-rank OLS.** Regress log(rank) on $\log(S)$
- **Gabaix–Ibragimov (2011).** Shift by 1/2 to remove finite-sample bias

$$\log(\text{rank} - 1/2) = a - \mu \log S, \qquad \mathrm{SE}(\hat{\mu}_{\text{GI}}) = \sqrt{2/N}\,\hat{\mu}_{\text{GI}}$$

- Recommendation: report all three, headline with GI. Hill is sharper when only the top tail follows the power law

back

Appendix: Simulation, Copper Calibration

back

μ	Median $\hat{\phi}_d$	Bias	RMSE	Coverage	CI Length	Median F
0.3	−0.1352	−0.0035	0.0689	0.946	0.0504	47.4
0.5	−0.1350	−0.0015	0.0165	0.952	0.0378	87.3
0.8	−0.1351	−0.0007	0.0092	0.942	0.0278	162.2
1.2	−0.1353	−0.0057	0.0423	0.941	0.0876	15.9
1.4	−0.1348	−0.0157	0.3186	0.929	0.1306	7.6
1.8	−0.1315	0.0287	4.5410	0.926	0.1979	3.3
2.5	−0.1163	0.0298	1.8983	0.918	0.3033	1.4
3.5	−0.0993	−0.0779	4.5276	0.925	0.4095	0.8
6.0	−0.0807	0.8334	52.6726	0.931	0.5169	0.5

- Same qualitative picture as the crude oil calibration
- Smaller T amplifies the deterioration: F at $\mu = 1.4$ is only 7.6 (vs 29.7 for crude oil)

Seung C Ahn and Alex R Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227, 2013. ISSN 0012-9682.

Iñaki Aldasoro, Paula Beltrán, Federico Grinberg, and Tommaso Mancini-Griffoli. The macro-financial effects of international bank lending on emerging markets. *Journal of International Economics*, 142:103733, 2023.

Bertille Antoine and Eric Renault. GMM with Nearly-Weak Identification. *Econometrics and Statistics*, 2021.

Timothy B Armstrong. Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica*, 84(5):1961–1980, 2016. ISSN 0012-9682.

Robert L Axtell. Zipf Distribution of U.S. Firm Sizes. *Science*, 293(5536):1818–1820, 2001. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1062081. URL `https://www.science.org/doi/10.1126/science.1062081`.

Jushan Bai. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1): 135–171, 2003. ISSN 0012-9682. doi: https://doi.org/10.1111/1468-0262.00392. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00392`.

Jushan Bai and Serena Ng. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221, 2002. ISSN 0012-9682. doi: https://doi.org/10.1111/1468-0262.00273. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00273`.

Saman Banafti and Tae-Hwy Lee. Inferential Theory for Granular Instrumental Variables in High Dimensions. *arXiv preprint arXiv:2201.06605*, 2022.

Christiane Baumeister and James D Hamilton. A Full-Information Approach to Granular Instrumental Variables. Technical report, Working Paper, UCSD, 2023.

Gabriel Chodorow-Reich, Andra Ghent, and Valentin Haddad. Asset insulators. *The Review of Financial Studies*, 34(3):1509–1539, 2021. ISSN 0893-9454.

References III

Xavier Gabaix. Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1): 255–294, 2009. doi: 10.1146/annurev.economics.050708.142940. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev.economics.050708.142940`.

Xavier Gabaix. The Granular Origins of Aggregate Fluctuations. *Econometrica*, 79(3):733–772, 2011. ISSN 0012-9682. doi: 10.3982/ECTA8769.

Xavier Gabaix and Ralph S J Koijen. Granular Instrumental Variables. *Journal of Political Economy*, 132(7):728743, 7 2024. ISSN 0022-3808, 1537-534X. doi: 10.1086/728743. URL `https://www.journals.uchicago.edu/doi/10.1086/728743`.

Sigurd Galaasen, Rustam Jamilov, Ragnar Juelsrud, and Hélène Rey. Granular Credit Risk. Technical Report w27994, National Bureau of Economic Research, Cambridge, MA, 10 2020. URL `http://www.nber.org/papers/w27994.pdf`.

Jinyong Hahn and Guido Kuersteiner. Discontinuities of weak instrument limiting distributions. *Economics Letters*, 75(3):325–331, 2002. ISSN 0165-1765.

References IV

Yueran Ma, Teodora Paligorova, and José-Luis Peydro. Expectations and bank lending. *University of Chicago, Unpublished Working Paper*, 2021.

Jim Pitman and Marc Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, pages 855–900, 1997. ISSN 0091-1798.

Eric Qian. Heterogeneity-robust granular instruments. *arXiv preprint arXiv:2304.01273*, 2023.

Douglas Staiger and James H Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586, 1997. ISSN 00129682, 14680262. doi: 10.2307/2171753. URL `http://www.jstor.org/stable/2171753`.