Linear Regression with Weak Exogeneity

Anna Mikusheva (MIT)
Mikkel Sølvsten (Aarhus University)

Intro: OLS with weak exogeneity

- Linear regression in time series

$$y_t = x_t'\beta + \varepsilon_t, \qquad t \in \{1, \ldots, T\}$$

- Object of interest is the linear contrast $\theta = r'\beta$
- The most commonly used assumption of weak exogeneity:

$$\mathbb{E}[\varepsilon_t \,|\, x_t, x_{t-1}, \ldots] = 0$$

- It is common to have feedback from y_t to x_{t+1}
- All good properties of OLS (no bias) derived with strict exogeneity:

$$\mathbb{E}[\varepsilon_t \,|\, \ldots, x_{t+1}, x_t, x_{t-1}, \ldots] = 0$$

Intro: OLS with weak exogeneity

- It is known that OLS is biased in time series
- Common belief: OLS is consistent, asymptotically Gaussian, bias is small
- Our claim: OLS may have large biases and be even inconsistent
- Factors leading to large bias of OLS:
  - violations of strict exogeneity (even mild, one-period)
  - regressors are auto-correlated (no strong persistence needed)
  - many regressors

Intro: OLS with weak exogeneity

Simulation setup:

- $\tilde{x}_t$ is K-dimensional AR (1) process with parameter ρ
- $\varepsilon_t \sim i.i.d. N(0,1)$ independent from $\tilde{X}$
- violation of strict exogeneity for one period: $x_{t+1} = \tilde{x}_{t+1} + \alpha\varepsilon_t$
- $y_t = x_t'\beta + \varepsilon_t$
- $\hat{\beta}^{\text{OLS}} = (X'X)^{-1}X'Y$
- $\hat{\theta}^{\text{OLS}} = r'\hat{\beta}^{\text{OLS}}$

Intro: OLS with weak exogeneity

Simulation results ($T = 200$):

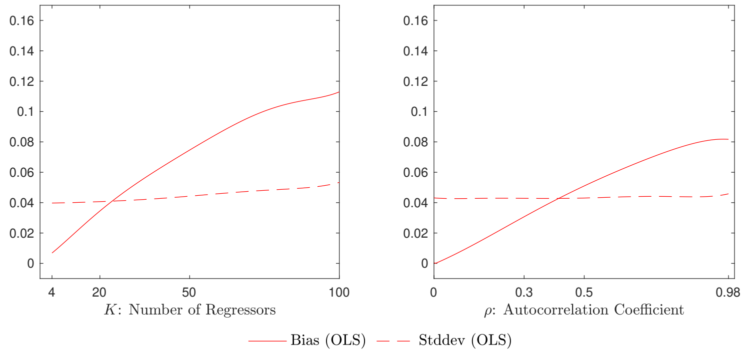

Figure 1: Absolute Bias and Standard Deviation of OLS

Same results with MA(1) process | Results with T=800

Intro: OLS with weak exogeneity

- We derive formula for OLS (feedback) bias
- We propose a new estimator that is nearly unbiased
  - Our estimator uses an oblique projection (IV motivated)
  - Uses an ‘invalid’ IV; an endogenous instrument
  - ‘Instrument’ is constructed (from regressors) so that

    endogeneity bias cancels with the feedback bias.
  - Our estimator is consistent and asymptotically Gaussian
  - In most settings, simulated changes to standard deviations are minimal in comparison to OLS

Intro: OLS with weak exogeneity

Comparison with new estimator ($T = 200$):

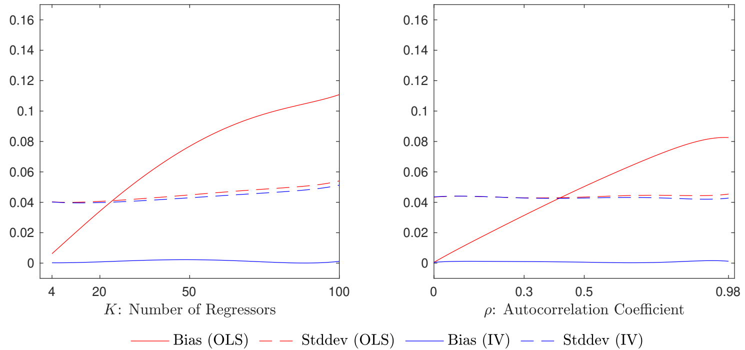

Figure 2: Absolute Bias and Standard Deviation of OLS and IV

Plan for This Talk

- Why is OLS biased?
- New estimator
- Consistency and asymptotic Gaussianity
- ~~Multi-period feedback~~

Why OLS is biased?

