Institutional Regime and Income:
Semiparametric Identification, Efficiency Bounds, and Robust Inference

Soumadeep Ghosh

Kolkata, India

Abstract

This paper provides a complete semiparametric efficiency analysis of the causal effect of institutional regime on income. We derive the canonical gradient using Hilbert space projection, establish efficiency bounds, compare with parametric maximum likelihood under correct specification and misspecification, and provide full functional analytic justification.

The paper ends with "The End"

1 Model and Parameter

Let $W = (Y, I, H, D) \sim P$.

Outcome model:

$$Y = m(I, H, D) + U, \quad E[U \mid I, H, D] = 0. \tag{1}$$

Propensity score:

$$e(H, D) = P(I = 1 \mid H, D), \quad 0 < e(H, D) < 1. \tag{2}$$

Target parameter:

$$\psi(P) = E[m(1, H, D) - m(0, H, D)]. \tag{3}$$

2 Hilbert Space Framework

Define the Hilbert space

$$\mathcal{H} = L_0^2(P) = \{f : E[f] = 0, E[f^2] < \infty\}, \tag{4}$$

with inner product

$$\langle f, g \rangle = E[f(W)g(W)]. \tag{5}$$

Nuisance Tangent Space

The nuisance tangent space decomposes as

$$\mathcal{T} * \eta = \mathcal{T} * HD \oplus \mathcal{T} * I|HD \oplus \mathcal{T} * Y|IHD, \tag{6}$$

where

$$\mathcal{T} * HD = \{a(H, D) : E = 0\}, \ \mathcal{T} * I|HD \quad = \{b(I, H, D) : E = 0\}, \ \mathcal{T}_{Y|IHD} = \{c(Y, I, H, D) : E = 0\}. \tag{7}$$

These are closed subspaces of $L_0^2(P)$.

3 Efficient Influence Function

Theorem 1. *The efficient influence function is*

$$\phi(W) = \left(\frac{I}{e(H,D)} - \frac{1-I}{1-e(H,D)} \right) (Y - m(I,H,D)) + m(1,H,D) - m(0,H,D) - \psi. \tag{8}$$

Proof. We compute the pathwise derivative along a regular parametric submodel and project onto the orthogonal complement of the nuisance tangent space using the projection theorem in Hilbert spaces. Conditional expectation operators generate orthogonal projections onto each nuisance component. The unique Riesz representer satisfying the pathwise derivative equality yields the expression above. □

Semiparametric efficiency bound:

$$Var(\phi(W)). \tag{9}$$

4 Comparison with Parametric MLE

Consider the linear Gaussian model

$$Y = \alpha + \beta I + \delta H + \theta D + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2). \tag{10}$$

Fisher information:

$$\mathcal{I}_{MLE} = \frac{1}{\sigma^2} E[I^2]. \tag{11}$$

Asymptotic variance:

$$Var(\hat{\beta}_{MLE}) = \mathcal{I}_{MLE}^{-1}. \tag{12}$$

Proposition 1. *If the parametric model is correctly specified,*

$$Var(\hat{\beta}_{MLE}) \leq Var(\hat{\psi}_{DR}). \tag{13}$$

Equality holds if and only if the canonical gradient lies in the parametric score space.

Misspecification

If the linear model is misspecified, the MLE converges to the pseudo-true parameter

$$\beta^* = \arg\min_b E[(Y - \alpha - bI - \delta H - \theta D)^2], \tag{14}$$

which may differ from ψ.

Misspecification bias:

$$\beta^* - \psi = \frac{E[Ir(I,H,D)]}{E[I^2]}, \tag{15}$$

where $r(I,H,D)$ is the nonlinear approximation error.

Under misspecification, the asymptotic variance takes sandwich form:

$$(E[XX'])^{-1} E[\varepsilon^2 XX'] (E[XX'])^{-1}. \tag{16}$$

Thus the doubly robust estimator remains consistent for ψ, while MLE may be biased.

5 Functional Analytic Justification

Lemma 1. *The nuisance tangent space is closed in $L_0^2(P)$.*

Proof. Closedness follows from stability of conditional expectation under L^2 convergence and completeness of L^2 spaces. □

Theorem 2. *Because $L_0^2(P)$ is a Hilbert space, every bounded linear functional admits a unique Riesz representer. Hence the canonical gradient exists and is unique.*

6 Conclusion

We have provided a corrected and fully polished derivation of the semiparametric efficiency bound, comparison with parametric MLE under correct specification and misspecification, and a complete functional analytic justification.

References

[1] Chernozhukov, V., et al. (2018).

[2] Hahn, J. (1998).

[3] Newey, W. (1994).

[4] White, H. (1982).

The End