Instrumental Variable Estimation via a Continuum of Instruments with an Application to Estimating the Elasticity of Intertemporal Substitution in Consumption

Carlos Velasco Xuexin Wang

UC3M

Xiamen University

Venice Time Series Workshop, April 2026

Motivating Example: The EIS Estimation

- The log-linearized Euler equation implied by an asset pricing model:

$$\mathbb{E}\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathcal{F}_t\right] = 0, \tag{1}$$

  or

$$\mathbb{E}\left[r_{t+1} - \varpi - 1/\psi \triangle c_{t+1} | \mathcal{F}_t\right] = 0, \tag{2}$$

  where ψ is the **Elasticity of Intertemporal Substitution**, $\triangle c_{t+1}$ is the growth rate of consumption, r_{t+1} is the gross return on some asset, α and ϖ are constants, and $\mathcal{F}_t$ is the agent's information set at time t.
- **Many exogenous variables**: lag of growth rate of consumption, asset returns (stock returns, interest rate), price-dividend ratio, ...
- **Weak instruments**: linear IVs are weak (Yogo, 2004; Ascari et al., 2021), what about **nonlinear IVs**?
- Many other examples in microeconomics, dynamic panel data models, etc.

Overview

- Introduce nuisance-parameter-free IV estimators for linear regressions by employing a **continuum of IVs** coupled with a novel **non-integrable weighting function** in a minimum distance objective function.
  - The new IV estimator is labeled as WCIV, Weighted Continuum of IVs.
  - WCIV has analytical formula and a natural jackknife form, resembling the conventional k-class IV estimators.
  - Robust to weak instruments and heteroskedasticity of unknown form.
  - Robust to the high-dimensionality of exogenous variables.
  - Justification for dependent data as for our application.
  - Excellent finite sample properties.
- **Estimate the EIS** by the newly proposed IV estimators, using macro data sets from the US.
  - The WCIV and WCIVF estimates of the EIS well exceed 1 and are statistically different from zero.
  - Findings persist across model transformations, distinct sets of IVs, data structures, and data ranges.
  - The results suggest a resolution to a long-standing discrepancy between EIS values in many model calibrations and those estimated via macro data sets.

Overview

- Introduce nuisance-parameter-free IV estimators for linear regressions by employing a **continuum of IVs** coupled with a novel **non-integrable weighting function** in a minimum distance objective function.
  - The new IV estimator is labeled as WCIV, Weighted Continuum of IVs.
  - WCIV has analytical formula and a natural jackknife form, resembling the conventional k-class IV estimators.
  - Robust to weak instruments and heteroskedasticity of unknown form.
  - Robust to the high-dimensionality of exogenous variables.
  - Justification for dependent data as for our application.
  - Excellent finite sample properties.
- **Estimate the EIS** by the newly proposed IV estimators, using macro data sets from the US.
  - The WCIV and WCIVF estimates of the EIS well exceed 1 and are statistically different from zero.
  - Findings persist across model transformations, distinct sets of IVs, data structures, and data ranges.
  - The results suggest a resolution to a long-standing discrepancy between EIS values in many model calibrations and those estimated via macro data sets.

Background: IV Estimation

- **2SLS** becomes unreliable in the case of weak IVs.
- The use of many weak instruments can help to enhance the estimation accuracy of some IV estimators, but the asymptotic normality approximation can be poor.
- **LIML** and bias-corrected 2SLS are consistent under many weak instruments (Chao and Swanson, 2005), but LIML can be inconsistent for many weak instruments under heteroskedasticity of unknown form (Hausman et al. 2012).
- **HLIM** and HFUL proposed by Hausman et al. (2012) are consistent in the presence of heteroskedasticity and many weak instruments.
- Determining an appropriate number of weak instruments for the standard many weak IV estimators is an extremely challenging task.
  - The linear combination of an increasing number of instruments approximates the reduced form as the sample size approaches infinity.
  - The asymptotic properties of these estimators crucially depend on the number of weak IVs and the unknown concentration parameter.
  - High-dimensional methods need regularization and bias-correction (d-d).

Related Literature

- **Consistent ICM tests** of specification: Bierens (1982), Bierens (1990) and Bierens and Ploberger (1997), Stute (1997), Escanciano (2006), Tan and Zhu (2019, 2021), Jiang and Tsyawo (2026).
- **Consistent estimation** with minimum identifying conditions for nonlinear models: Domínguez and Lobato (2004), Hsu and Kuan (2011), Wang (2018).
- **Continuum GMM**: Carrasco and Florens (2000).
- Integrable weighting function for a continuum of IVs: Escanciano (2018), Antoine and Lavergne (2014).
- Testing with non-integrable weighting function: Distance Covariance (DC: Székely et al., 2007; Fokianos and Pitsillou, 2017; Davis et al., 2018); Martingale Difference Divergence (MDD) measure (Shao and Zhang, 2014).
- Estimation w/ non-integrable weighting function: Tsyawo (2023) uses MDD for a related ill-posed estimate. Velasco and Wang (2025) for time series.

Plan of talk

- Model and estimates
- Distance measures
- Asymptotics
- Comparison with other estimates
- Simulations
- EIS estimation

Model Setup

- Linear Structural Form equation

$$y_t = \alpha_0 + \mathbf{Y}_t' \boldsymbol{\beta}_0 + \varepsilon_{0t}, \quad t = 1, ..., n.$$

- The true parameter $\boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\beta}_0')' \subset \mathbb{R}^{1+p}$.
- $\mathbf{X}_t$ is a $q \times 1$ vector of exogenous instruments, such that

$$\mathbb{E}\left(\varepsilon_{0t} | \mathbf{X}_t\right) = 0 \ \ a.s.$$

- $q \geq p$ or $q < p$ (unknown nonlinear RF).
- $\mathbf{Y}_t$ may contain included exogenous variables. Correspondingly, $\mathbf{X}_t$ may contain these exogenous variables, in addition to (excluded) IVs.
- The exogeneity condition is a martingale difference: Linearized Euler equation; the new Keynesian Phillips curve; dynamic panel data models.
- Continuum of IVs:

$$\exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right), \quad \text{for all } \boldsymbol{\tau} \in \mathbb{R}^q.$$

- Continuum of unconditional moment restrictions:

$$\mathbb{E}\left\{\left[y_t - \mu_y - \boldsymbol{\beta}_0' \left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right] \exp\left(i \left\langle \boldsymbol{\tau}, \mathbf{X}_t \right\rangle\right)\right\} = 0, \text{ for all } \boldsymbol{\tau} \in \mathbb{R}^q, \qquad (3)$$

where $\mu_y = \mathbb{E}\left(y_t\right)$ and $\boldsymbol{\mu}_Y = \mathbb{E}\left(\mathbf{Y}_t\right)$.

