MEC: Machine-Learning-Assisted Generalized Entropy Calibration for Semi-Supervised Mean Estimation

Se Yoon Lee   Jae-Kwang Kim

ICML 2026

International Conference On Machine Learning

Motivation: labels are costly, covariates are abundant

Data structure

$$\{X_i\}_{i=1}^N \sim P_X, \qquad \{(X_j, Y_j) : j \in S\} \sim P,$$

$$|S| = n, \qquad f = \frac{n}{N}.$$

Target

$$\theta_0 = \mathbb{E}(Y).$$

- Labels Y are expensive.
- Unlabeled covariates X are cheap.
- Goal: valid confidence intervals using both samples.

Inference, not just prediction.

Prediction-powered inference: elegant but not always efficient

PPI / CF-PPI form

$$\widehat{\theta}_{\mathrm{PPI}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{m}(X_i) + \frac{1}{n}\sum_{j\in S}\{Y_j - \widehat{m}(X_j)\}.$$

Problem 1: label reuse

If the same labels train $\widehat{m}$ and evaluate residuals, flexible ML can overfit.

Problem 2: efficiency shortfall

Even with cross-fitting, misspecified predictions can inflate variance.

MEC keeps cross-fitting and adds calibration to resolve two problems.

Angelopoulos, A. N., Bates, S., Fannjiang, C., Jordan, M. I., & Zrnic, T. (2023). Prediction-powered inference. *Science*, 382(6671), 669–674.
Zrnic, T., & Candès, E. J. (2024). Cross-prediction-powered inference. *Proceedings of the National Academy of Sciences*, 121(15).

MEC idea: use ML predictions as calibration information

Cross-fitted (out-of-fold) predictor basis

$$h(X) = \{1, \widehat{m}^{(-)}(X)\}.$$

Calibration constraints

$$\sum_{j \in S} \widehat{\omega}_j = N, \qquad \sum_{j \in S} \widehat{\omega}_j \widehat{m}^{(-)}(X_j) = \sum_{i=1}^{N} \widehat{m}^{(-)}(X_i).$$

MEC estimator

$$\widehat{\theta}_{\mathrm{MEC}} = \frac{1}{N} \sum_{j \in S} \widehat{\omega}_j Y_j.$$

MEC balances the out-of-fold prediction between the labeled sample and the full covariate sample.

Why MEC helps?

Key idea

MEC replaces a condition on the *raw prediction error* with a condition on the *projection error* onto the calibration subspace.

CF-PPI

Raw prediction error:

$$\left\|\widehat{m}^{(-)} - m_0\right\|_{L_2(P_X)} = o_p(1).$$

MEC

Projection error:

$$\mathcal{W} = \operatorname{span}\{1, \widehat{m}^{(-)}\}, \qquad m_{0,\perp} = m_0 - \Pi_{\mathcal{W}} m_0.$$

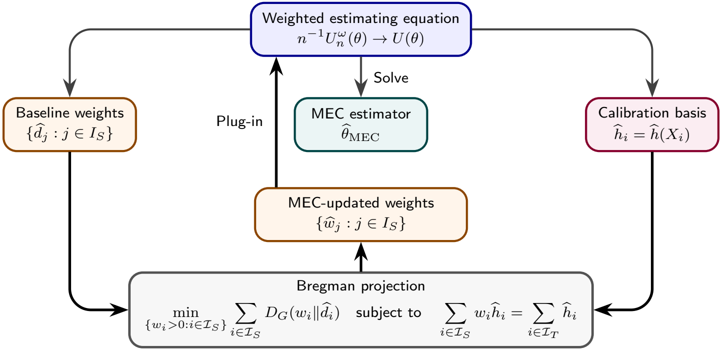
$$\left\|m_{0,\perp}\right\|_{L_2(P_X)} = o_p(1).$$

Why this is good?

$$\left\|m_0 - \Pi_{\mathcal{W}} m_0\right\|_{L_2(P_X)} \le \left\|\widehat{m}^{(-)} - m_0\right\|_{L_2(P_X)}.$$

Thus, MEC can remain efficient when $m_0 \approx a + b\widehat{m}^{(-)}$, even if $\widehat{m}^{(-)}$ is not perfectly precise.

Generalization



Semi-supervised mean

$$U_{n,N}^{\omega}(\theta; m) = \sum_{i\in\mathcal{I}_T} m(X_i) - \theta + \sum_{j\in\mathcal{I}_S} \omega_j(Y_j - m(X_j)) = 0$$

Average treatment effect

$$U_n^{\omega_1,\omega_0}(\theta) = \sum_{i\in\mathcal{I}_{S,1}} \omega_{1i}Y_i \sum_{i\in\mathcal{I}_{S,0}} \omega_{0i}Y_i - |\mathcal{I}_T|\theta = 0$$

ATT-marginal hazard

$$U_n^{\omega}(\theta) = \sum_{i\in\mathcal{I}_T\cup\mathcal{I}_S} \int \omega_i(A_i - \bar{A}_{\omega}(t;\theta))dN_i(t) = 0$$

Practical considerations

Choice of calibration basis

The calibration basis h determines both the estimator and its robustness properties.

Default choice (this paper):

$$h(x) = \{1, \widehat{m}^{(-)}(x)\}.$$

Richer choices are also possible:

$$h(x) = \{1, \widehat{m}^{(-)}(x), \widehat{m}^{(-)}(x)^2\},$$

or multiple ML predictors:

$$h(x) = \{1, \widehat{m}_1^{(-)}(x), \widehat{m}_2^{(-)}(x)\}.$$

This can improve robustness, but may increase weight instability.

When efficiency gains are limited

Efficiency gains from MEC are not automatic for every target parameter.

MEC is most useful when the calibration basis captures outcome- or score-relevant variation in the estimating equation.

For semi-supervised mean estimation, this is natural because X helps predict Y.

If the estimating equation has little component predictable from X, then calibration has limited room to reduce variance.

Takeaway

One-sentence summary

MEC is a cross-fitted calibration framework that uses machine-learning predictions to construct a calibration basis for updating weights, thereby enabling valid and efficient inference.

Easy to implement

MEC does not require a problem-specific derivation of the efficient influence function for implementation.

Efficient

Efficiency gains can arise when the ML-based calibration basis captures outcome- or score-relevant variation.

Generalizable

Extends naturally to a broad class of weighted estimating equations, including missing-data problems, causal inference problems, and weighted Cox regression.

Thank you!