Discussion of "Minimizing Sensitivity to Model Misspecification" by Stéphane Bonhomme and Martin Weidner

Kirill Evdokimov

M.I.T.

Setup

- *Reference* model: η and $\theta(\eta)$
- *True* model: $\theta_0 : f_{\theta_0}(Y)$
- $\Gamma_\epsilon(\eta) = \{\theta_0 \in \Theta : d(\theta_0, \theta(\eta))\}$
- Find $\tilde{\delta}$ that minimizes Worst Case MSE:

$$\text{WCMSE}\left(\tilde{\delta}\right) = \sup_{\theta_0 \in \Gamma_\epsilon(\eta)} E_{\theta_0}\left[\left(\tilde{\delta} - \delta(\theta_0)\right)^2\right]$$

- Choice of ϵ
  - prior knowledge?
  - $\epsilon = \epsilon(p)$ the probability of a model detection error

Also In the Paper

- Bayesian interpretation with Gaussian priors
  - Small ϵ analysis
  - Also least favorable priors

- Example of treatment effect estimation under unconfoundedness
  - the "double-robust" moments arise from the procedure

Comments

- Two cases: regular identification and not
- Regular case: one-step $\hat{\theta}_{\text{"MLE"}}$
  - can compute, why not?
  - MSE
- Multiple potentially "true" models?
- Multiple reference models?

Comments

- Two cases: regular identification and not
- One-step $\hat{\theta}_{\text{“MLE”}}$
- In principle $\tilde{\delta}_n \equiv \hat{\lambda}_n \delta\left(\hat{\theta}_{\text{“MLE”}}\right) + \left(1 - \hat{\lambda}_n\right) \delta\left(\theta\left(\hat{\eta}\right)\right)$
- Minimizing Worst Case MSE:
  - when $\epsilon n \to c$ should have

$$\text{WCMSE}\left(\tilde{\delta}\right) = \sup_{\theta_0 \in \Gamma_\epsilon(\eta)} E_{\theta_0}\left[\left(\tilde{\delta} - \delta\left(\theta_0\right)\right)^2\right] \lesssim \frac{1}{n} \tag{1}$$

  - even when $nE\left(\delta\left(\hat{\theta}_{\text{“MLE”}}\right) - \delta\left(\theta_0\right)\right)^2 \to \infty$
- The paper links $\delta\left(\hat{\theta}_{\text{“MLE”}}\right)$ and $\delta\left(\theta\left(\hat{\eta}\right)\right)$ nonlinearly, via a regularization
- Choice of ϵ in the irregular case

Comments

- Asymptotically linear estimators - very natural class of estimators to consider

- But what if identification is irregular/partial/weak?

- IV example
  - known v.s. unknown Π
  - $|\rho_{UV}| \sim \sqrt{\epsilon} \sim n^{-1/2}$

- Instead?
  - Short $Y_i = X_i \eta + U_i$
  - Long $Y_i = X_i \delta + W_i \beta_W + V_i$