The WCIV and WCIVF Estimators

- Let $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{t=1}^{n} Y_t$

$$\tilde{\mathbf{Y}} = \left(\tilde{\mathbf{Y}}_1, ..., \tilde{\mathbf{Y}}_n\right)' = \left(\mathbf{Y}_1 - \bar{\mathbf{Y}}, ..., \mathbf{Y}_n - \bar{\mathbf{Y}}\right)',$$

$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t$,

$$\tilde{\mathbf{y}} = (\tilde{y}_1, ..., \tilde{y}_T)' = (y_1 - \bar{y}, ..., y_T - \bar{y})'.$$

- Let $\mathbf{D}$ be a square matrix of size n and D_{jk} denote its (j,k)th element,

$$D_{jk} = -\|\mathbf{X}_j - \mathbf{X}_k\|, \quad j,k = 1, ..., n.$$

- The estimates minimizing our objective function are

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{WCIV} &= \left[\tilde{\mathbf{Y}}' \left(\mathbf{D} - \hat{\lambda}_{WCIV} \mathbf{I}_n\right) \tilde{\mathbf{Y}}\right]^{-1} \left[\tilde{\mathbf{Y}}' \left(\mathbf{D} - \hat{\lambda}_{WCIV} \mathbf{I}_n\right) \tilde{\mathbf{y}}\right] \quad (4) \\ \hat{\alpha}_{WCIV} &= \bar{y} - \bar{\mathbf{Y}} \hat{\boldsymbol{\beta}}_{WCIV}. \end{aligned}$$

where $\mathbf{I}_n$ is the identity matrix of size n.

- $\hat{\lambda}_{WCIV}$ is the minimum value of the objective function, which is computed as the smallest eigenvalue of $\left(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}\right)^{-1} \check{\mathbf{Y}}'\mathbf{D}\check{\mathbf{Y}}$ with $\check{\mathbf{Y}} = \left[\tilde{\mathbf{y}} \ \tilde{\mathbf{Y}}\right]$.
- WCIVF writes for a constant C as (4) with

$$\hat{\lambda}_{WCIVF} = \left[\hat{\lambda}_{WCIV} - \left(1 - \hat{\lambda}_{WCIV}\right) C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{WCIV}\right) C/n\right].$$

IVs estimates

- WCIV and WCIVF estimators have a *Jackknife* form automatically ($\mathbf{D}$ is diagonal: $D_{jj} = 0$), so are robust to heteroskedasticity of unknown form.
- They resemble conventional **k-class IV estimators**: WCIV to HLIM, WCIVF to HFUL:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{Y}^*\right]^{-1}\left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{y}\right],$$

where $\mathbf{Y}^* = [\boldsymbol{\iota}, \mathbf{Y}]$, with $\boldsymbol{\iota}$ representing a vector of ones, and $\mathbf{P}$ is a matrix that depends on an $n \times m$ matrix $\mathbf{X}$ of IV observations, $rank\,(\mathbf{X}) = q \geq p + 1$.

IVs estimates

- WCIV and WCIVF estimators have a *Jackknife* form automatically ($\mathbf{D}$ is diagonal: $D_{jj} = 0$), so are robust to heteroskedasticity of unknown form.
- They resemble conventional **k-class IV estimators**: WCIV to HLIM, WCIVF to HFUL:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{Y}^{*}\right]^{-1}\left[\mathbf{Y}^{*\prime}\left(\mathbf{P} - \hat{\lambda}\mathbf{I}_n\right)\mathbf{y}\right],$$

where $\mathbf{Y}^{*} = [\boldsymbol{\iota}, \mathbf{Y}]$, with $\boldsymbol{\iota}$ representing a vector of ones, and $\mathbf{P}$ is a matrix that depends on an $n \times m$ matrix $\mathbf{X}$ of IV observations, $rank\,(\mathbf{X}) = q \geq p + 1$.

  - **2SLS** $\rightarrow$ $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, with $\hat{\lambda} = 0$;
  - **JIVE** $\rightarrow$ $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' - diag\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$ and $\hat{\lambda} = 0$;
  - **LIML** $\rightarrow$ $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ and $\hat{\lambda}$ is the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^{*}\right)^{-1}\check{\mathbf{Y}}^{*\prime}\mathbf{P}\check{\mathbf{Y}}^{*}$ with $\check{\mathbf{Y}}^{*} = [\mathbf{y}, \mathbf{Y}^{*}]$;
  - **HLIM** $\rightarrow$ $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' - diag\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)$ and $\hat{\lambda}$ is the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^{*}\right)^{-1}\check{\mathbf{Y}}^{*\prime}\mathbf{P}\check{\mathbf{Y}}^{*}$.
  - **HFUL** employs in HLIM expression

$$\hat{\lambda}_{HFUL} = \left[\hat{\lambda}_{HLIM} - \left(1 - \hat{\lambda}_{HLIM}\right)C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{HLIM}\right)C/n\right].$$

WMD estimates: Antoine and Lavergne (2014)

- WCIV and WCIVF bear some similarities with the Weighted Minimum Distance (WMD) estimators.
- $\mathbf{K}$ is a $n \times n$ matrix, such that

$$K_{jk} = \exp\left(-\left\|\mathbf{X}_j - \mathbf{X}_k\right\|^2/2\right), \; j \neq k,$$

  and $K_{jj} = 0$ for $j, k = 1, ..., n$.
- $\mathbf{Y}^* = [\boldsymbol{\iota}, \mathbf{Y}]$, with $\boldsymbol{\iota}$ representing a vector of ones:

$$\hat{\boldsymbol{\theta}}_{WMD} = \left[\mathbf{Y}^{*\prime}\left(\mathbf{K} - \hat{\lambda}_{WMD}\mathbf{I}_n\right)\mathbf{Y}^*\right]^{-1}\left[\mathbf{Y}^{*\prime}\left(\mathbf{K} - \hat{\lambda}_{WMD}\mathbf{I}_n\right)\mathbf{y}\right], \quad (5)$$

  where $\hat{\lambda}_{WMD}$ is the minimum value of the objective function, which can be computed as the smallest eigenvalue of $\left(\check{\mathbf{Y}}^{*\prime}\check{\mathbf{Y}}^*\right)^{-1}\check{\mathbf{Y}}^{*\prime}\mathbf{K}\check{\mathbf{Y}}^*$, $\check{\mathbf{Y}}^* = [\mathbf{y}, \mathbf{Y}^*]$.
- WMDF, the Fuller-style variant of WMD, uses

$$\hat{\lambda}_{FWMD} = \left[\hat{\lambda}_{WMD} - \left(1 - \hat{\lambda}_{WMD}\right)C/n\right] / \left[1 - \left(1 - \hat{\lambda}_{WMD}\right)C/n\right].$$

Distance Measures

- Population moments indexed by parameter value $\boldsymbol{\beta}$

$$h\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)=\mathbb{E}\left\{\left[y_t-\mu_y-\boldsymbol{\beta}'\left(\mathbf{Y}_t-\boldsymbol{\mu}_Y\right)\right]\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right)\right\},$$

  and its sample analog

$$h_n\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)=\frac{1}{n}\sum_{t=1}^{n}\left(\tilde{y}_t-\boldsymbol{\beta}'\tilde{\mathbf{Y}}_t\right)\exp\left(i\left\langle\boldsymbol{\tau},\mathbf{X}_t\right\rangle\right), \tag{6}$$

  where $\tilde{y}_t=y_t-\bar{y}$.

- To fully utilize the continuum of moments (6), we formulate the following **distance measure**:

$$\int_{\mathbb{R}^q}\left|h\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)\right|^2W\left(d\boldsymbol{\tau}\right),$$

  where $W\left(\cdot\right)$ is an arbitrary positive weighting function for which the integrals exist, and its sample analog is

$$\int_{\mathbb{R}^q}\left|h_n\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)\right|^2W\left(d\boldsymbol{\tau}\right).$$

A Novel Non-integrable Weighting Function

- Székely, Rizzo and Bakirov (2007):

$$W\left(d\boldsymbol{\tau}\right) = \frac{\Gamma\left(\left(q+1\right)/2\right)}{\pi^{(q+1)/2}}\left(\left\|\boldsymbol{\tau}\right\|^{q+1}\right)^{-1} d\boldsymbol{\tau}.$$

- $$\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)\right|^2 \frac{(q-1)!!}{(2\pi)^{q/2}\left\|\boldsymbol{\tau}\right\|^{q+1}} d\boldsymbol{\tau} = \int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta},\boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) \qquad (7)$$

  is the Martingale Difference Divergence (MDD) proposed by Shao and Zhang (2014), enjoying a scale invariance property.