Why OLS is biased?

Special case (only first regressor is weakly exogenous)

- $Y = X\beta + \varepsilon$
- X are $T \times K$ observed regressors
- $\tilde{X}$ are strictly exogenous variables:

$$\mathbb{E}[\varepsilon_t \,|\, \tilde{X}] = 0, \qquad \mathbb{E}[\varepsilon_t \varepsilon_s \,|\, \tilde{X}] = \sigma^2 1\{s = t\}$$

- Feedback: $x_{1,t} = \tilde{x}_{1,t} + a\varepsilon_{t-1}$,
- All other regressors are strictly exogenous $X_{-1} = \tilde{X}_{-1}$
- Normalization: $\tilde{X}'\tilde{X}/T = I_K$

Why OLS is biased?

- Frisch-Waugh theorem:

$$\hat{\beta}_1^{\mathsf{OLS}} = \frac{X_1' M_{-1} Y}{X_1' M_{-1} X_1}; \qquad \hat{\beta}_1^{\mathsf{OLS}} - \beta_1 = \frac{X_1' M_{-1} \varepsilon}{X_1' M_{-1} X_1}$$

$M_{-1} = I - \tilde{X}_{-1}(\tilde{X}'_{-1}\tilde{X}_{-1})^{-1}\tilde{X}'_{-1}$ is projection orthogonal to $\tilde{X}_{-1}$

- $x_{1,t} = \tilde{x}_{1,t} + a\varepsilon_{t-1}$

$$X_1' M_{-1} \varepsilon = \sum_{s,t} M_{st} X_s \varepsilon_t = \tilde{X}_1' M_{-1} \varepsilon + a \sum_{s,t} M_{st} \varepsilon_{s-1} \varepsilon_t$$

- Partialling out mixes up timing!!!
- $\mathbb{E}\left[\sum_{s,t} M_{s,t}\varepsilon_{s-1}\varepsilon_t \,|\, \tilde{X}\right] = \sigma^2 \sum_t M_{t+1,t}$

Why OLS is biased?

$$\hat{\beta}_1^{\mathsf{OLS}} - \beta_1 = \frac{X_1' M_{-1} \varepsilon}{X_1' M_{-1} X_1}$$

- The denominator is $T + a^2\sigma^2(T-K) + o_p(T)$
- The order of the bias in $\hat{\beta}_1^{\mathsf{OLS}}$ is

$$\frac{a\sigma^2 \sum_t M_{t+1,t}}{T + a^2\sigma^2(T-K)}$$

Why OLS is biased?

- $\sum_t M_{t+1,t} = -\sum_t P_{t+1,t}$, where $P = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$
- Normalization: $\tilde{X}'\tilde{X}/T = I_K$, then

$$\sum_t M_{t+1,t} = -\frac{1}{T}\sum_t \tilde{X}'_{t+1}\tilde{X}_t \approx -\mathbb{E}\tilde{X}'_{t+1}\tilde{X}_t$$

- $\sum_t M_{t+1,t}$ measures a linear connection between $\tilde{X}_t$ and $\tilde{X}_{t+1}$
- In stationary time series we should expect

$$\sum_t M_{t+1,t} \approx -\rho K,$$

where ρ is average of the first order auto-correlation coefficients

Why OLS is biased?

Summary

- If only first regressor is weakly exogenous: $x_{1,t} = \tilde{x}_{1,t} + a\varepsilon_{t-1}$
- Then
  - the coefficient $\hat{\beta}_1^{\mathrm{OLS}}$ is biased by

$$\frac{a\sigma^2 \sum_t M_{t+1,t}}{T + a^2\sigma^2(T-K)} \approx -\frac{a\sigma^2\rho K}{T + a^2\sigma^2(T-K)}$$

  - Inconsistency when: $K/T \to const$; a and ρ separated from zero
  - Bias comparable to standard error when: $K^2/T \to const$; a and ρ separated from zero
  - All other coefficients are nearly unbiased
- Special case is more general than you may think – rotations!

