Poverty Targeting with Imperfect Information

Juan C. Yamin
Brown University

2025 World Congress of the Econometric Society

August 22, 2025

Motivation

- Policymakers do not observe *true household incomes*.

- Instead, they use *income estimates* constructed from observable data (e.g., housing quality, assets, location).

- **Problem:** These estimates are noisy $\Rightarrow$ large targeting errors (Brown et al., 2018)

- This paper shifts focus:
    - Not: producing better estimates
    - But: designing transfer allocations that are *optimal* given noisy estimates

Goal: Use existing income estimates more effectively...without new data or added costs.

This Paper

1. **Decision-theoretic framework** for poverty targeting with *income estimates*
   - Planner chooses transfers subject to: (i) no taxation; (ii) a budget constraint
   - Allocations based only on noisy household income estimates
2. **Plug-in rule is inadmissible**
   - Plug-in = projection of estimated poverty gaps onto the budget simplex
   - A James–Stein type estimator dominates $\Rightarrow$ plug-in cannot be *optimal*
3. **Empirical Bayes rule**
   - Even though the optimal allocation is a *nonsmooth function* of posterior means, the EB rule achieves Bayes regret rates that match those for estimating the posterior mean

Roadmap of Talk

Targeting Model

Inadmissibility of Plug-in Rule

(Empirical) Bayes Rule

Simulation

Bibliography

Poverty Targeting Under Perfect Information

Planner's Problem (Full Information)

- Planner observes true incomes $\boldsymbol{y} = (y_1, \ldots, y_n)$
  - Fixed poverty line z
  - Abstract from labor responses, household size, and spillovers
- Objective: allocate a fixed budget B through nonnegative transfers $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$
  - Constraints: $\tau_i \geq 0$ and $\sum_{i=1}^{n} \tau_i \leq B$
- The planner wants to prioritize the poorest of the poor (Foster et al., 1984):

$$\min_{\boldsymbol{\tau}=(\tau_1,\ldots,\tau_n)} \frac{1}{n} \sum_{i=1}^{n} \left(z - y_i - \tau_i\right)^2 \text{ subject to (i) } \sum_{i}^{n} \tau_i \leq B, \text{(ii) } \tau_i \geq 0, \forall i \qquad (1)$$

Loss function

Perfect Information: Optimal Policy

Proposition (Complete Information Policy)

The optimal transfer rule can be described in two equivalent ways:

1. ***KKT conditions:***

$$\tau_i^\star = \max\left\{0,\ z - y_i - \tfrac{\lambda}{2}\right\}, \quad i = 1, \ldots, n$$

   where λ is the shadow price of the budget.

2. ***Geometric projection:***

$$\boldsymbol{\tau}^\star = \arg\min_{\boldsymbol{\tau} \in \mathcal{A}} \left\| \boldsymbol{\tau} - (\mathbf{1}z - \mathbf{y}) \right\|^2, \quad \mathcal{A} = \{\boldsymbol{\tau} \in \mathbb{R}_+^n : \sum \tau_i \leq B\}.$$

$\Rightarrow$ Transfers go to the poorest first; only the bottom households receive transfers.

Imperfect Information

Statistical Decision Theory

For each i, the planner observes an estimate and has to determine transfer τ_i

Statistical Model: Motivated by CLT, estimates $\hat{Y}_i \mid \mu_i, \sigma_i \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$, where σ_i is known and $\mu_i = \mathbb{E}[y_i \mid W_i] \in \mathbb{R}_+$

- Geographic: Target based on average income ▸ Details
- Proxy means test: Target based on prediction from linear regression ▸ Details

Unknown parameters:

- Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ be thouseholds' conditional mean income
- $\boldsymbol{\mu}$ belongs to the parameter space $\Theta = \{\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^\top \mid \boldsymbol{\mu} \in \mathbb{R}_+^n\}$

Action space: No taxation + Budget constraint

$$\mathcal{A} = \left\{ \boldsymbol{t} \in \mathbb{R}_+^n : \mathbf{1}^\top \boldsymbol{t} \leq B \right\}.$$