The Non-integrable Weighting function: small $\boldsymbol{\tau}$

- The weighting values of $\boldsymbol{\omega}(\boldsymbol{\tau})$ in a neighborhood of the origin tend to infinite.
- $\mathbb{E}\left(\left|\sqrt{n}h_n\left(\boldsymbol{\beta}_0, \boldsymbol{\tau}\right)\right|^2\right) \to 1 - |\varphi_{\mathbf{X}}(\boldsymbol{\tau})|^2 \approx C\left\|\boldsymbol{\tau}\right\|^2 \to 0$ as $\boldsymbol{\tau} \to 0$.

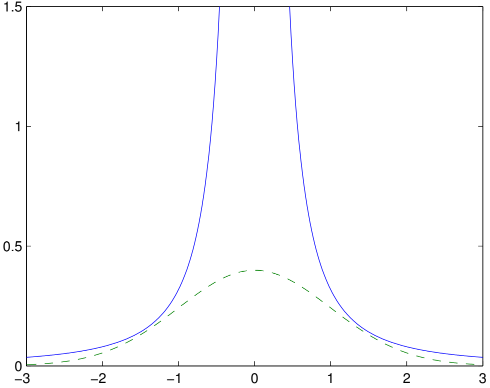

Figure: Standard normal density function (dashed curve) vs. $1/(\pi\|\boldsymbol{\tau}\|^2)$ (solid curve)

The Non-integrable Weighting function: large q

- The weighting function $\boldsymbol{\omega}$ is an increasing function of q in a neighborhood of the origin of $\boldsymbol{\tau}$, indicating a "bless of dimensionality" in terms of estimation efficiency:

$$\boldsymbol{\omega}(\boldsymbol{\tau}) \approx \frac{c\sqrt{e}}{\|\boldsymbol{\tau}\|}\left(\frac{q-1}{2\pi e\,\|\boldsymbol{\tau}\|^2}\right)^{q/2} \to \infty$$

  as $q \to \infty$, where $c = \sqrt{\pi}$ for $q-1$ is even, and $c = \sqrt{2}$ for $q-1$ is odd.

- A q-dimensional **standard normal** *pdf* is a decreasing function of q, given a $\|\boldsymbol{\tau}\|$ and

$$\max_{\boldsymbol{\tau}} \phi_q(\boldsymbol{\tau}) = \phi_q(\mathbf{0}) = (2\pi)^{-q/2} \to 0$$

  as $q \to \infty$.

The Non-integrable Weighting function: MDD

Lemma 1

When $\mathbb{E}\left(\varepsilon_t^2\right) < \infty, \mathbb{E}\left\|\mathbf{X}_t\right\|^2 < \infty,$

$$\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) = -\mathbb{E}\left[\begin{array}{c} \left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right) \\ \times\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\| \end{array}\right].$$

where $\left(y_t^+, \mathbf{Y}_t^+, \mathbf{X}_t^+\right)$ *is an i.i.d. copy of* $(y_t, \mathbf{Y}_t, \mathbf{X}_t)$*.*

IV transformation interpretation:

$$\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right) = \mathbb{E}\left[\left(y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right) m(\mathbf{X}_t; \boldsymbol{\beta})\right] \geq 0,$$

where

$$m(\mathbf{X}_t; \boldsymbol{\beta}) := -\mathbb{E}^+\left[\left(y_t^+ - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t^+ - \boldsymbol{\mu}_Y\right)\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right].$$

The Non-integrable Weighting function: Identification

Beneath the specialties of the non-integrable weighting function lies a fundamental property regarding parameter identification.

Proposition 2

When $\mathbb{E}\left(\varepsilon_{0t}^2\right) < \infty, \mathbb{E}\left\|\mathbf{X}_t\right\|^2 < \infty$, *for a fixed* q *or* $q \to \infty$,

$$\mathbb{E}\left(\varepsilon_{0t} | \mathbf{X}_t\right) = 0 \ a.s.$$

if and only if

$$\mathbb{E}\left[\varepsilon_{0t}\varepsilon_{0t}^{+}\left\|\mathbf{X}_t - \mathbf{X}_t^{+}\right\|\right] = 0 \ and \ \mathbb{E}\left(\varepsilon_{0t}\right) = 0.$$

where $\left(\varepsilon_{0t}^{+}, \mathbf{X}_t^{+}\right)$ *is an i.i.d. copy of* $\left(\varepsilon_{0t}, \mathbf{X}_t\right)$.

This proposition indicates that $\boldsymbol{\beta}_0$ is always identifiable even as $q \to \infty$, implying the high-dimensional robustness of WCIV and WCIVF.

Minimum Distance Objective Functions

From previous identification result,

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} \frac{\int_{\mathbb{R}^q} \left|h\left(\boldsymbol{\beta}, \boldsymbol{\tau}\right)\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right)}{\mathbb{E}\left(\left[y_t - \mu_y - \boldsymbol{\beta}'\left(\mathbf{Y}_t - \boldsymbol{\mu}_Y\right)\right]^2\right)},$$

$$\alpha_0 = \mu_y - \boldsymbol{\beta}_0'\boldsymbol{\mu}_Y.$$

Then, the WCIV estimator is defined as

$$\hat{\boldsymbol{\beta}}_{WCIV} = \arg\min_{\boldsymbol{\beta}} \left[\frac{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)' \mathbf{D} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}{\left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)' \left(\tilde{\mathbf{y}} - \tilde{\mathbf{Y}}\boldsymbol{\beta}\right)}\right],$$

$$\hat{\alpha}_{WCIV} = \bar{y} - \hat{\boldsymbol{\beta}}_{WCIV}'\bar{\mathbf{Y}}.$$

Model: Many Weak Instruments (Semi-Strong identification)

$$y_t = \alpha_0 + \mathbf{Y}_t' \boldsymbol{\beta}_0 + \varepsilon_{0t}.$$

Assumption (Hausman et al. (2012))

$$\mathbf{Y}_t = \frac{\mathbf{R}_n \mathbf{f}(\mathbf{X}_t)}{\sqrt{n}} + \boldsymbol{\eta}_t, \qquad \mathbf{R}_n = \tilde{\mathbf{R}} diag(r_{1,n}, ..., r_{q,n})$$

- $\tilde{\mathbf{R}}$ *is a* $q \times q$ *nonsingular matrix, for each* j, $r_{j,n} = \sqrt{n}$ *or* $r_{j,n}/\sqrt{n} \to 0$, $r_n = \min_{1 \le j \le q} r_{j,n} \to \infty$.

