Many Weak Instruments

Anna Mikusheva

MIT

Stylized Setting

- Linear IV model with one endogenous variable

$$Y_t = \beta X_t + e_t, \quad \mathbb{E}[e_t Z_t] = 0$$

  - Many instruments: K is large when compared to T
  - Weak identification
  - Time series: instruments are weakly-exogenous, errors may be autocorrelated
- Lessons can be learned from cross-sectional study of many weak IV
- What challenges are specific to time series?

Rational Expectation Models

- **Example 1**: New Keynesian Phillips curve

$$\pi_t = \lambda x_t + \gamma_f \mathbb{E}_t[\pi_{t+1}] + \gamma_b \pi_{t-1} + e_t$$

- Common estimation method is via IV:

$$\mathbb{E}\left[(\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) Z_t\right] = 0,$$

  where Z_t is any variable in the information set at time $t-1$
- Gali and Gertler (1999) used 4 lags of 6 variables (24 instruments)
- Kleibergen and Mavroeidis (2009): NKPC is weakly identified
- Mavroeidis et al (survey, 2014): uncertainty is too high to provide informative estimates for all practical purposes

Rational Expectation Models

- **Example 2**: Euler equation (linearized version with external habits)

$$\mathbb{E}_t \Delta c_{t+1} = \gamma \Delta c_t + \sigma(1-\gamma) r_t$$

- Many available instruments
- Stock and Wright (2000), Yogo (2004): identification is very weak
- Ascari et al (survey, 2020): uncertainty does not allow us to distinguish specifications

Rational Expectation Models

- **Example 3**: Taylor rule

$$r_t = \beta \mathbb{E}_t \pi_{t+1} + \gamma \mathbb{E}_t x_{t+1} + \rho r_{t-1} + e_t$$

- Clarida et al (1998): IV estimation on monthly data, 37 instruments
- Mavroeidis (2004): identification tends to be very weak

Example Outside Rational Expectations

- **Example 4**: Factor pricing

$$\mathbb{E}r_{it} = \lambda\beta_i; \quad \beta_i = \Sigma_F^{-1} cov(r_{it}, F_t)$$

- Fama-MacBeth procedure = TSLS
  - Estimate β_i via OLS of r_{it} on F_t
  - Estimate λ via OLS of $\bar{r}_i$ on $\hat{\beta}_i$
- Settings with many instruments (proportional to the number of assets)
- Macro factors produce weak identification

Stylized Setting

- Linear IV model with one endogenous variable

$$Y_t = \beta X_t + e_t, \quad \mathbb{E}[e_t Z_t] = 0$$

- Simplifications made here (problems not discussed in this talk), that will bring additional challenges in applications:
    - We consider linear IV, not non-linear GMM
    - No controls (partialled out)
    - Single endogenous regressor
    - Strong persistence (unit root)

Overview

1. Cross-sectional: many weak IV
    - Review of many instruments
    - Is estimated optimal instrument exogenous?
    - Is consistent estimation possible?
    - Are inferences standard?
    - Pretest for Weak IV?
    - Robust testing
2. Time Series

Cross-sectional: Many Instruments

- Moment restriction:

$$Y_i = \beta X_i + e_i, \quad \mathbb{E}[z_i e_i] = 0$$

  for all $z_i \in \mathcal{Z}_i$ (set of allowable instruments: groups, sieve)

- Optimal instrument (Chamberlain, 1987) achieves minimal variance

$$f_i = \frac{\mathbb{E}[X_i|\mathcal{Z}_i]}{\mathbb{E}[e_i^2|\mathcal{Z}_i]}$$

- How to reach a semi-parametric efficient estimator (under homoscedasticity)?

Cross-sectional: Many Instruments

- Non-parametric estimator of optimal instrument on the first stage: sieve or k-nearest neighbors (Newey, 1990)
- Bias of the IV estimator increases with the number of moment conditions/instruments (Bekker, 1994, Newey and Smith, 2004)
- Solution: regularized procedure on the first stage that does selection and estimation of $f_i = \mathbb{E}[X_i|\mathcal{Z}_i]$ (prediction task)
- Second step: do IV with estimated optimal instrument $\hat{f}_i$:

$$\widehat{\beta} = \frac{\sum_i \hat{f}_i Y_i}{\sum_i \hat{f}_i X_i}$$

Cross-sectional: Many Instruments

Some suggestions that lead to semi-parametric efficiency under some assumptions:

- Instrument selection procedure (Donald and Newey, 2001)
- LASSO on the first stage (Belloni et al., 2012)
- Ridge (Okui, 2011)
- Tikhonov's regularization, spectral cut-off (Carrasco, 2012)

There are usually two results:

- Consistency and semi-parametric efficiency.
- Finite-sample bias-variance trade-off to choose regularization parameter

Cross-sectional: Many Instruments

- Angrist and Frandsen (2020) simulation designs mimic two applications:
  - Return to education, instruments -quarter of birth (Angrist and Krueger, 1991)
  - Effect of opening weekend on subsequent movie-going, instruments - weather (Gilchrist and Sands, 2016)
- Comparisons between ML first stage (LASSO, random forest) and econometric estimators (LIML, JIVE, sample-split)
- ML selection on the first stage delivers large bias to the IV estimator

Weak Instruments

- Consider (infeasible) IV with optimal instrument:

$$\begin{cases} Y_i = \beta X_i + e_i, \\ X_i = \mathbb{E}[X_i|\mathcal{Z}_i] + v_i = f_i + v_i. \end{cases}$$

- Weak identification = uncertainty from v_i is empirically important as measured by the signal-to-noise ratio (homoscedastic formula)

$$\mu^2 = \frac{n\mathbb{E}[f_i^2]}{\sigma_v^2}$$

- (Infeasible) optimal IV estimator:

$$\widehat{\beta}_o - \beta = \frac{\sum_{i=1}^n f_i Y_i}{\sum_{i=1}^n f_i X_i} - \beta = \frac{\sum_{i=1}^n f_i e_i}{\sum_{i=1}^n f_i^2 + \sum_{i=1}^n f_i v_i}$$

- Red term is important and endogenous

Weak Instruments

- If the concentration parameter is low, then
  - TSLS is very biased
  - confidence sets and tests are unreliable
- If the model is just identified (f_i is known) then:
  - First stage F-test can be used in two-step procedure as a pre-test
  - Robust tests (AR) are asymptotically efficient

Many Weak IV: Is Estimated Instrument Exogenous?

- If many regressors in the first stage, they might ‘overfit’ the noise
- Estimated optimal instrument is endogenous $E[\hat{f}_i e_i] \neq 0$
- For homoscedastic TSLS: $\widehat{f_i} = X'Z(Z'Z)^{-1}Z_i = f_i + V'Z(Z'Z)^{-1}Z_i$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\widehat{f_i} - f_i)e_i\right] = \mathbb{E}[v_i e_i] trace(Z(Z'Z)^{-1}Z') = K\sigma_{ev}$$

- Endogeneity is growing in K!
    - Leads to bias
    - May destroy consistency

Many Weak IV: Is Estimated Instrument Exogenous?

Suggestions on how to remove endogeneity:

- Sample splitting (Angrist and Krueger, 1995):
  - split sample to halves
  - select/estimate optimal instrument on one half
  - estimate β on the other half
- Jackknife (Angrist et al., 1999)
  - estimate optimal instrument for observation i on sample excluding i
  - use estimated optimal instrument
  - can be done to many estimators: JIVE-LIML and JIVE-Fuller (Hausman et al., 2012), JIVE-ridge (Hansen and Kozbur, 2014)
- In simulations (Angrist and Frandsen, 2020): split-sample and JIVE have superior performance to LASSO and random forest

Many Weak IV: Is Consistent Estimation Possible?

- How large signal-to-noise $\mu^2 = \frac{n\mathbb{E}[f_i^2]}{\sigma_v^2}$ is needed for consistency?
- Depends on what we know about 'optimal' instrument
- The **best** possible scenario: f is known
  - Condition for consistency $\mu^2 \to \infty$
- The **most agnostic** scenario: any linear combination of $(K < n)$ instruments may be optimal
  - Condition for consistency $\frac{\mu^2}{\sqrt{K}} \to \infty$
  - Necessary condition (Mikusheva and Sun, 2020): if $\frac{\mu^2}{\sqrt{K}}$ is bounded then no consistent discrimination between $\beta_0 \neq \beta$ is possible uniformly over all directions of the optimal instrument
  - Sufficient condition: JIVE, JIVE-LIML, JIVE-Fuller are consistent under mild assumptions when $\frac{\mu^2}{\sqrt{K}} \to \infty$

Many Weak IV: Is Consistent Estimation Possible?

There is a trade-off between the quality of first stage estimation and the strength of identification needed for consistency.

Lemma 1.

Under mild regularity conditions if $\mathbb{E}\left[(\widehat{f}_i - f_i)^2\right] = O_p(\frac{r_n}{n})$ *and* $\frac{\mu^2}{\sqrt{r_n}} \to \infty$, *then* $\widehat{\beta}_{SS}$ *and* $\widehat{\beta}_{CFSS}$ *are consistent for* β.

Many Weak IV: Is Consistent Estimation Possible?