Imperfect Information

Statistical Decision Theory

Loss function: $L : \mathcal{A} \times \Theta \to \mathbb{R}$ ▸ Loss equivalence

$$L(\boldsymbol{t}, \boldsymbol{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \left(z - \mu_i - t_i\right)^2$$

Decision rule: $\boldsymbol{\tau} : \mathbb{R}^n \to \mathcal{A}$

- For each observed value of the data $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ decides what action of $\mathcal{A}$ to take
- Set of decision rules: $\mathcal{D} = \{\boldsymbol{\tau} : \mathbb{R}^n \to \mathcal{A}\}$

Statistical Decision Problem: $\{\Theta, \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^{\mathbf{2}}), \mathcal{A}, L\}$

Risk Function: $\mathcal{R} : \mathcal{D} \times \Theta \to \mathbb{R}$

$$\mathcal{R}(\boldsymbol{\tau}, \boldsymbol{\mu}) = \mathbb{E}_{\hat{\boldsymbol{Y}}}[L(\boldsymbol{\tau}(\hat{\boldsymbol{Y}}), \boldsymbol{\mu})].$$

Roadmap of Talk

Inadmissibility of Plug-in Rule

Plug-in Rule

- Define the observed poverty gap:

$$X_i = z - \hat{Y}_i \sim N(\theta_i = z - \mu_i,\ \sigma_i^2), \quad i = 1, \ldots, n$$

- **Decision problem:** Choose rule $\tau(X) \in \mathcal{A}$ to minimize

$$\| \boldsymbol{\theta} - \tau(X) \|^2,$$

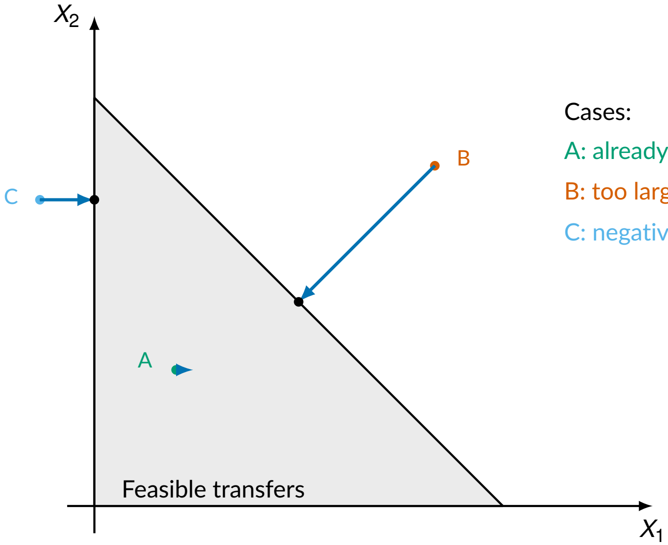
  given $X \sim N(\boldsymbol{\theta}, \mathrm{diag}(\sigma^2))$.

- The plug-in rule assumes the observed X is the truth and optimizes as if $\theta = X$

- Formally, the plug-in estimator is the projection of X onto $\mathcal{A}$:

$$\tau^{\text{plug}}(X) = \arg\min_{w \in \mathcal{A}} \| w - X \|.$$

Plug-in Rule = Project the observed gaps onto the feasible set



Cases:

A: already feasible

B: too large $\Rightarrow$ trim to budget line

C: negative coord. $\Rightarrow$ clip to axis

Intuition: Plug-in treats the estimate as truth, then projects to the nearest feasible allocation.

Inadmissibility of Plug-in Rule

Why plug-in is problematic: It takes noisy income gaps at face value. A few extreme estimates can absorb most of the budget, leading to high variance allocations.

Proposition

Under homoskedasticity, if $\mathbb{P}\left(s > 3,\ 0 < \sigma^2(s-3) < \|\tau^{\text{plug}}(X)\|^2\right) > 0$, then the James–Stein shrinkage rule

$$\tau^{\text{JS}}(X) = \left(1 - \frac{\sigma^2(s-3)}{\|\tau^{\text{plug}}(X)\|^2}\right)_{\text{shrink}} \cdot \tau^{\text{plug}}(X),$$

where $(x)_{\text{shrink}} = x\ (0 < x < 1),\ 1$ otherwise, strictly dominates τ^{plug}. Hence, the plug-in rule is inadmissible.