$\to$ *(Unknown) and different rates of convergence.*

- $\mathbf{f} : \mathbb{R}^q \to \mathbb{R}^p$ *is a measurable function:*

$\to$ *(Unknown) and not estimated Reduced Form.*

$\frac{1}{n^2} \sum_{j,k} \tilde{\mathbf{f}}(\mathbf{X}_j) D_{jk} \tilde{\mathbf{f}}(\mathbf{X}_k)'$ *is finite and p.d.*, $\tilde{\mathbf{f}}(\mathbf{X}_t) = \mathbf{f}(\mathbf{X}_t) - \frac{1}{n}\sum_{j=1}^{n} \mathbf{f}(\mathbf{X}_j)$.
$\vartheta_{\min}\left(\sum_{t=1}^{n} \mathbf{f}(\mathbf{X}_t)\mathbf{f}(\mathbf{X}_t)'/n\right) \ge 1/C$ *for* n *sufficiently large.*

Assumption

$\left\{(\varepsilon_{0t}, \mathbf{X}_t', \boldsymbol{\eta}_t')'\right\}$ *is strictly stationary and ergodic,* $\mathbb{E}(\varepsilon_{0t}|\mathbf{I}_t) = 0, \mathbb{E}(\boldsymbol{\eta}_t|\mathbf{X}_t) = 0$.
$\mathbb{E}\left(\|\mathbf{X}_t\|^2\right) < C$, $\sup_t \mathbb{E}\left(\varepsilon_{0t}^2|\mathbf{X}_t\right) < C$, $\sup_t \mathbb{E}\left(\|\boldsymbol{\eta}_t\|^2 |\mathbf{X}_t\right) < C$.
$Var\left((\varepsilon_{0t}, \boldsymbol{\eta}_t')' |\mathbf{X}_t\right) = diag(\Phi_t^*, 0)$, *and* $\vartheta_{\min}\left(\sum_{t=1}^{n} \Phi_t^*/n\right) \ge 1/C$ *for* n *large.*

Estimates Consistency

Theorem 3

Under Assumptions 1-2, for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$, $\hat{\alpha} = \hat{\alpha}_{WCIV}$ *or* $\hat{\alpha}_{WCIVF}$,

$$\mathbf{R}_n' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0,$$

$$\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\boldsymbol{\beta}}) \xrightarrow{p} \boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\beta}_0).$$

Asymptotic Normality

Variance of the score of the loss function, $n^{-1}\sum_{j\neq k}^{n}\bar{\varepsilon}_j(\boldsymbol{\theta})\bar{\varepsilon}_k(\boldsymbol{\theta})\left\|\mathbf{X}_j-\mathbf{X}_k\right\|$:

$$\mathbf{V}\left(\boldsymbol{\theta}\right)=\mathbf{V}_1\left(\boldsymbol{\theta}\right)+\mathbf{V}_2\left(\boldsymbol{\theta}\right)-\mathbf{V}_3\left(\boldsymbol{\theta}\right)-\mathbf{V}_3\left(\boldsymbol{\theta}\right)',$$

where

$$\mathbf{V}_1\left(\boldsymbol{\theta}\right)=\mathbb{E}\left[\left(y_t-\alpha-\boldsymbol{\beta}'\mathbf{Y}_t\right)^2\left(\mathbf{f}\left(\mathbf{X}_t^{+}\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)\left(\mathbf{f}\left(\mathbf{X}_t^{++}\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)'\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\left\|\mathbf{X}_t-\mathbf{X}_t^{++}\right\|\right]$$

$$\mathbf{V}_2\left(\boldsymbol{\theta}\right)=\mathbb{E}\left[\left(y_t-\alpha-\boldsymbol{\beta}'\mathbf{Y}_t\right)^2\right]\mathbb{E}\left[\left(\mathbf{f}\left(\mathbf{X}_t\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\right]\mathbb{E}\left[\left(\mathbf{f}\left(\mathbf{X}_t\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)'\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\right]$$

$$\mathbf{V}_3\left(\boldsymbol{\theta}\right)=\mathbb{E}\left[\left(y_t-\alpha-\boldsymbol{\beta}'\mathbf{Y}_t\right)^2\left(\mathbf{f}\left(\mathbf{X}_t^{+}\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\right]\mathbb{E}\left[\left(\mathbf{f}\left(\mathbf{X}_t\right)-\boldsymbol{\mu}_{\mathbf{f}}\right)'\left\|\mathbf{X}_t-\mathbf{X}_t^{+}\right\|\right]$$

with $\boldsymbol{\mu}_{\mathbf{f}}=\mathbb{E}[\mathbf{f}\left(\mathbf{X}_t\right)]$.

Assumption

$\sup_t\mathbb{E}\left(\varepsilon_{0t}^4|\mathbf{X}_t\right)<C$, $\sup_t\mathbb{E}\left(\left\|\boldsymbol{\eta}_t\right\|^4|\mathbf{X}_t\right)<C$ *a.s.*, $\mathbb{E}\left\|\mathbf{X}_t\right\|^4<\infty$, $\mathbb{E}\left\|\mathbf{f}\left(\mathbf{X}_t\right)\right\|^4<\infty$.
$\mathbf{V}\left(\boldsymbol{\theta}_0\right)$ *is positive definite.*

- Accounts for centering of data.
- No restrictions on conditional heteroskedasticity (or dynamics).

Asymptotic Normality

Theorem 4

Under Assumptions 1-3, for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$,

$$\mathbf{R}_n' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \xrightarrow{d} \mathbf{N} \left(0, \boldsymbol{\Pi}^{-1} \mathbf{V} \left(\boldsymbol{\theta}_0 \right) \boldsymbol{\Pi}^{-1} \right),$$

where

$$\boldsymbol{\Pi} = -\mathbb{E} \left[\left(\mathbf{f} \left(\mathbf{X}_t \right) - \boldsymbol{\mu}_{\mathbf{f}} \right) \left(\mathbf{f} \left(\mathbf{X}_t^+ \right) - \boldsymbol{\mu}_{\mathbf{f}} \right)' \left\| \mathbf{X}_t - \mathbf{X}_t^+ \right\| \right].$$

- Elements in $\mathbf{R}_n$ may diverge at different speeds.
- In practice, no need to estimate or assume the rates at which $\mathbf{R}_n \to \infty$ as $n \to \infty$ for valid asymptotic inference.
- Distance transformation replaces the role of projections in usual IV estimates:

$$\boldsymbol{\Pi} := -\mathbb{E} \left[\left(\mathbf{f} \left(\mathbf{X}_t \right) - \boldsymbol{\mu}_{\mathbf{f}} \right) \mathbf{h} \left(\mathbf{X}_t \right)' \right] = -\mathbb{E} \left[\left(\mathbf{Y}_t - \boldsymbol{\mu}_{\mathbf{Y}} \right) \mathbf{h} \left(\mathbf{X}_t \right)' \right],$$

for

$$\mathbf{h} \left(\mathbf{X}_t \right) := \mathbb{E}^+ \left[\left(\mathbf{f} \left(\mathbf{X}_t^+ \right) - \boldsymbol{\mu}_{\mathbf{f}} \right) \left\| \mathbf{X}_t - \mathbf{X}_t^+ \right\| \right] \not\to \mathbf{f} \left(\mathbf{X}_t \right) - \boldsymbol{\mu}_{\mathbf{f}} := \mathbb{E} \left[\mathbf{Y}_t - \boldsymbol{\mu}_{\mathbf{Y}} | \mathbf{X}_t \right].$$

