Functional Generalized Empirical Likelihood Estimation for Conditional Moment Restrictions

Heiner Kremer[1], Jia-Jie Zhu[2], Krikamol Muandet[1] and Bernhard Schölkopf[1]

[1]Max-Planck-Institute for Intelligent Systems, Tübingen, Germany
[2]Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Conditional Moment Restrictions

Conditional moment restrictions identify a parameter θ_0 via:

$$E[\psi(X;\theta_0)\,|\,Z] = 0 \quad P_Z\text{-a.s.},$$

with $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^n$ being an integrable function.

Conditional Moment Restrictions

Conditional moment restrictions identify a parameter θ_0 via:

$$E[\psi(X;\theta_0)\,|\,Z] = 0 \quad P_Z\text{-a.s.},$$

with $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^n$ being an integrable function.

Examples:

- Instrumental variable regression
- Off-policy evaluation in RL
- Double/Debiased ML

Conditional Moment Restrictions

Conditional moment restrictions identify a parameter θ_0 via:

$$E[\psi(X;\theta_0) \mid Z] = 0 \quad P_Z\text{-a.s.},$$

with $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^n$ being an integrable function.

Examples:

- Instrumental variable regression
- Off-policy evaluation in RL
- Double/Debiased ML

Equivalent unconditional moment restrictions:

$$E[\psi(X;\theta_0)^\top h(Z)] = 0 \quad \forall h \in \mathcal{H}$$

$\Rightarrow$ Requires methods which can handle continua of moment restrictions

Functional GEL

$E_{P_0}[\psi(X;\theta_0)^\top h(Z)] = 0 \ \forall h \in \mathcal{H}$ $+$ Sample $\hat{P}_n$ from P_0

Functional GEL

$E_{P_0}[\psi(X;\theta_0)^\top h(Z)] = 0 \ \forall h \in \mathcal{H}$

$+$

Sample $\hat{P}_n$ from P_0

$\Downarrow$

Primal GEL problem:

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \inf_{P \in \mathcal{P}} D_f(P || \hat{P}_n) \quad \text{s.t.} \quad E_P[\psi(X;\theta)^\top h(Z)] = 0 \quad \forall h \in \mathcal{H}$$

Functional GEL

$$E_{P_0}[\psi(X;\theta_0)^\top h(Z)] = 0 \ \forall h \in \mathcal{H}$$

$+$

Sample $\hat{P}_n$ from P_0

$\Downarrow$

Primal GEL problem:

$$\hat{\theta} = \operatorname*{argmin}_{\theta\in\Theta} \inf_{P\in\mathcal{P}} D_f(P||\hat{P}_n) \quad \text{s.t.} \quad E_P[\psi(X;\theta)^\top h(Z)] = 0 \quad \forall h \in \mathcal{H}$$

$\Downarrow$ Relaxation

$$\hat{\theta} = \operatorname*{argmin}_{\theta\in\Theta} \inf_{P\in\mathcal{P}} D_f(P||\hat{P}_n) \quad \text{s.t.} \quad \sup_{h\in\mathcal{H},\ \|h\|\leq 1} E_P[\psi(X;\theta)^\top h(Z)] \leq \lambda$$

Functional GEL

$$E_{P_0}[\psi(X;\theta_0)^\top h(Z)] = 0 \ \forall h \in \mathcal{H}$$

$+$

Sample $\hat{P}_n$ from P_0

$\Downarrow$

Primal GEL problem:

$$\theta = \operatorname*{argmin}_{\theta\in\Theta} \inf_{P\in\mathcal{P}} D_f(P||\hat{P}_n) \quad \text{s.t.} \quad E_P[\psi(X;\theta)^\top h(Z)] = 0 \quad \forall h \in \mathcal{H}$$

$\Downarrow$ Relaxation

$$\theta = \operatorname*{argmin}_{\theta\in\Theta} \inf_{P\in\mathcal{P}} D_f(P||\hat{P}_n) \quad \text{s.t.} \quad \sup_{h\in\mathcal{H},\ \|h\|\leq 1} E_P[\psi(X;\theta)^\top h(Z)] \leq \lambda$$

$\Downarrow$ Duality

$$\theta^{\text{FGEL}} = \operatorname*{argmin}_{\theta\in\Theta} \sup_{h\in\mathcal{H}} E_{\hat{P}_n}\left[\phi\Big(\psi(X;\theta)^\top h(Z)\Big)\right] - \frac{\lambda_n}{2}\|h\|^2_{\mathcal{H}},$$

Properties of FGEL Estimators

- Consistency, asymptotic normality
- Allows using f-divergences beyond χ^2 ($\hat{=}$ GMM)
- Contains variants of related estimators as special cases
- Can leverage arbitrary ML models (we investigate kernel- and neural network-based implementations)
- Exhibits SOTA performance on conditional moment restrictions tasks