Why OLS is biased; one period feedback

Assumption 1

(i) The regressors x_t satisfy $x_t = \tilde{x}_t + \alpha\varepsilon_{t-1}$ where $\tilde{X}$ has full rank

(ii) The errors $\{\varepsilon_t\}_{t=0}^{T}$ are *i.i.d.* conditionally on $\tilde{X}$ with $\mathbb{E}[\varepsilon_t \,|\, \tilde{X}] = 0$, $\sigma^2 = \mathbb{E}[\varepsilon_t^2 \,|\, \tilde{X}]$ and $\mathbb{E}[\varepsilon_t^4] < \infty$

(iii) The non-random vectors $r,\ \alpha \in \mathbb{R}^K$ satisfy $r'(\tilde{X}'\tilde{X}/T)^{-1}r = O_p(1)$ and $\alpha'(\tilde{X}'\tilde{X}/T)^{-1}\alpha = O_p(1)$.

(iv) The number of regressors K may diverge with the sample size T; $T - K$ diverges to infinity

Why OLS is biased; one period feedback

Theorem: Asymptotic bias of OLS

Suppose Assumption 1 holds. Then,

$$r'\hat{\beta}^{\mathsf{OLS}} - r'\beta = \sigma^2 r'\bar{S}^{-1}\alpha \sum_t \tilde{M}_{t+1,t} + o_p(1),$$

where

- $\bar{S} = \tilde{X}'\tilde{X} + \alpha\alpha'\sigma^2(T-K)$
- $\tilde{M} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$.

- Value of $r'\bar{S}^{-1}\alpha = O(1/T)$ is not known and hard to assess,
- Value of $\sum_t \tilde{M}_{t+1,t} \approx \sum_t M_{t+1,t}$ is observed

Is it empirically important?

- Stock and Watson (2016) data set: quarterly observations from 1964 to 2013 (T = 200) on 108 US macro indicators
- Extracted cyclical component = a two-year-ahead forecast error based on a AR(4) forecast (as in Hamilton (2018))
- 100 experiments:
  - randomly draw a regression with K regressors
  - estimate feedback
  - keep exogenous part of regressors, simulate outcome with feedback
  - calculate bias based on 1000 draws

Is it empirically important?

Simulation results with US macro data for $\tilde{x}_t$ $(T = 200)$:

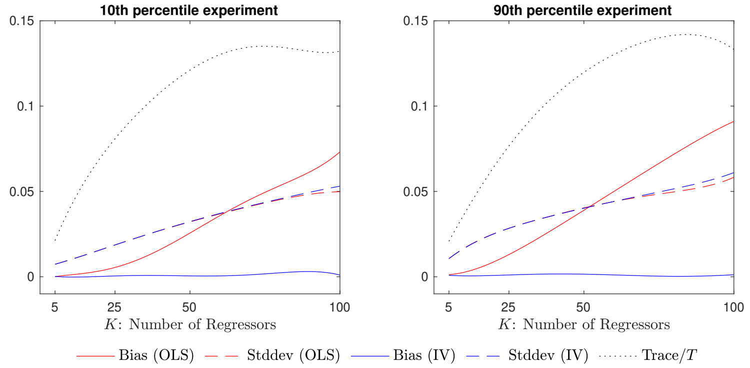

Figure 3: Absolute Bias and Standard Deviation of OLS

New Estimator

New Estimator: one period case

- Assume 1-period violation of strict exogeneity
- Oblique projection:

$$P_{Z,X} = X(Z'X)^{-1}Z', \quad M_{Z,X} = I - P_{Z,X}$$

- $M_{Z,X}^2 = M_{Z,X}$ but not symmetric
- Direction of projection: $z_{t,\gamma} = z_t = x_t - \gamma x_{t+1}$

$$\hat{\beta}^{\mathsf{IV}}(\gamma) = (Z'X)^{-1}Z'Y$$

New Estimator: one period case

- Generalization of Frisch-Waugh:

$$\hat{\beta}_1^{\mathsf{IV}}(\gamma) - \beta_1 = \frac{Z_1' M_{Z_{-1},X_{-1}} \varepsilon}{Z_1' M_{Z_{-1},X_{-1}} X_1}$$