- Shrinkage $\Rightarrow$ reduces variance more than it increases bias.
- Keeps the same households, but with moderated transfer sizes.

Roadmap of Talk

From Inadmissibility to Empirical Bayes

- The plug-in rule is inadmissible. This rules out the natural benchmark; the question is what to do instead.

- In related estimation problems, a successful approach has been *Empirical Bayes (EB)*: treat the unknown means μ_i as draws from a common distribution G, and allocate based on their posterior expectations.

- To adapt EB here, we condition on known variances σ_i^2 and assume:

$$\mu_i \mid \sigma_i \overset{\text{i.i.d.}}{\sim} G, \quad i = 1, \ldots, n.$$

Oracle Bayes Rule

Proposition (Known prior G, known variances)

Assume (i) σ_i^2 are known and fixed; (ii) $\mu_i \mid \sigma_i \overset{\text{i.i.d.}}{\sim} G$. For each realization $\hat{\boldsymbol{y}}$, the Bayes action is

$$t_i^*(\hat{\boldsymbol{y}};\lambda) = \max\left\{0,\ z - \mathbb{E}_G[\mu_i \mid \hat{y}_i, \sigma_i^2] - \tfrac{\lambda}{2}\right\}, \quad i = 1, \ldots, n,$$

where λ is chosen to satisfy the budget constraint:

$$\lambda^* = \inf\left\{\lambda \geq 0 : \sum_{i=1}^{n} t_i^*(\hat{\boldsymbol{y}};\lambda) \leq B\right\}.$$

- Transfers depend on posterior means $\mathbb{E}_G[\mu_i \mid \hat{y}_i, \sigma_i^2]$
- Allocation goes to households with the largest *posterior expected poverty gaps*

Empirical Bayes Rule (Feasible Implementation)

- Oracle Bayes rule requires the true prior G, which is unknown.
- **Prior estimation:** Use the *nonparametric MLE (NPMLE)*:

$$\hat{G} = \arg\max_{G \in \mathcal{G}} \frac{1}{n} \sum_{i=1}^{n} \log \int \varphi\left(\frac{\hat{y}_i - \mu}{\sigma_i}\right) \frac{1}{\sigma_i}\, dG(\mu),$$

  where φ is the standard normal density.

- For each $\hat{\boldsymbol{y}}$,

$$\hat{t}_i(\hat{\boldsymbol{y}}; \lambda) = \max\left\{0,\ z - \mathbb{E}_{\hat{G}}[\mu_i \mid \hat{y}_i, \sigma_i^2] - \frac{\lambda}{2}\right\}.$$

- EB mirrors the oracle's structure, but shrinks using $\hat{G}$

Bayes Regret: Definition

- Goal: evaluate the performance gap between the feasible EB rule $\hat{\tau}$ and the oracle Bayes rule τ^*.

- Define:

$$\text{Bayes Regret}(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*) = \mathbb{E}\left[L(\hat{\boldsymbol{\tau}}, \boldsymbol{\mu})\right] - \mathbb{E}\left[L(\boldsymbol{\tau}^*, \boldsymbol{\mu})\right].$$

- Measures the additional expected loss from using an estimated prior $\hat{G}$ instead of the true prior G in constructing the decision rule.

Regret Rates

Proposition (Regret rate)

Suppose (i) $\hat{G}$ is an approximate NPMLE of the prior, (ii) the true prior G has bounded p-th moment for some $p \geq 2$, and (iii) variances σ_i^2 are treated as fixed. Then

$$\textit{Bayes Regret}\left(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}^*\right) \lesssim n^{-p/(2+2p)}(\log n)^{(7p+8)/(4+4p)}.$$

- Lighter tails ($p \to \infty$): convergence accelerates toward the parametric rate $n^{-1/2}$.
- Heavy tails (finite p) $\Rightarrow$ slower convergence
- Intuition: the Bayes rule is *1-Lipschitz* in the posterior means, so projecting onto the feasible set cannot increase error. Thus, the Bayes regret inherits the same rates as estimating the posterior mean itself.