Wald Statistic

- Nonlinear restrictions
$$H_0 : \mathbf{g}\left(\boldsymbol{\beta}_0\right) = 0, \tag{8}$$
where $\mathbf{g}\left(\cdot\right)$ is a function from $\mathbb{R}^p$ on $\mathbb{R}^m$ with $m \leq p$.
- The Wald test statistic is constructed as
$$W_n\left(\hat{\boldsymbol{\theta}}\right) = n\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right)'\right)^{-1}\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right), \tag{9}$$
where $\mathbf{G}\left(\hat{\boldsymbol{\beta}}\right) = \partial\mathbf{g}\left(\hat{\boldsymbol{\beta}}\right)/\partial\boldsymbol{\beta}'$,
$$\hat{\mathbf{V}}\left(\boldsymbol{\theta}, \lambda\right) = \hat{\boldsymbol{\Upsilon}}\left(\lambda\right)^{-1}\hat{\boldsymbol{\Omega}}\left(\boldsymbol{\theta}, \lambda\right)\hat{\boldsymbol{\Upsilon}}\left(\lambda\right)^{-1}.$$
- For
$$\hat{\boldsymbol{\Upsilon}}\left(\lambda\right) = \frac{1}{n^2}\tilde{\mathbf{Y}}'\tilde{\mathbf{D}}\left(\lambda\right)\tilde{\mathbf{Y}} = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\tilde{D}_{jk}\left(\lambda\right).$$
- $\hat{\boldsymbol{\Omega}}\left(\boldsymbol{\theta}, \lambda\right) = \hat{\mathbf{S}}_1\left(\boldsymbol{\theta}, \lambda\right) + \hat{\mathbf{S}}_2\left(\boldsymbol{\theta}, \lambda\right) - \hat{\mathbf{S}}_3\left(\boldsymbol{\theta}, \lambda\right) - \hat{\mathbf{S}}_3'\left(\boldsymbol{\theta}, \lambda\right)$,
$$\hat{\mathbf{S}}_1\left(\boldsymbol{\theta}, \lambda\right) = \frac{1}{n^3}\sum_{l=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\varepsilon_l\left(\boldsymbol{\theta}\right)^2\tilde{\mathbf{Y}}_j\tilde{\mathbf{Y}}_k'\tilde{D}_{jl}\left(\lambda\right)\tilde{D}_{kl}\left(\lambda\right), \quad \ldots$$

Asymptotic Distribution of the Wald Statistic

Theorem 5

Under Assumptions 1-3, if $\mathbf{g}(\cdot)$ *is continuously differentiable twice and* $\mathbf{G}(\boldsymbol{\beta}_0)$ *is of full rank, for testing the null (8), considering* $\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right) = \left(\hat{\boldsymbol{\theta}}_{WCIV}, \hat{\lambda}_{WCIV}\right)$ *or* $\left(\hat{\boldsymbol{\theta}}_{WCIVF}, \hat{\lambda}_{WCIVF}\right)$,

$$n\mathbf{R}_n^{-1}\hat{\mathbf{\Omega}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{V}\left(\boldsymbol{\theta}_0\right),$$

$$n\mathbf{R}_n^{-1}\hat{\mathbf{\Upsilon}}\left(\hat{\lambda}\right)\mathbf{R}_n^{-1\prime} \xrightarrow{p} \mathbf{\Pi},$$

and

$$W_n\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right) \xrightarrow{d} \chi^2_m.$$

- Under the null, $W_n\left(\hat{\boldsymbol{\theta}}\right)$ has a convenient chi-squared distribution asymptotically, despite the fact that the degrees of identification are unknown.
- In particular, we can obtain t-statistics by treating $\hat{\boldsymbol{\beta}}$ as if it were normally distributed with mean $\boldsymbol{\beta}_0$ and variance $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)/n$. Under the null, the t-statistic $\left(\hat{\beta}_j - \beta_{0j}\right)/\sqrt{\hat{V}_{jj}\left(\hat{\boldsymbol{\theta}}, \hat{\lambda}\right)/n}$ will be asymptotically normal.

WCIV with many instruments: $q \to \infty$

Assumption

$\left\{(\varepsilon_{0t}, \mathbf{X}_t', \boldsymbol{\eta}_t')'\right\}$ *is strictly stationary and ergodic, and absolutely regular with mixing coefficients* $\gamma(n) = O\left(n^{-r}\right)$ *for some* $r > 7\left(1 + 2\upsilon^{-1}\right)$ *with some* $\upsilon > 0$.

$\mathbb{E}\left(\varepsilon_{0t}|\mathbf{X}_t\right) = 0, \mathbb{E}\left(\boldsymbol{\eta}_t|\mathbf{X}_t\right) = 0$, *a.s.*.

$\mathbb{E}\left\|\mathbf{X}_t - \boldsymbol{\mu}_X\right\|^{3(1+\upsilon)} < C$, $\mathbb{E}\left|\varepsilon_{0t}\right|^{3(1+\upsilon)} < C$, $\mathbb{E}\left|\boldsymbol{\eta}_t\right|^{3(1+\upsilon)} < C$,
$\mathbb{E}\left(\left\|\mathbf{X}_t - \boldsymbol{\mu}_X\right\|^{2+\upsilon}\left|\varepsilon_{0t}\right|^{2+\upsilon}\right) < C$, $\mathbb{E}\left(\left\|\mathbf{X}_t - \boldsymbol{\mu}_X\right\|^{2+\upsilon}\left|\boldsymbol{\eta}_t\right|^{2+\upsilon}\right) < C$.
$Var\left((\varepsilon_{0t}, \boldsymbol{\eta}_t')'\,|\mathbf{X}_t\right) = diag\left(\Phi_t^*, 0\right)$, *a.s., and* $\vartheta_{\min}\left(\sum_{t=1}^n \Phi_t^*/n\right) \geq 1/C$.

When $q \to \infty$, we rely on stronger moment conditions and V-statistics theory, instead of empirical process theory, and accommodate non-MDS scenarios.

Theorem 6

For $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{WCIV}$ *or* $\hat{\boldsymbol{\beta}}_{WCIVF}$, $\hat{\alpha} = \hat{\alpha}_{WCIV}$ *or* $\hat{\alpha}_{WCIVF}$, *when* $q \to \infty$ *and* $n \to \infty$,

$$\mathbf{R}_n'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)/r_n \xrightarrow{p} 0,$$

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0, \hat{\alpha} \xrightarrow{p} \alpha_0.$$

Comparison with WMD, Antoine and Lavergne (2014)

- The minimum distance objective function of WMD is based on a population moment with a (Gaussian) *pdf* weighting which can be bounded as

$$\left|\mathbb{E}\left[\left(y_t - \boldsymbol{\theta}'\mathbf{Y}_t^*\right)\left(y_t^+ - \boldsymbol{\theta}'\mathbf{Y}_t^{*+}\right)\exp\left(-\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|^2/2\right)\right]\right|$$
$$\leq C\left(\boldsymbol{\theta}\right)\left\{\mathbb{E}\left[\exp\left(-\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|^2\right)\right]\right\}^{1/2} \underset{q\to\infty}{\longrightarrow} 0.$$

- Assuming the components of $\mathbf{X}_t = (X_{1t}, ..., X_{qt})'$ are i.i.d., then

$$\mathbb{E}\left[\exp\left(-\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|^2\right)\right] = \left\{\underbrace{\mathbb{E}\left[\exp\left(-\left(\mathbf{X}_{1t} - \mathbf{X}_{1t}^+\right)^2\right)\right]}_{<1}\right\}^q \underset{q\to\infty}{\longrightarrow} 0.$$

- θ_0 is not identifiable when q goes to infinity! It is confirmed by the poor estimation accuracy of WMD in simulations and poor finite-sample properties of the ICM test of Bierens (1992) in high-dimensional situations (Tan and Zhu, 2021).

