Discussion of paper “Leave out estimation of variance components” by Patrick Kline, Raffaele Saggio and Mikkel Solvsten

Anna Mikusheva

MIT, Department of Economics

Typical structure:

- Data is used to arrive at estimator $\widehat{\beta}$ for a parameter β.
- Estimator $\widehat{\beta}$ is close to gaussian (OLS or averages):

$$\widehat{\beta} - \beta \approx N(0, \Sigma).$$

- Parameter of interest $\theta = F(\beta)$.
- Example (IV): $\theta_{IV} = \frac{\beta_1}{\beta_2}$.
- This paper $\theta = \beta' A \beta$.
- One will have difficulties with statistical analysis (bias, unusual inference) when F is significantly **non-linear** in the area of uncertainty of β.

Bias

- Simplistic example: $\widehat{\beta} = \beta + \sigma_n \xi$, $\xi \sim N(0,1)$, parameter of interest $\theta = \beta^2$.
- Naive estimate

$$\widehat{\theta} = \left(\widehat{\beta}\right)^2 = (\beta + \sigma_n \xi)^2 = \theta + 2\beta\sigma_n\xi + \sigma_n^2 \xi^2.$$

- It contains bias: $E\widehat{\theta} = \theta + \sigma_n^2$.
- It is usually called "finite-sample bias": if $\sigma_n = \frac{const}{\sqrt{n}}$, the bias is of order $\frac{c}{n}$.
  - It does not have to be small.
  - Relative bias $\frac{\sigma_n^2}{\theta} = \left(\frac{\beta}{\sigma_n}\right)^{-2}$ connected to uncertainty about β in relation to its impact.
  - If $\beta = (\beta_1, ..., \beta_k)$ and $\theta = \sum_{i=1}^k \beta_i^2$, bias accumulates: $E\widehat{\theta} = \theta + \sum_{i=1}^k \sigma_{ni}^2$

Bias correction

- $\widehat{\beta} = \beta + \sigma_n \xi$, $\xi \sim N(0,1)$, $\theta = \beta^2$

$$E\widehat{\theta} = \theta + \sigma_n^2$$

- Natural way to correct bias is $\widetilde{\theta} = \left(\widehat{\beta}\right)^2 - \widehat{\sigma}^2$
- The paper discusses when β is many-dimensional, $\theta = \beta' A \beta$ and data is heteroskedastic: what a good estimate of σ's is.
- The solution is leave-one-out:
  - Very clean and convincing argument
  - Approach very successfully used in many weak IV literature

Inferences

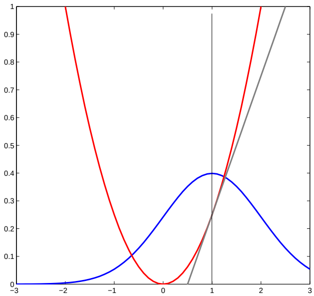
- Corrected estimator

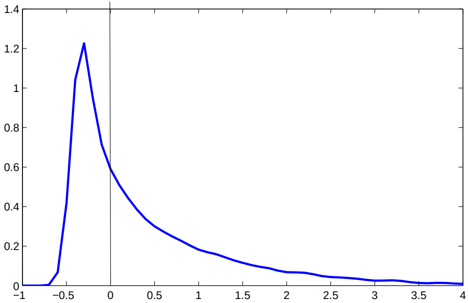
$$\widetilde{\theta}-\theta=\left\{\left(\widehat{\beta}\right)^2-\sigma_n^2\right\}-\theta=\underbrace{2\beta\sigma_n\xi}_{gaussian}+\underbrace{\sigma_n^2(\xi^2-1)}_{centered\quad\chi_1^2}$$