Roadmap of Talk

Targeting Model

Inadmissibility of Plug-in Rule

(Empirical) Bayes Rule

Simulation

Bibliography

Simulation: Setup

- Data: Household surveys from 9 African countries (Brown et al. 2018).

- Policymaker observes noisy estimates of conditional expectations:

$$\hat{Y}_{ic} \sim \mathcal{N}\left(\mathbb{E}[y_{ic} \mid W_{ic}],\ \sigma_{ic}^2\right), \quad \sigma_{ic} \sim U[\sigma_c/2, 1.5\sigma_c].$$

- Poverty line $z_c =$ median consumption; Budget $B_c = 10\%$ of poverty gap.

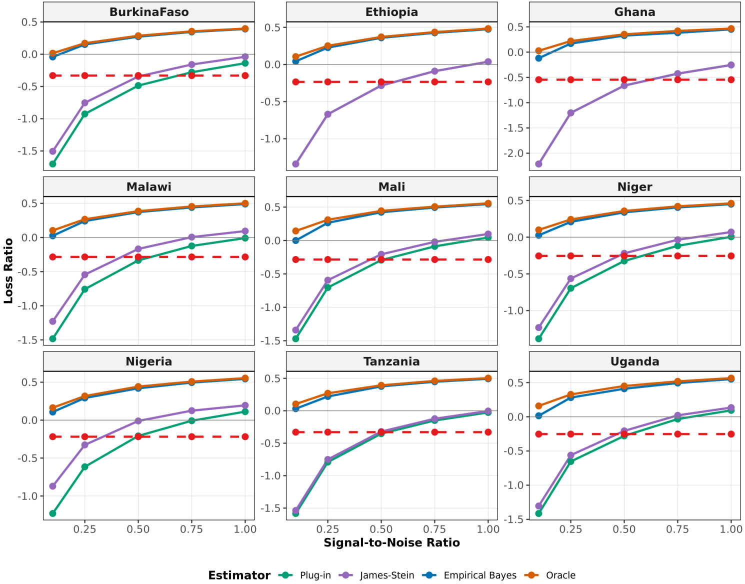
- Targeting rules compared:
  1. Plug-in
  2. James–Stein shrinkage
  3. Oracle Bayes (knows G_c)
  4. Empirical Bayes (NPMLE $\hat{G}_c$)
  5. UBI (uniform transfers B_c/n_c)

- Performance metric: **Loss Ratio**

$$\mathrm{LR}(\tau) = \frac{L(\tau, y) - L(0, y)}{L(\tau^\star, y) - L(0, y)}.$$

Interpretation: $1 =$ optimal; $0 =$ no transfer; $(0, 1) =$ partial gain; $< 0 =$ worse

Simulation Results: Varying Signal-to-Noise Ratios



Simulation Results: Additional Metrics (Burkina Faso)

	Plug-in	J-S	EB	Oracle	UBI
Av. transfer $\mid > 0$	11.58	10.58	1.30	1.27	0.28
% Pop. reached	0.02	0.02	0.22	0.22	1.00
Inclusion error	0.30	0.30	0.32	0.32	0.50
Exclusion error	0.97	0.97	0.70	0.70	0.00

- Plug-in: aggressive — large transfers to $\sim$2% of households.
- EB/Oracle: smaller transfers, reach $\sim$22% of the population; reduce exclusion error sharply while keeping inclusion error stable.
- UBI: universal coverage, but very high inclusion error.

Conclusion

- Takeaway: I study how to optimally allocate transfers with noisy data, a decision-theoretic complement to the targeting literature.
- My contribution: Establish that the plug-in rule is inadmissible and propose empirical Bayes approach
- Future research: This paper is only a first step — extending the framework to dynamics, general equilibrium effects, and behavioral responses

Roadmap of Talk

Targeting Model

Inadmissibility of Plug-in Rule

(Empirical) Bayes Rule

Simulation

Bibliography

Bibliography I

Abadie, A. and Kasy, M. (2019). Choosing among regularized estimators in empirical economics: The risk of machine learning. *Review of Economics and Statistics*, 101(5):743–762.