Comparison with Tsyawo (2023)

- Tsyawo (2022) introduces the so-called Generalized Martingale Difference Divergence measure:

$$GMDD = \int_{\mathbb{R}^q} \left|\mathbb{E}\left[\left(y_t - \boldsymbol{\theta}'\mathbf{Y}_t\right) \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t\right\rangle\right)\right]\right|^2 \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right). \tag{10}$$

- However, GMDD is not well-defined because of the non-integrability of $\boldsymbol{\omega}(\boldsymbol{\tau})$. Only when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\mathbb{E}\left(y_t - \boldsymbol{\theta}_0'\mathbf{Y}_t | \mathbf{X}_t\right) = 0$, it coincides with MDD.
- Correspondingly, the population **objective function** in Tsyawo (2022),

$$-\mathbb{E}\left[\left(y_t - \boldsymbol{\theta}'\mathbf{Y}_t\right)\left(y_t^+ - \boldsymbol{\theta}'\mathbf{Y}_t^+\right)\left\|\mathbf{X}_t - \mathbf{X}_t^+\right\|\right],$$

is not the analytic form of (10).
- Instead, this objective function equals

$$-\int_{\mathbb{R}^q} \mathbb{E}\left[\left(y_t - \boldsymbol{\theta}'\mathbf{Y}_t\right)\left(y_t^+ - \boldsymbol{\theta}'\mathbf{Y}_t^+\right)\left(1 - \exp\left(i\left\langle \boldsymbol{\tau}, \mathbf{X}_t - \mathbf{X}_t^+\right\rangle\right)\right)\right] \boldsymbol{\omega}\left(d\boldsymbol{\tau}\right),$$

which is not of the form of a squared L^2 norm.
- This minimization problem is ill-posed and it is challenging to adapt to a LIML-type modification.

Efficiency of estimates

- Not easy to make general statements on efficiency with a continuum of moments estimates, as for WMD.
- Efficiency requires to **weight moments** in terms of their conditional variance (Carrasco and Florens, 2000; Laverge and Patilea, 2013).
- This, in turn, deteriorates finite sample properties of estimates, specially under weak/semi-strong identification (CUE, etc.).

Efficiency of estimates

- Not easy to make general statements on efficiency with a continuum of moments estimates, as for WMD.
- Efficiency requires to **weight moments** in terms of their conditional variance (Carrasco and Florens, 2000; Laverge and Patilea, 2013).
- This, in turn, deteriorates finite sample properties of estimates, specially under weak/semi-strong identification (CUE, etc.).
- We provide a simple comparison in a toy example (strong identification, homoskedasticity, small q).
- We set $p \leq q \leq 2$, $\mathbb{E}\left(\varepsilon_{0t}^2|\mathbf{X}_t\right) = \sigma_\varepsilon^2$ and assume

$$(\mathbf{Y}_t, \mathbf{X}_t) \sim i.i.d.N(\mathbf{0}, \Sigma_{\mathbf{YX}})$$

  and $Var\left(\mathbf{X}_t\right) = \Sigma_{\mathbf{XX}} = \sigma_{\mathbf{X}}^2 \mathbf{I}_q$.
- Same results for Gaussian $\mathbf{X}_t$ and linear $\mathbb{E}[\mathbf{Y}_t|\mathbf{X}_t]$.

Asymptotic Relative Efficiency of estimates

- The asymptotic variance of **WCIV** is

$$\mathbf{\Pi}^{-1}\mathbf{V}\left(\boldsymbol{\theta}_0\right)\mathbf{\Pi}^{-1} = ARE_{WCIV}\left(q\right)\cdot\sigma_{\varepsilon}^2\left(\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\right)^{-1},$$

  where $ARE_{WCIV}\left(q\right)$ is the asymptotic relative efficiency of WCIV wrt 2SLS:

$$\begin{aligned} ARE_{WCIV}(1) &= \frac{\pi}{3} \approx 1.047 \\ ARE_{WCIV}(2) &\approx 1.035. \end{aligned}$$

- This holds independently of any other parameter beyond q, in particular of $\sigma_{\mathbf{X}}^2$: WCIV is only slightly less efficient than 2SLS, which is the optimal estimate in this set up, and efficiency loss shrinks with q.

Asymptotic Relative Efficiency of estimates

- The asymptotic variance of **WCIV** is

$$\boldsymbol{\Pi}^{-1}\mathbf{V}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Pi}^{-1} = ARE_{WCIV}\left(q\right)\cdot\sigma_{\varepsilon}^{2}\left(\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\right)^{-1},$$

  where $ARE_{WCIV}\left(q\right)$ is the asymptotic relative efficiency of WCIV wrt 2SLS:

$$\begin{aligned} ARE_{WCIV}(1) &= \frac{\pi}{3} \approx 1.047 \\ ARE_{WCIV}(2) &\approx 1.035. \end{aligned}$$

- This holds independently of any other parameter beyond q, in particular of $\sigma_{\mathbf{X}}^{2}$: WCIV is only slightly less efficient than 2SLS, which is the optimal estimate in this set up, and efficiency loss shrinks with q.
- The ARE of **WMD** with **normal weighting** wrt 2SLS depends on q and $\sigma_{\mathbf{X}}^{2}$,

$$\begin{aligned} ARE_{WMD}\left(1,\sigma_{\mathbf{X}}^{2}\right) &= \frac{\left(1+2\sigma_{\mathbf{X}}^{2}\right)^{3}}{\left(1+\sigma_{\mathbf{X}}^{2}\right)^{3/2}\left(1+3\sigma_{\mathbf{X}}^{2}\right)^{3/2}} \in (1,\ 1.540) \\ ARE_{WMD}\left(2,\sigma_{\mathbf{X}}^{2}\right) &\in (1,\ 1.778). \end{aligned}$$

- Inefficiency of WMD increases with q for fixed $\sigma_{\mathbf{X}}^{2}$, while the lowest $ARE = 1$ is only approached as $\sigma_{\mathbf{X}}^{2} \to 0$, cf. Lavergne and Patilea (2013).

Monte Carlo Simulation Evidences: Setup 1

$$y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t$$

- $$M_1 : Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t,$$

- $$M_2 : Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \eta_t,$$

- $$M_3 : Y_t = 1\left\{ \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t > 0 \right\},$$

  where $1\{\cdot\}$ denotes the indicator function.

Monte Carlo Simulations: Setup 1 (Hausman et al. 2012)

- To mimic empirical situations, $\mathbf{X}_t = (X_{1,t}, ..., X_{q,t})'$ follows a VAR(1)

$$X_{j,t} = 0.5X_{j,t-1} + \frac{e_{0,t} + e_{j,t}}{\sqrt{2}}, \; j = 1, ..., q,$$

  where $(e_{0,t}, e_{1,t}, ..., e_{q,t}) \sim i.i.d.N\left(0, I_{q+1}\right)$.

- ε_t is allowed to be heteroskedastic [LIML is not consistent],

$$\varepsilon_t = \rho\mu_t + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^4}}\left(\phi\eta_{1,t} + 0.86\eta_{2,t}\right), \eta_{1,t} \sim N\left(0, X_{1,t}^2\right), \eta_{2,t} \sim N\left(0, 0.86^2\right).$$

- We set $\alpha_0 = \beta_0 = 0$ without losing generality, sample size $n = 250$, $c = 10$ and $\rho = 0.6$. We consider $q = 3, 10, 15$ and $\phi = 0$ or 0.5.
- For HFUL estimation, $\left(1, X_t', \left(X_t^2\right)', \left(X_t^3\right)', \left(X_t^4\right)', X_t'd_1, ..., X_t'd_{L-4}\right)'$, where $d_l \in \{0, 1\}$, $\Pr\left(d_l = 1\right) = 1/2$. We consider $L = 1$, 4 or 9: HFUL1, HFUL4, and HFUL9, respectively.
- The comparisons among these estimators are in terms of Median Biases, Ranges between the 0.05 and 0.95 quantiles, and empirical Rejection Frequencies for t-statistics at the 5% nominal level.
- The number of Monte Carlo simulations is $10,000$, $n = 250$.