- Relative importance of two components is $\frac{\beta\sigma_n}{\sigma_n^2}=\frac{\beta}{\sigma_n}$ connected to the size of β relative to its uncertainty.
- Standard inferences based on Delta-method:

$$\widetilde{\theta}-\theta\approx\underbrace{2\beta\sigma_n\xi}_{gaussian}.$$

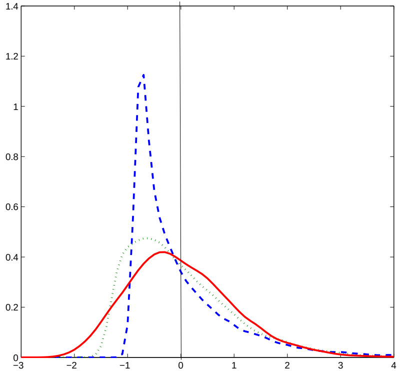
Inferences





Here $\widehat{\beta} \sim N(1, 1), \theta = \frac{1}{4}\beta^2$.
The distribution on the right is properly centered (at zero) and normalized to have variance 1.

Inferences



Change uncertainty about β : $\widehat{\beta} \sim N(1, \sigma^2)$.
Blue ($\sigma^2 = 1$), Green ($\sigma^2 = 0.2$), Red ($\sigma^2 = 0.04$)

Inference

- In the application β is very multi-dimensional, $\theta = \beta' A \beta$.
- High dimension of β may help: if $\theta = \sum_{i=1}^{k} \beta_i^2$ and all $\widehat{\beta}_i$ are stochastically of the same size (asymptotically negligible), the CLT will bring gaussianity back (as $k \to \infty$).
- Problem occurs when:
  θ strongly depends on a few linear combinations of β that are imprecisely estimated relative to overall uncertainty in $\widehat{\theta}$.

Application

- Two-way fixed effect

$$y_{gt} = \alpha_g + \psi_{j(g,t)} + \varepsilon_{gt}.$$

- Individual fixed effect α_g and firm fixed effect ψ_j cannot be separately identified, – they come as a sum
- If there are workers who moved between Firm 1 and Firm 2, then $\psi_1 - \psi_2$ is identified by $E(y_{gt_1} - y_{gt_2})$.
- Uncertainty of $\widehat{\psi}_1 - \widehat{\psi}_2$ is connected to how many workers moved.
- If some workers moved between Firm 1 and Firm 2, and some moved between Firm 2 and Firm 3, then $\psi_1 - \psi_3$ is identified even if none moved between Firm 1 and Firm 3.

Application

- This way you can uncover variation $\psi_j - \psi_{j^*}$ over all "connected" firms. But the structure of uncertainty is cumbersome.
- Goal: to estimate $\theta = Var(\psi_j)$. It is identified if all firms are "connected".

Application

- Imagine that there are two "clusters" of firms; firms are tightly connected within cluster but not between

$$\theta = \omega_1 Var(\psi_j; j \in cl_1) + \omega_2 Var(\psi_j; j \in cl_2) + a(\overline{\psi}_{cl_1} - \overline{\psi}_{cl_2})^2.$$

- Between-cluster-difference $\overline{\psi}_{cl_1} - \overline{\psi}_{cl_2}$ has strong influence on θ.
- If only a few workers moved between clusters ('bottleneck'), this component is poorly estimated (a.k.a. weakly identified).
- Problem occurs when:
  θ strongly depends on a few linear combinations of β that are imprecisely estimated relative to overall uncertainty in $\widehat{\theta}$.

Application

- Complication: the network structure of "connected" firms is complicated.
- Potential problem depends on many unknowns: the effect of a linear component, its uncertainty, its relation to other components.
- Method of inference (Andrews and Mikusheva (2016)) measure the curvature of F function in relation to covariance, choosing the direction of 'worst curvature' (determining the problematic direction); then adjust critical values accordingly.
- Method is agnostic - it does not require knowledge of the problem location.
- This method is nicely executed; it demonstrates problem exactly where you would expect it: no curvature when estimation is done over one well-connected region and pronounced curvature when estimation is done over two poorly connected regions.