Areias, A. and Wai-Poi, M. (2022). Machine learning and prediction of beneficiary eligibility for social protection programs.

Baez, J., Kshirsagar, V., and Skoufias, E. (2019). Adaptive safety nets for rural africa: Drought-sensitive targeting with sparse data. *World Bank Policy Research Working Paper*, (9071).

Brown, C., Ravallion, M., and Van de Walle, D. (2018). A poor means test? econometric targeting in africa. *Journal of Development Economics*, 134:109–124.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766.

McBride, L. and Nichols, A. (2018). Retooling poverty targeting using out-of-sample validation and machine learning. *The World Bank Economic Review*, 32(3):531–550.

Roadmap of Talk

Appendix

Why Use the Squared Poverty Gap for All Households?

- **Applies to everyone.** Each household contributes $(z - y_i - \tau_i)^2$, even if $y_i \geq z$.
  - Under full information, rich households still receive $\tau_i^\star = 0$, so this is innocuous.

- **Policy.** With imperfect information, giving to the nonpoor is costly.
  - The term for $y_i \geq z$ penalizes *inclusion errors* and captures the *opportunity cost* of misallocating budget.
  - It also proxies real governance costs (legitimacy, audits, political backlash) that arise when nonpoor households receive transfers.

- **Statistical.** Smooth loss enables clean risk comparisons.
  - Squared loss lets us apply Stein-type arguments for uniform risk reduction.
  - A truncated loss $L^+ = (z - y_i - \tau_i)_+^2$ introduces kinks that complicate the analysis.

- **Practical.** Conclusions are the pretty much the same using $L^+ = (z - y_i - \tau_i)_+^2$.
  - In simulations, EB still outperforms the plug-in rule under L^+; the simple JS shrinkage gains are smaller but remain directionally similar.

▸ Back to Full Information Slide

Plug-in Rule

Geographic Targeting

- Use representative survey data to estimate the average income and then target using that estimate. Specifically, the noisy estimate is given by

$$\hat{y}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij}, j = 1, \ldots, J,$$

where j denotes administrative regions, y_{ij} the income or consumption of households in region j, and N_j the number of households in region j.

- The normal distribution assumption makes sense as long as the population from which the samples are drawn has a finite variance and the effective sample size for each administrative region is relatively large.

Back

Plug-in Rule

Proxy Means Test

- Use observable household characteristics, such as household assets or housing quality, as proxies for household income or consumption:

$$y_i = W_i' \beta + \epsilon_i, \, i = 1, ..., N.$$

- Then, using a dataset of the entire relevant population - usually a census or administrative registry - the predicted income, $\hat{y}_i = W_i' \hat{\beta}$, is used for targeting
- If the support of covariates is discrete and the regression is fully saturated $\Rightarrow$ the predicted income is the average income for each of the different types of households
- If the support of covariates is infinite or the regression is not saturated, the normality of $\hat{y}_i$ rests on the normality of the OLS estimator as the sample size grows to infinity.

▸ Back

Loss Equivalence Between $\boldsymbol{y}$ and $\boldsymbol{\mu}$

- Model: $y_i = \mu_i + \varepsilon_i$, with $\mu_i = \mathbb{E}[y_i \mid W_i]$ and $\mathbb{E}[\varepsilon_i \mid W_i] = 0$.
- Quadratic loss decomposes as

$$L_i(\tau_i, y_i) \ = \ L_i(\tau_i, \mu_i) + \mathbb{E}[\varepsilon_i^2].$$

- Summing:

$$L(\boldsymbol{\tau}, \boldsymbol{y}) \ = \ L(\boldsymbol{\tau}, \boldsymbol{\mu}) + \sum_i \mathbb{E}[\varepsilon_i^2].$$

- The additive term $\sum_i \mathbb{E}[\varepsilon_i^2]$ is independent of $\boldsymbol{\tau}$.
- $\Rightarrow$ Minimizing with respect to true incomes $\boldsymbol{y}$ is equivalent to optimizing with respect to conditional means $\boldsymbol{\mu}$.