$M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \mu_t$

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med bias	0.0015	0.0016	0.0017	0.0047	0.0330	0.0339	0.0464
	DecR	0.8347	0.8341	1.0607	1.0452	0.7504	0.8806	1.1013
	Rej	0.0537	0.0537	0.0547	0.0554	0.0336	0.0510	0.0797
					$q = 15$			
	Med bias	0.0000	0.0003	-0.0004	0.0337	0.0084	0.0126	0.0293
	DecR	0.3782	0.3780	0.7192	0.6159	0.3877	0.5617	0.9736
	Rej	0.0472	0.0474	0.0507	0.0664	0.0040	0.0197	0.0741
$\phi = 0.5$					$q = 3$			
	Med bias	-0.0031	-0.0030	-0.0061	-0.0034	0.0361	0.0400	0.0524
	DecR	0.9336	0.9337	1.1291	1.1103	0.8453	0.9939	1.2257
	Rej	0.0503	0.0503	0.0500	0.0509	0.0362	0.0574	0.0882
					$q = 15$			
	Med bias	0.0002	0.0005	-0.0018	0.0317	0.0111	0.0141	0.0228
	DecR	**0.4054**	**0.4048**	0.6831	0.5891	0.4186	0.5765	0.9542
	Rej	0.0500	0.0500	0.0487	0.0632	0.0080	0.0285	0.0768

$$M_2 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\ Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t}^2 + \mu_t$$

		WCIV	**WCIVF**	**WMD**	**WMDF**	**HFUL1**	**HFUL4**	**HFUL9**
$\phi = 0$					$q = 3$			
	Med bias	0.0000	0.0009	-0.0020	0.0028	0.5617	0.0267	0.0272
	DecR	1.0401	1.0326	1.2310	1.1989	1.5104	0.7683	0.9729
	Rej	0.0398	0.0398	0.0512	0.0516	0.5472	0.0436	0.0701
					$q = 15$			
	Med bias	-0.0009	0.0012	-0.0026	0.0576	0.4480	0.0142	0.0470
	DecR	0.4831	0.4772	1.1197	0.7513	2.0637	0.6847	1.2875
	Rej	0.0366	0.0379	0.0574	0.0808	0.5118	0.0260	0.0804
$\phi = 0.5$					$q = 3$			
	Med bias	-0.0189	-0.0165	-0.0118	-0.0069	0.5668	0.0317	0.0379
	DecR	1.1754	1.1458	1.2764	1.2380	1.5758	0.8506	1.0585
	Rej	0.0359	0.0362	0.0474	0.0482	0.5313	0.0555	0.0808
					$q = 15$			
	Med bias	-0.0030	-0.0003	-0.0030	0.0590	0.4479	0.0168	0.0399
	DecR	**0.5380**	**0.5282**	1.0702	0.7366	2.0453	0.7144	1.2552
	Rej	0.0408	0.0415	0.0550	0.0780	0.5172	0.0302	0.0850

$$M_3 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_t,\, Y_t = 1\left\{\sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \mu_t > 0\right\}$$

		WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL4	HFUL9
$\phi = 0$					$q = 3$			
	Med bias	-0.0062	-0.0057	0.0006	0.0091	0.0843	0.1167	0.1554
	DecR	2.3123	2.3006	3.1884	3.0489	1.9550	2.5684	3.5802
	Rej	0.0442	0.0442	0.0457	0.0465	0.0233	0.0478	0.0829
					$q = 15$			
	Med bias	-0.0026	-0.0017	-0.0114	0.0759	0.0248	0.0412	0.1182
	DecR	1.1017	1.1006	2.3307	1.8374	1.2120	1.9415	3.7257
	Rej	0.0483	0.0487	0.0478	0.0606	0.0046	0.0309	0.0800
$\phi = 0.5$					$q = 3$			
	Med bias	-0.0028	-0.0023	-0.0121	-0.0043	0.1153	0.1288	0.1585
	DecR	2.5261	2.5255	3.3285	3.2036	2.1789	2.8813	3.8007
	Rej	0.0417	0.0418	0.0441	0.0445	0.0302	0.0562	0.0855
					$q = 15$			
	Med bias	-0.0055	-0.0049	-0.0116	0.0761	0.0298	0.0315	0.1004
s	**DecR**	**1.1689**	**1.1680**	2.3200	1.7693	1.3057	2.0461	3.8185
optma	**Rej**	0.0459	0.0460	0.0471	0.0585	0.0075	0.0328	0.0806

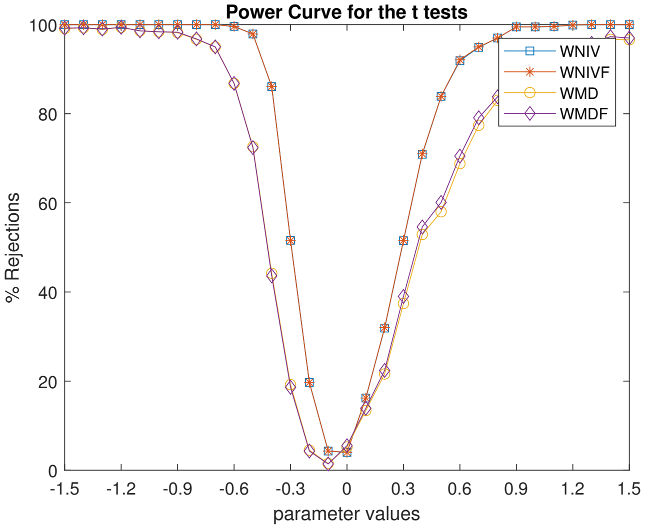

Figure: Power Curve of the t tests of WCIV, WCIVF, WMD and WMDF. The linear IV model $M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t$ with $\phi = 0.5$, $c = 10$ and $q = 10$. The sample size is 250. The null hypothesis is $H_0 : \beta_0 = 0$.

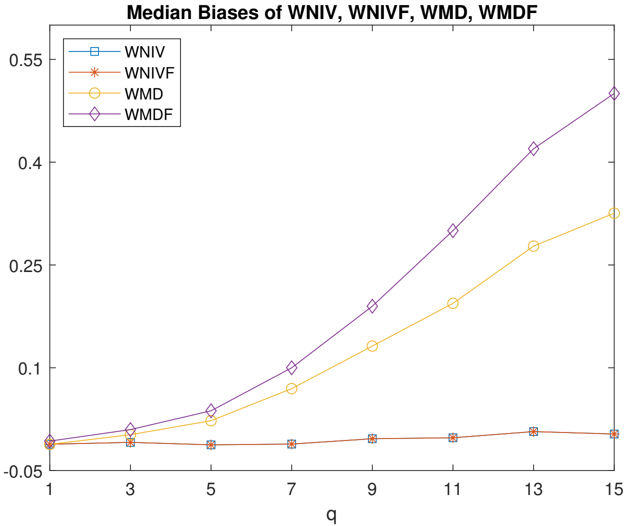

Figure: Median biases of WCIV, WCIVF, WMD and WMDF. The linear IV model $M_1 : y_t = \alpha_0 + \beta_0 Y_t + \varepsilon_{0t}$, $Y_t = \sqrt{\frac{c/q}{n}} \sum_{j=1}^{q} X_{j,t} + \eta_t$ with $\phi = 0.5$, $c = 10$. $(X_{1t}, ..., X_{qt})$ are i.i.d.. The sample size is 250.