- Special case with 1 weakly exogenous regressor: the same logic
- $M_{Z_{-1},X_{-1}}$ is exogenous
- $x_{1,t} = \tilde{x}_{1,t} + a\varepsilon_{t-1}$
- $z_{1,t} = \tilde{z}_{1,t} + a(\varepsilon_{t-1} - \gamma\varepsilon_t)$

$$Z_1' M_{Z_{-1},X_{-1}} \varepsilon = \tilde{Z}_1' M_{Z_{-1},X_{-1}} \varepsilon + a \sum_{s,t} (\varepsilon_{s-1} - \gamma \varepsilon_s) M_{\gamma,st} \varepsilon_t$$

$$\mathbb{E}\left[Z_1' M_{Z_{-1},X_{-1}} \varepsilon \middle| \tilde{X} \right] = a\sigma^2 \sum_t (M_{\gamma,t+1,t} - \gamma M_{\gamma,tt})$$

New Estimator: one period case

- When the direction of projection: $z_t = x_t - \gamma x_{t+1}$,

$$\hat{\beta}^{\mathsf{IV}}(\gamma) = (Z'X)^{-1} Z'Y$$

- Main bias term:

$$\mathbb{E}\left[Z_1' M_{Z_{-1},X_{-1}} \varepsilon\right] = a\sigma^2 \sum_t (M_{\gamma,t+1,t} - \gamma M_{\gamma,tt})$$

- Idea: choose γ in such a way to make this zero
- Solution is possible since diagonal elements are larger than lower diagonal
- Estimator has IV interpretation: technical 'instrument' $z_t = x_t - \gamma x_{t+1}$ is invalid if/when x_t is not strictly exogenous
- γ is selected s.t. endogeneity bias of IV cancels OLS bias of weak exogeneity

New Estimator: one period case

$$\sum_t (M_{\gamma,t+1,t} - \gamma M_{\gamma,tt}) = 0$$

- Non-linear equation

  Lemma: Existence and uniqueness

- If $X'X$ has rank K and $T \geq 5K$, then equation always has a solution
- If $|\sum_t M_{t+1,t}| \leq \mu K$ and $T > K(1 + (\sqrt{\mu} + 1)^2)$, then equation has a solution $|\gamma| < \frac{\sqrt{\mu}}{\sqrt{\mu}+1}$
- Solution is a fixed point of a contraction

Asymptotics

New estimator: consistency

Theorem: Consistency of IV

Suppose Assumption 1 holds.

(i) If $|\gamma| < 1$ is fixed, then

$$r'\hat{\beta}^{\mathsf{IV}}(\gamma) - r'\beta = \sigma^2 r'\bar{S}_\gamma^{-1}\alpha \sum_t (\tilde{M}_{\gamma,t+1,t} - \gamma\tilde{M}_{\gamma,tt}) + o_p(1),$$

- $\bar{S}_\gamma = \tilde{Z}'\tilde{X} + \sigma^2\alpha\alpha' \sum_t (\tilde{M}_{\gamma,tt} - \gamma\tilde{M}_{\gamma,t,t+1})$
- $\tilde{M}_\gamma = I - \tilde{X}(\tilde{Z}'\tilde{X})^{-1}\tilde{Z}'$

(ii) If $\hat{\gamma}$ solves equation $\sum_t (M_{\gamma,t+1,t} - \gamma M_{\gamma,tt}) = 0$, then

$$r'\hat{\beta}^{\mathsf{IV}}(\hat{\gamma}) - r'\beta = o_p(1).$$

Asymptotics: gaussianity

Assumption 2 $\max_t \|(\tilde{X}'\tilde{X})^{-1/2}\tilde{x}_t\| = o_p(1)$

Theorem: Gaussianity

Suppose Assumptions 1 and 2 hold. If $\hat{\gamma}$ solve trace equation, then

$$\frac{r'\hat{\beta}^{\mathsf{IV}}(\hat{\gamma}) - r'\beta}{\hat{\sigma}_T} \xrightarrow{d} N(0,1) \text{ as } T \to \infty$$

where

- $\hat{\sigma}_T^2 = \hat{\sigma}^2(\hat{\gamma})\|r'(Z_{\hat{\gamma}}'X)^{-1}Z_{\hat{\gamma}}'\|^2$
- $\hat{\sigma}^2(\hat{\gamma}) = \frac{y'(I-\hat{\gamma}D)M_{\hat{\gamma}}y}{\mathrm{tr}\left[(I-\hat{\gamma}D)\tilde{M}_{\hat{\gamma}}\right]}$ where D is 'lag operator matrix'