- If we assume that the first stage is approximately sparse

$$f_i = Z_i' \pi_0 + r_i, \quad \|\pi_0\|_0 \leq s,$$

  where $s = o(n/\log(K))$ and $\sqrt{\frac{1}{n}\sum_i r_i^2} \leq C\sqrt{\frac{s}{n}}$
- Estimate first stage via LASSO (Belloni et al, 2010)

$$\|\widehat{\pi} - \pi_0\|^2 = O_p\left(\frac{s\log(K \vee n)}{n}\right)$$

- Sample-split IV employing LASSO is consistent if

$$\frac{\mu^2}{\sqrt{s\log(K \vee n)}} \rightarrow \infty$$

Many Weak IV: Are Inferences Standard?

$$\widehat{\beta} - \beta = \frac{\sum_{i=1}^{n} f_i e_i + \sum_{i=1}^{n} (\widehat{f_i} - f_i) e_i}{\sum_{i=1}^{n} f_i X_i + \sum_{i=1}^{n} (\widehat{f_i} - f_i) X_i}$$

- Consistency: when $\sum_{i=1}^{n} f_i X_i = O_p(\mu^2)$ dominates other terms

$$\frac{\mu^2}{\sqrt{r_n}} \to \infty$$

- Standard Gaussian inferences (with usual standard errors): when $\sum_{i=1}^{n} f_i e_i = O_p(\mu)$ dominates $\sum_{i=1}^{n} (\widehat{f_i} - f_i) e_i = O_p(\sqrt{r_n})$

$$\frac{\mu^2}{r_n} \to \infty$$

- Whether asymptotic inferences are standard depends on the estimation rate of the first stage method

Many Weak IV: Are Inferences Standard?

For JIVE-type estimators:

- First stage mistake is average of v_j's: $\widehat{f}_i - f_i \approx \sum_{j\neq i} \widetilde{P}_{ij} v_j$
- Quadratic form CLT (Chao et al, 2012)

$$\frac{1}{\sqrt{K}} \sum_{i=1}^{n} (\widehat{f}_i - f_i) e_i \approx \frac{1}{\sqrt{K}} \sum_{i=1}^{n} \sum_{j\neq i} \widetilde{P}_{ij} v_j e_i \Rightarrow N(0, \Sigma)$$

- Different formulas for asymptotic variance of $\widehat{\beta}_{JIVE}$ (robust- converge to the standard ones if $\frac{\mu^2}{r_n} \to \infty$)

Many Weak IV: Are Inferences Standard?

Split-sample: $\widehat{\beta}_{SS} = \frac{\sum_{i \in I_2} \widehat{f}(\mathcal{A}_1, Z_i) Y_i}{\sum_{i \in I_2} \widehat{f}(\mathcal{A}_1, Z_i) X_i}$,

- Use conditioning argument (on the first subsample $\mathcal{A}_1$ and all instruments $\mathcal{Z}_2$)
- Second stage is just identified: once the estimator is consistent, it is Gaussian with the usual (conditional) standard errors
- Effectively using only half of the sample (efficiency loss?)

Many Weak IV: Are Inferences Standard?

Cross-fit: $\widehat{\beta}_{CFSS} = \frac{\sum_{i\in I_1}\widehat{f}(\mathcal{A}_2,Z_i)Y_i+\sum_{i\in I_2}\widehat{f}(\mathcal{A}_1,Z_i)Y_i}{\sum_{i\in I_1}\widehat{f}(\mathcal{A}_2,Z_i)X_i+\sum_{i\in I_2}\widehat{f}(\mathcal{A}_1,Z_i)X_i}$

- Gaussianity is an open question: conditioning is not possible
- Complicated cross-dependence of terms: $\sum_{i\in I_1}(\widehat{f}(\mathcal{A}_2, Z_i) - f_i)e_i + \sum_{i\in I_2}(\widehat{f}(\mathcal{A}_1, Z_i) - f_i)e_i$

Many Weak IV: Pretest for Weak IV?

- Practitioners want to know if Gaussian inferences are reliable
  - Use a pretest to choose between Gaussian inferences (strong enough signal) and robust (weak)
- One approach:
  - Derive the distribution of t-statistic, when the estimator in question is NOT consistent
  - Find the parameter that governs size distortions
  - Create an empirical indicator that assesses size distortions
- For JIVE this is done in Mikusheva and Sun (2020)
- For Sample-split we can do (heteroscedasticity-robust) first stage F pretest by conditioning argument

Many Weak IV: Robust Testing

- Identification-robust tests control size when the signal is low
- Problem of testing many moment conditions:

$$H_0 : \quad \mathbb{E}[(Y_i - \beta_0 X_i) Z_i] = 0, \quad Z_i \in \mathcal{Z}$$