EIS Estimation

- The EIS in consumption is an important parameter in macroeconomics and finance. It measures how much consumers change their expected consumption growth rate in response to changes in the expected return on any asset.
- It is required to be large in the calibrations of macro and finance models in recent literature.
  - Kydland and Prescott (1982): $EIS = 0.66$;
  - King and Rebelo (1990), $EIS = 1$;
  - Bansal and Yaron (2004) $EIS = 1.5$;
  - Barro (2009), Ai (2010), and Colacito and Croce (2011), $EIS = 2$.
- The EIS value has been estimated using different macro data sets.
  - Hall (1988): for the US, the EIS is is unlikely to be much above 0.1 and may well be zero.
  - Campbell (2003) and Yogo (2004): similar conclusions for 11 developed countries.
  - Ascari et al. (2021): similar conclusions for the US, using inference procedures robust to weak instruments and an updated data set.
  - Chen et al. (2025): generalized Bierens maximum statistic, $[-0.30, 0.15]$ for US.

The US Quarterly Data from Ascari et al. (2021)

$$\mathbb{E}\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathcal{F}_t\right] = 0, \tag{1}$$

$$\mathbb{E}\left[r_{t+1} - \varpi - 1/\psi \triangle c_{t+1} | \mathcal{F}_t\right] = 0, \tag{2}$$

- Macro data from Q4 1955 to Q1 2018.
- Nominal interest rate i_t: three-month treasury bill rate;
  Nominal stock market return s_t: the S&P 500 return;
  c_t: log of consumption of nondurables.
- We consider $i_t - \pi_{t+1}$, which is the ex-post real interest rate, and $s_t - \pi_{t+1}$, which is the ex-post real stock return.
- The EIS is estimated using the real interest rate as the asset return.
- $\mathbf{X}_t$ comprises lag terms of the real interest rate, real stock return, consumption growth and the first-difference of the log dividend-price ratio:

$$\begin{aligned} \mathbf{Z}_t &= (i_t - \pi_{t+1}, s_t - \pi_{t+1}, \triangle c_t, \triangle dp_t) \\ \mathbf{X}_t &= (\mathbf{Z}_t, \ldots, \mathbf{Z}_{t-L}), \quad L = 0, 1, 2. \end{aligned}$$

- For HFUL, we consider HFUL1, HFUL2, and HFUL6.

The US Quarterly Data from Ascari et al. (2021)

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	2SLS
				lags 1				
ψ	1.03*	1.01*	1.19	0.77*	0.15	0.22**	0.19	0.14
	(0.62)	(0.60)	(0.97)	(0.45)	(0.14)	(0.11)	(0.17)	(0.10)
$1/\psi$	0.98	0.97	0.84	0.67	4.77	3.95	4.03	0.62**
	(0.61)	(0.61)	(0.71)	(0.51)	(4.32)	(2.05)	(2.41)	(0.31)
				lags 1 to 2				
ψ	1.15*	1.14*	1.27	0.90*	0.21	0.23	0.19	0.17*
	(0.64)	(0.63)	(0.90)	(0.49)	(0.16)	(0.16)	(0.15)	(0.09)
$1/\psi$	0.87*	0.86*	0.78	0.68	3.94	3.65	3.28	0.50**
	(0.50)	(0.50)	(0.58)	(0.47)	(3.00)	(2.45)	(1.67)	(0.23)
				lags 1 to 3				
ψ	1.62*	1.60*	1.82	1.36**	0.23	0.24	0.26	0.18**
	(0.94)	(0.92)	(1.16)	(0.67)	(0.18)	(0.21)	(0.27)	(0.09)
$1/\psi$	0.62*	0.61*	0.55*	0.52*	3.67	3.45	2.98	0.46**
	(0.34)	(0.34)	(0.33)	(0.31)	(2.81)	(2.95)	(2.55)	(0.22)

Table: Data: 1955Q4-2018Q1. EIS is estimated from $\mathbb{E}\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathbf{X}_t\right] = 0$. The reciprocal of the EIS is estimated from $\mathbb{E}\left[r_{t+1} - \mu - 1/\psi \triangle c_{t+1} | \mathbf{X}_t\right] = 0$.

- $\mathbf{X}_t$ comprises lags of the real interest rate, real stock return, consumption growth, and the first difference of the log D-P ratio up to 3 lags.

The US Quarterly Data from Beeler and Campbell (2012)

	WCIV	WCIVF	WMD	WMDF	HFUL1	HFUL2	HFUL6	2SLS
				Set 1				
ψ	1.21**	1.19**	1.10*	0.83**	0.30**	0.29**	0.27**	0.29**
	(0.61)	(0.59)	(0.59)	(0.37)	(0.13)	(0.13)	(0.14)	(0.13)
$1/\psi$	0.83*	0.81*	0.91*	0.57**	2.90**	2.91**	2.97**	1.13***
	(0.46)	(0.45)	(0.50)	(0.26)	(1.32)	(1.32)	(1.46)	(0.31)
				Set 2				
ψ	0.90**	0.87**	0.95**	0.34***	0.30**	0.29**	0.27*	0.29***
	(0.36)	(0.35)	(0.48)	(0.12)	(0.13)	(0.13)	(0.15)	(0.12)
$1/\psi$	1.11*	1.02*	1.06*	0.19**	2.89**	2.90**	2.82**	1.12***
	(0.65)	(0.56)	(0.63)	(0.07)	(1.31)	(1.32)	(1.32)	(0.31)
				Set 3				
ψ	1.21**	1.19**	1.10*	0.84**	0.30**	0.30**	0.32**	0.29**
	(0.61)	(0.60)	(0.60)	(0.38)	(0.13)	(0.14)	(0.14)	(0.13)
$1/\psi$	0.83*	0.81*	0.91*	0.59**	2.89**	2.86**	2.71**	1.13***
	(0.46)	(0.45)	(0.50)	(0.27)	(1.31)	(1.30)	(1.38)	(0.31)

Table: Data: 1947Q2-2008Q4. EIS is estimated from $\mathbb{E}\left[\triangle c_{t+1} - \alpha - \psi r_{t+1} | \mathbf{X}_t\right] = 0$. The reciprocal of the EIS is estimated from $\mathbb{E}\left[r_{t+1} - \mu - 1/\psi \triangle c_{t+1} | \mathbf{X}_t\right] = 0$.

- **Set 1**: first lag terms of real interest rate, real stock return, consumption growth.
- **Set 2**: first lag terms of real interest rate, consumption growth, and the first difference of the log D-P ratio.
- **Set 3**: first lag terms of real interest rate, real stock return, consumption growth, and the first difference of the log D-P ratio.

Discussion

- WCIV and WCIVF($C = 1$) estimates of the EIS (ψ) appear to be large, and statistically significant at the 10% significance level.
- These findings hold true over model transformation, with the WCIV and WCIVF($C = 1$) estimates of $1/\psi$ aligning with those of ψ.
- HFUL estimates of the EIS are generally greater than the 2SLS estimates, but much less than the WCIV and WCIVF($C = 1$) estimates.
- These findings are robust to distinct sets of $\mathbf{X}_t$, model transformations, various data structures and different data ranges.

Conclusions

- New IV estimates valid under many weak instruments in a dynamic context.
- No need of specifying a set of IVs to approximate a complex RF.
- Good bias and efficiency properties compared to other proposals.
- Robust asymptotic inference to heteroskedasticity and many weak instruments.

Thank you!