Size in simulation

Simulation results with the US macro data for $\tilde{x}_t$ ($T = 200$):

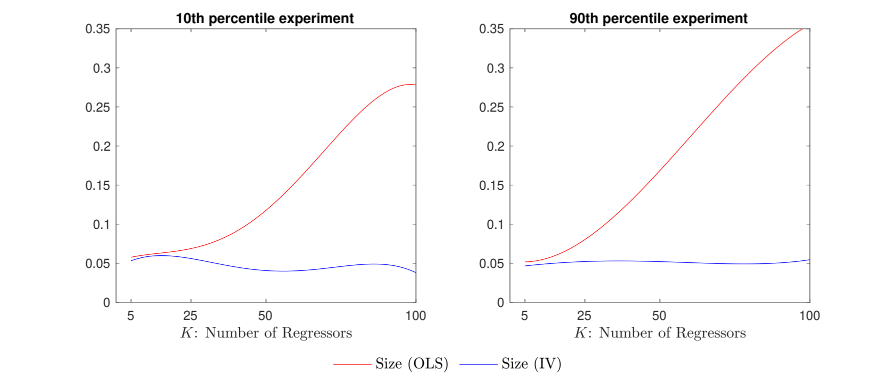

Figure 4: Size of Nominal 5% two-sided tests using OLS and IV with $T = 200$

Summary

- We showed that the typical time series OLS estimator with large number of regressors is prone to large biases
- Factors leading to bias:
  - Weak exogeneity (feedback)
  - Autocorrelation of regressors
  - Number of regressors
- Potential for bias can be assessed by the size of lower traces of M

Summary

- We proposed a new estimator
  - Relies on oblique projection
  - Correction is made on regressors without looking on the outcome or assessing the direction of feedback
  - Consistent under mild assumptions, asymptotically gaussian
  - Standard deviation is comparable to that of OLS

Intro: OLS with weak exogeneity

Simulation results with MA(1)-process for $\tilde{x}_t$ $(T = 200)$:

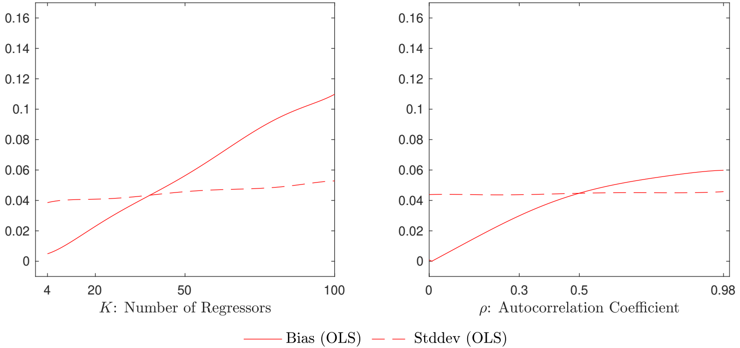

Figure 5: Absolute Bias and Standard Deviation of OLS

Back to presentation

Intro: OLS with weak exogeneity

Simulation results ($T = 800$):

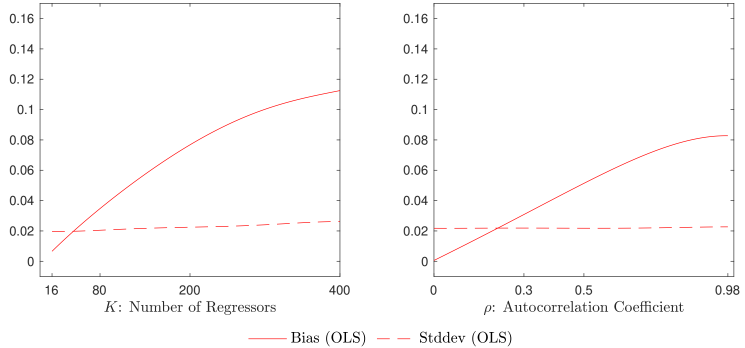

Figure 6: Absolute Bias and Standard Deviation of OLS

Back to presentation