- How to combine moments in the most informative way?
- Max score test works great for sparse models: power comes from max coefficient (Belloni at al, 2012)
- JIVE-type quadratic form works well when all directions of instruments are possible (Mikusheva and Sun, 2020)
- New suggestion: sample split
  - find optimal instrument on one subsample
  - use identification robust test (AR) on the other

Many Weak IV: Summary

- Flexible first stage may lead to endogenously estimated instrument. Mitigating approaches: sample-split, jackknife
- There is a trade-off between information from first and second stage
- Asymptotics can be cumbersome and depend on the asymptotics of the first stage
- Sample-split has the cleanest inferences due to conditioning argument

Overview

1. Cross-sectional: many weak IV
2. Time Series
    - Challenges in Time Series
    - Factor Models
    - To Do List

Challenges in Time Series

1. Structural errors e_t are autocorrelated
   - Need HAC-robust standard errors
   - Concept of optimality is complicated: exploit dependence (Hansen, 1985, Anatolyev, 2007)
2. Weak exogeneity: $\mathbb{E}[e_t | Z_t, Z_{t-1}, ...] = 0$
   - In TS strict exogeneity: $\mathbb{E}[e_t|$ all $Z_s] = 0$ almost never holds
   - Strict exogeneity allows inferences 'conditional on instruments' i.e. treat instruments as fixed
   - With weak exogeneity we should not mix up observations from different time periods (no GLS!)

Time Series: Is Estimated Instrument Exogenous?

Three approaches to getting exogenously estimated instrument

1. Estimate the optimal instrument without using X_t. YES! Well developed. Factor Models and Factor IV
2. Sample split: should work. Not much work is done
3. Jackknife: may be. New way of jackknifing - increasing window

Time Series: Factor Models

$$Z_{it} = \lambda_i' F_t + \epsilon_{it}$$

- Estimation by Principle Components (PCA), test for number of factors (Bai and Ng, 2002)
- 7 dynamic factors for US economy (Stock and Watson, 2005)
- Factor IV (Bai and Ng, 2010): (1) do PCA and (2) use factors as instruments
- Main motivating assumption: if

$$X_t = \mu' F_t + v_t,$$

then Factor IV attains semi-parametric efficiency

Time Series: Factor IV

- Selection of instruments (factors) is done by Principle components on Z, without using X_t
  - Pros: selected instruments are exogenous (almost); Weak IV literature results are applicable (Kapetanos and Mercellino, 2010)
  - Cons: factors that best explain variation in Z are not always best in explaining X
- To bring back predictive power, solutions in consideration include (Bai and Ng, 2009):
  - Boosting
  - Ordering instruments by predictive power
  - Information criteria
- Revive the question: is estimated instrument exogenous?

Time Series: Is Estimated Instrument Exogenous?

Two potential solutions:

1. Sample-split (one application is Anatolyev and Mikusheva, 2020)
    - Use the past subsample for instrument selection, and the recent for structural estimation
    - $\widehat{f}_t$ is in the correct information set (exogenous)
    - Just identified model $\Rightarrow$ pretest for weak IV, robust tests
2. Jackknife
    - Direct form of JIVE is inapplicable (weak exogeneity, cannot use future values of instruments)
    - Increasing window: for $\widehat{f}_t$ can use the sample up to $t-1$ (or some lag to account for autocorrelation)

Time Series: To Do List

To do item 1: Find good methods for the optimal instrument selection. Promising approaches:

- Partial Least Squares and Ridge (Carrasco and Rossi, 2016, contains asymptotic results on the speed of convergence)
- Boosting (Luo and Spindler, 2016, contains asymptotic results on the speed of convergence for cross-sectional)
- LASSO in time series (Babii et al.,2019)
- Bayesian Model Averaging
- Kernel weighted IV (Kuersteiner, 2001)

Time Series: To Do List

To do item 2: Inference (Sample-split)

- Inferences conditional on the initial sample
- Pretest on weak ID is available
- Robust inference is available
- Concern: not full use of the sample

Time Series: To Do List

To do item 3: New asymptotic theory needed for increasing window jackknifing

- For theoretical justification we need asymptotic results on $\sum_t \widehat{f}_t e_t$. Wishful thinking - CLT
- For increasing window jackknife with OLS first stage - CLT for quadratic forms in time series (U-statistics)
- Li and Liao (2020) - strong approximation in time series and non-parametrics

Summary

- Very flexible first stage comes with costs (overfitting leads to endogenously estimated optimal instrument)
- Sample-split and JIVE are good ways to mitigate endogeneity
- In time series first stage should respect weak exogeneity condition
- Good first stage forecasting may help with identification
- There is a need for asymptotic results for ML