▸ Back

(Empirical) Bayes Rule

Feasible rule under truncated normal

- Suppose $\sigma_i = 1$ for $i = 1, \ldots, N$ and G is a normal truncated at zero with μ and γ^2
- The posterior mean for y_i given $\hat{y}_i$ is:

$$\mathbb{E}[y \mid \hat{y}_i] = \delta(\hat{y}_i) + \frac{\varphi\left(-\frac{\delta(\hat{y}_i)}{\nu}\right)}{1 - \Phi\left(-\frac{\delta(\hat{y}_i)}{\nu}\right)}\nu, \text{ where } \delta(\hat{y}_i) = \left(\frac{\gamma^2}{\gamma^2+1}\right)\hat{y}_i + \left(\frac{1}{\gamma^2+1}\right)\mu,$$

and $\nu^2 = \gamma^2/(\gamma^2+1)$.

- The empirical Bayes approach uses estimates of μ and γ:

$$\hat{\tau}_i^* = \max\left\{0, z - \hat{\delta}(\hat{y}_i) - \frac{\varphi\left(-\frac{\hat{\delta}(\hat{y}_i)}{\hat{\nu}}\right)}{1 - \Phi\left(-\frac{\hat{\delta}(\hat{y}_i)}{\hat{\nu}}\right)}\hat{\nu} - \frac{\hat{\lambda}^*}{2}\right\},$$

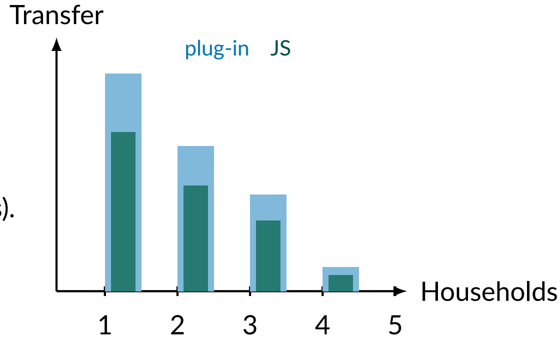
$\Rightarrow$ the optimal transfer is a decreasing function of the posterior mean that shrinks noisy estimates toward the overall mean and an adjustment term because of G's truncation Back

Two Approaches to Constructing Signals

- **Two distinct setups.**
  - (i) *Unbiased noisy signals*: $\hat{y}_i$ is an approximately normal sample mean.
    - Examples: geographic targeting (village means), PMT with saturated cells, etc.
  - (ii) *Predictions from covariates* X_i: ML predictions (ridge, BART, XGBoost), which build in shrinkage and treat $\hat{y}_i$ as known.
- **Why focus on (i)?**
  - Very common in practice: governments routinely target using village averages or PMT-type cell means.
  - Exactly the "noisy but unbiased" signals our EB framework is designed for.
- **Why not (ii)?**
  - ML plug-ins correspond to Bayesian priors on $E[Y|X]$.
  - By Prop. 4, if you are willing to commit to such a prior, posterior plug-in is optimal.
  - But priors are rigid (Gaussian, Laplace, tree-based) and often misspecified.
- **Why EB?**
  - Nonparametric EB flexibly learns the latent distribution from the signals.
  - NPEB dominates parametric shrinkage when p is large, there is no sparsity, and the parameter distribution is highly dispersed (Abadie and Kasy , 2019).
  - These are precisely the conditions of the poverty targeting problem: large admin data, wide income dispersion, no sparsity.

Toy Example

- Example: $\hat{Y} = (20, 35, 45, 60, 100)$, $z = 100$, $B = 100$, $\sigma^2 = 1$.
- Plug-in projection: $(45, 30, 20, 5, 0)$.
- Here $s = 4$ (four households receive transfers).
- Shrink factor $\approx 0.73 \Rightarrow$ JS allocation: $(32.9, 21.9, 14.6, 3.7, 0)$.
- Note: JS rule may not fully spend the budget.
- **Takeaway:** Same recipients, moderated magnitudes, and even with unused budget $\Rightarrow$ lower expected loss.



Back

Simulation Results: One SNR Level

Signal-to-noise ratio = 0.25